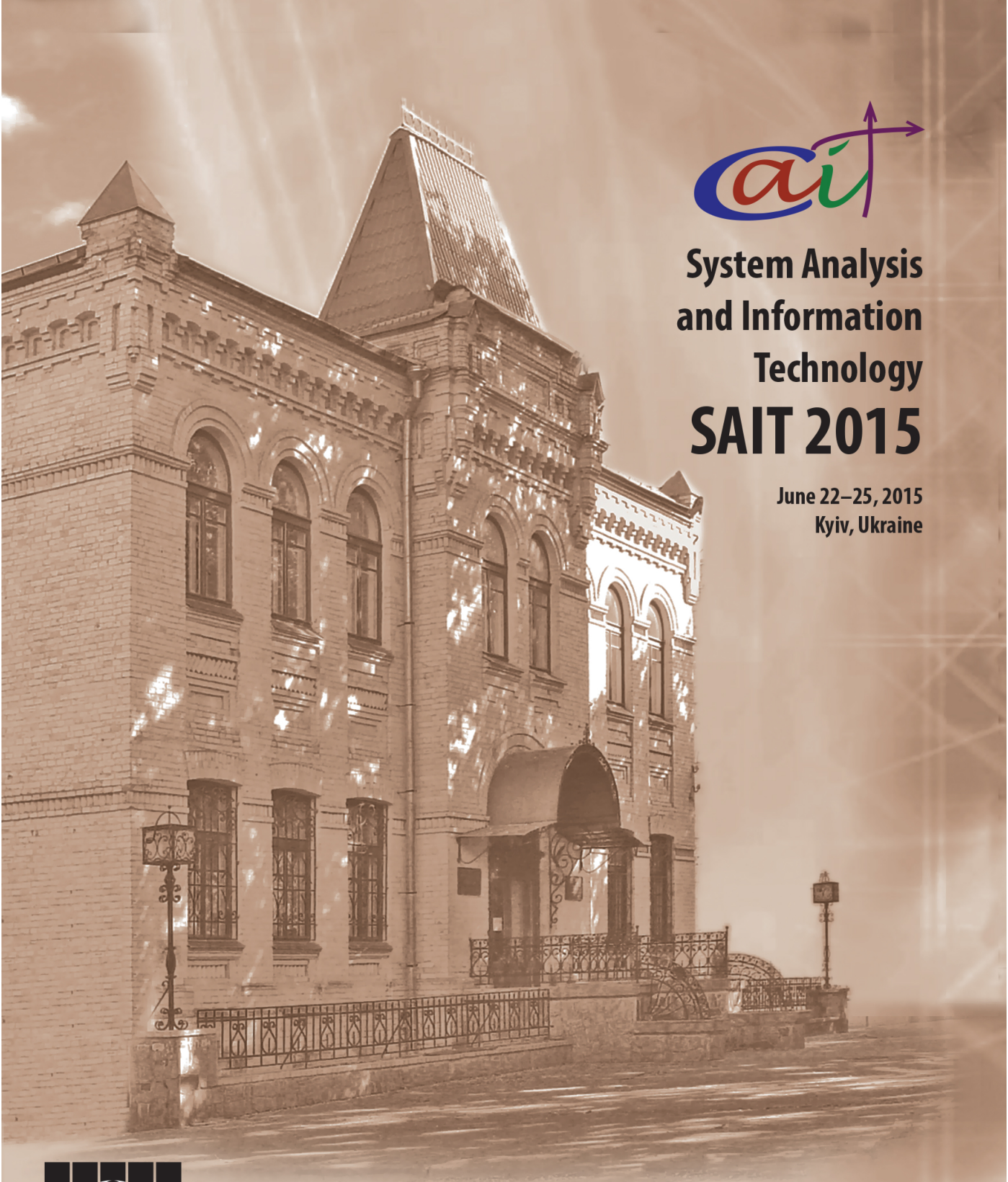

System Analysis and Information Technology

SAIT 2015

June 22–25, 2015
Kyiv, Ukraine

IASA

Institute for Applied System Analysis

National Academy of Sciences of Ukraine
Ministry of Education and Science of Ukraine
National Technical University of Ukraine "Kyiv Polytechnic Institute"

Nataliya D. Pankratova (Ed.)

System Analysis and Information Technologies

17-th International Conference SAIT 2015
Kyiv, Ukraine, June 22-25, 2015

Proceedings

IASA

Institute for Applied System Analysis
of National Technical University of Ukraine
"Kyiv Polytechnic Institute"

UDC [519.7/.8:(004+007)](100)(06)
ББК 22.18я43+72я43
С40

Volume editor:
Nataliya D. Pankratova, Dr.Sc., Prof.

Editorial board:
Petro I. Bidyuk, Dr.Sc., Prof.
Nataliya D. Pankratova, Dr.Sc., Prof.
Anatoliy I. Petrenko, Dr.Sc., Prof.
Yuriy P. Zaichenko, Dr.Sc., Prof.
Elena L. Oparina

Revising:
Dmytro G. Didenko, Ph.D.
Gennadii D. Kiselyov, Ph.D.
Elena L. Oparina
Illia O. Savchenko, Ph.D.
Nataliia V. Shapoval

Design and typesetting:
Mykhailo P. Makukha

System analysis and information technology: 17-th International conference SAIT 2015, Kyiv, Ukraine, June 22–25, 2015. Proceedings. – ESC "IASA" NTUU "KPI", 2015. – 304 p.

С40 **Системный анализ и информационные технологии**: материалы 17-й Международной научно-технической конференции SAIT 2015, Киев, 22-25 июня 2015 г. / УНК "ИПСА" НТУУ "КПИ". – К.: УНК "ИПСА" НТУУ "КПИ", 2015. – 304 с. – Текст: укр., рус., англ.

С40 **Системний аналіз та інформаційні технології**: матеріали 17-ї Міжнародної науково-технічної конференції SAIT 2015, Київ, 22-25 червня 2015 р. / ННК "ІПСА" НТУУ "КПІ". – К.: ННК "ІПСА" НТУУ "КПІ", 2015. – 304 с. – Текст: укр., рос., англ.

This book of abstracts includes issues connected with the research and development of complex systems of various nature in conditions of uncertainty and multifactor risks, Grid and high performance computing in science and education, intelligent systems for decision-making, progressive information technologies for needs of science, industry, economy, and environment. The problems of sustainable development and global threats estimation, forecast and foresight in tasks of planning and strategic decision making are investigated.

В сборнике рассматриваются вопросы, связанные с разработкой и исследованием сложных систем разной природы в условиях неопределенности и многофакторных рисков, Grid и систем высокопроизводительных вычислений в науке и образовании, интеллектуальных систем поддержки принятия решений, прогрессивных информационных технологий для потребностей науки, промышленности, экономики, окружающей среды. Исследуются вопросы устойчивого развития и оценивания глобальных угроз, прогноза и предвидения в задачах планирования и принятия стратегических решений на уровне регионов, больших городов, предприятий.

У збірнику розглядаються питання, що пов'язані з розробкою та дослідженням складних систем різної природи в умовах невизначеності та багатофакторних ризиків, нових інформаційних технологій, Grid і систем високопродуктивних обчислень в науці і освіті, інтелектуальних систем підтримки прийняття рішень, прогресивних інформаційних технологій для потреб науки, промисловості, економіки та навколишнього середовища. Досліджуються питання сталого розвитку та оцінювання глобальних загроз, прогнозу та передбачення в задачах планування та прийняття стратегічних рішень на рівні регіонів, великих міст, підприємств.

© Institute for Applied System Analysis NTUU "KPI", 2015
ISBN 978-966-2748-68-0 (print)
ISBN 978-966-2748-69-7 (ebook)
http://sait.kpi.ua

Table of contents · Содержание · Зміст

Plenary talks · Пленарные доклады · Пленарні доповіді **12**

Pankratova N.D. Modelling and mitigation of social disasters of various nature 13

Валькман Ю.Р. Когнитивная семиотика: определения понятия и проблемы 15

Згуровский М.З. Большая Солнечная спираль "разогрева" системных мировых конфликтов 18

Section 1. System analysis of complex systems of various nature
Секция 1. Системный анализ сложных систем разной природы
Секція 1. Системний аналіз складних систем різної природи **20**

Bakhrushin V.E. Educational legislation from the standpoint of system analysis 22

Bilushchak Yu.I., Chernukha O.Yu., Gera B.V., Goncharuk V.Ye. Software for description of diffusion by two ways with cascade particle decay 23

Chaplya Ye.Ya., Chernukha O.Yu., Davydok A.Ye., Pabyrivskyi V.V. Diffusion equation for mass flux and problem of initial and boundary conditions 25

Chereda H.O., Nikolaiev S.S. Automatic annotation of digitized ECG signals with wavelets 27

Cojocaru S., Gaindric C., Secrieru Iu., Popcova O. The SonaRes platform and its applications 29

Ilchenko K.O., Pyshnograiev I.O. Knowledge Society Assessment as a Composite Index . . 30

Ivanchuk M.A., Malyk I.V. Linear separation using ε-nets 31

Koval O.V., Volkova K.A. Computer modelling of management information systems with Actionable analytics 32

Makukha M.P., Oparina E.L. Situational analytical center for monitoring, modeling, and mitigation of social disasters caused by catastrophes and terrorism 33

Paliichuk L.S. On the limit cycles for controlled piezoelectric fields 34

Pankratova N.D., Pyvovar V.V. Big Data Overview 35

Parkhomchuk D.M., Tymoshenko Yu.O. Analysis of second order dynamic method 36

Petic M., Cojocaru S. Vocabulary enriching for text analysis 37

Ptuha Yu.A., Slashcheva T.G., Slyota M.R., Tarasiuk T.S. Modern problems in prevention and treatment of cardiovascular disease 39

Savastiyanov V.V. Discovering of potential positive and negative factors of social disaster using sentiment analysis 41

Savchenko I.O. Modeling complex socio-technical systems using networks of morphological tables 43

Schuka R.V., Ivanov S.S. Analysis of crime in Ukraine with using methods of regression analysis in SAS Enterprise Guide system . . . 44

Shcherban M.M., Bidyuk P.I. Modeling and forecasting macroeconomic processes . . . 45

Tripolsky O.V. Probabilistic methods for studying of optimal gambling systems . . . 47

Белошицкая О.К. Исследование поведения нейронной сети головного мозга при различных патологиях с помощью методов нелинейной динамики . . . 48

Бояринова Ю.Е., Калиновский Я.А., Туренко А.С. Обобщение кватернионов вращения и связь с процедурой удвоения Грассмана-Клиффорда . . . 50

Васильев В.И., Виштальь Д.М., Любашенко Н.Д. Оценка стационарных характеристик надежности коммуникационных сетей . . . 51

Винокуров Д.С. Деякі підходи до модифікації методу Динамічного програмування Белламана . . . 53

Віжевський П.В., Чертов О.Р. Порівняльний аналіз моделей швидкості осадження частинок бруду під час очистки води . . . 54

Візнюк О.В. Ідентифікація динамічних систем прямими методами оптимізації в умовах пасивних експериментів . . . 56

Гоблик А.В. Комп'ютерне моделювання поля градієнта інтенсивності тяжіння населення до об'єктів громадського обслуговування . . . 57

Горбань І.М., Хоменко О.В. Керування пристінковим потоком за допомогою інтерцептора і відбору рідини . . . 59

Горбань Н.В. Сильний глобальний атрактор тривимірної модифікованої системи Нав'є-Стокса в необмеженій каналоподібній області . . . 60

Горелова Г.В. Предпроектные когнитивные исследования для разработки интеллектуализированных систем поддержки принятия решений . . . 61

Грайворонська А.М., Ланде Д.В. Дослідження інформаційних потоків, як динамічних мультиагентних систем . . . 62

Громова Ю.А. Особливості інноваційного розвитку України . . . 64

Довгаль К.І., Чертов О.Р. Байесові мережі для моделей оцінки якості програмного забезпечення . . . 66

Донченко В.С., Назарага І.М., Тарасова О.В. Регресії з матричними представниками : теорія та приклади застосування . . . 68

Заец Р.В. Концептуализация системных представлений и анализа научно-технической политики . . . 69

Зражевська Н.Г. Оптимізаційна модель для оцінювання параметра Херста . . . 71

Кірік О.Є. Декомпозиційний підхід до аналізу великих просторово розподілених об'єктів мережевої структури . . . 72

Козуля Т.В., Білова М.О. Підвищення ефективності комплексної методики оцінки екологічності системних об'єктів різного рівня дослідження . . . 74

Козуля Т.В., Ємельянова Д.І. Методичне забезпечення комплексної оцінки стану природно-техногенних систем . . . 76

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Курмак И.С. Модель популяции “хищник-жертва” с нелокальным потенциалом 78

Лазарович О.М. Дослідження деяких властивостей гаусівської багатопараметричної півгрупи 79

Лисецкий Ю.М., Снитюк В.Е. Формальное представление корпоративной интегрированной информационной системы как совокупности математических моделей . . . 80

Лопатин А.К., Черненко О.Б. Вероятностные модели оценки фазы финансовых рынков на основе методов нелинейной динамики 82

Маслянко П.П., Маркова Н.М. Спосіб оптимізації маршрутів доставки товарно-матеріальних цінностей Рівненської атомної електростанції 84

Маслянко П.П., Оріщенко А.О. Компонентна модель кластерного аналізу організаційних систем 86

Надточий А.В. Оптимизация инвестиционного портфеля в условиях неопределенности 88

Найдьонов І.М. Побудова оптимальних маршрутів транспортних засобів з урахуванням топології дорожньої мережі 89

Недашковская Н.И. Метод M_Outflow поиска наиболее несогласованных элементов матрицы парных сравнений 90

Пазинич І.С., Ланде Д.В. Дослідження алгоритмів інформаційного пошуку у піринговых мережах 91

Панкратов В.А. Выявление основных отраслей экономики Украины до 2020 г. 93

Панкратова Н.Д., Пинчук В.А., Шишко Н.В. К решению задачи структурной оптимизации сложных объектов 94

Печурин Н.К., Кондратова Л.П., Печурин С.Н. Модель распределения трафика в беспроводных компьютерных сетях 95

Потебня А.В., Погорілий С.Д. Дослідження процесів поширення інформації у наукометричних базах даних 97

Путренко В.В. Використання глобальних моделей інтерполяції у інтелектуальному аналізі геопросторових даних 103

Романенко В.Д., Милявский Ю.Л. Инструментарий адаптивного и координирующего управления импульсными процессами когнитивных моделей сложных систем . . . 105

Романкевич А.М., Романкевич В.А., Морозов К.В. Об одном условии модификации реберных функций GL-модели 106

Романчук К.Г., Стефанишин Д.В. Про використання сценарного підходу та методу Байєса при оцінці ризиків системних аварій на гідровузлах 107

Савін Є.В., Олефір О.С. Моделювання залежних збитків від нещасних випадків на сонові копула-моделей 108

Стефанишин Д.В., Ходневич Я.В., Щодро О.Є. Про ідентифікацію шорсткості природних русел для математичного моделювання руслових потоків в нижніх б’єфах гідроспоруд 110

Стефанишина-Гаврилюк Ю.Д., Стефанишин Д.В. Про один метод прийняття рішень в природокористуванні з врахуванням ризику невикористаних можливостей на основі попарного порівняння альтернатив . . . 111

Стефанович Т.О., Щербовських С.В. Аналіз причин непрацездатності системи із містковою структурою, в якій внутрішній елемент функціонує у ненавантаженому резерві . . . 112

Стецюк Р.В., Валькман Ю.Р. Проектирование онтологии для правовой предметной области . . . 114

Ус С.А., Станіна О.Д. Алгоритм розв'язування задачи оптимального розбиття множин з додатковими зв'язками . . . 116

Шаронова Н.В., Козуля М.М. Інформаційні та об'єктивні загальносистемні особливості функціонування природно-техногенних систем . . . 117

Шелест А.В. До побудови стратегії сталого розвитку економіки України за допомогою кластерного аналізу . . . 119

Шумада К.О. Цифрова система управління штучної вентиляції легенів . . . 120

Яковлева А.П., Курдуп І.О. Задача про розподіл ресурсів між підприємствами . . . 121

Ясинський В.К., Юрченко І.В. Поведінка сильного розв'язку стохастичних рівнянь в частинних похідних з марковськими параметрами . . . 122

Section 2. Intelligent systems for decision-making
Секция 2. Интеллектуальные системы принятия решений
Секція 2. Інтелектуальні системи прийняття рішень **124**

Fridman R., Sagan V.Yu., Slashcheva T., Tymoshenko Yu.O. Approach to automated ECG analysis . . . 126

Ivchenko D.A., Bondarenko A.S., Mazur V.G. Big Data Analysis in Health Care: Perspectives and Potential . . . 128

Korshunova I.A., Kopychko S.M. Detecting epileptic seizures from raw EEG data using convolutional neural networks . . . 130

Malinovskiy A.D., Nikolaiev S.S. Classification of the ECG features for automatic heart diseases detection . . . 131

Sokolov P.Ye., Bidyuk P.I. Model selection and combining forecast estimates . . . 133

Tkachenko D.A. Automated university course timetabling using genetic algorithm . . . 135

Zaychenko Yu.P., Ovi Nafas Aghaei Agh Ghamish Banks bankruptcy risk forecasting with application of fuzzy methods . . . 137

Zyma D.S., Bidyuk P.I. Information system for forecasting nonlinear nonstationary processes 139

Акимов В.С., Сергеєв-Горчинський О.О. Визначення оптимального об'єму даних для побудови моделі класифікації . . . 140

Амосов О.В. Проблема вибору базисного вейвлета в задачі ідентифікації музичних нот звукового сигналу . . . 141

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Базалий О.Ю., Бидюк П.И. Интелектуальные методы анализа данных в оценивании кредитоспособности физических лиц . . . 142

Баклан В.С. Анализ инвестиционной привлекательности акций аукционных домов с применением нечетких нейронных сетей Мамдани . . . 143

Барсученко В.В. Применение нейронной сети Back Propogation для стабилизации изображения на смартфонах в условиях нестационарной динамики изменения их положения 144

Бідюк П.І., Устимчик М.В. Використання регресійних моделей для розв'язання задач медіапланування . . . 146

Виклюк Я.І., Леко Т.Б. Розробка simulink моделі для вивчення залежності між появою сонячної активності та виникненням ураганів . . . 147

Вініченко А.М. Семантичні композитні веб-додатки в наукових дослідженнях . . . 148

Гоцко У.І. Використання "fingerprinting" методу для ідентифікації сигналів звукового оточення людини . . . 149

Даниляк О.С., Бідюк П.І. Проектування системи підтримки прийняття рішень для оцінювання і прогнозування стану підприємств . . . 150

Дембовский О.Ю. Проблема операционализации понятия устойчивого развития в системах поддержки решений по планированию развития городов (регионов) . . . 152

Карпенко Д.І. Дослідження методів згладжування даних зі збереженням фази . . . 154

Клунько Д.А. Технология проектирования интеллектуальных систем обучения . . . 155

Кондратенко Н.Р., Снігур О.О. Нечітка логічна система типу 2 для оцінювання перспективності артезіанської свердловини . . . 157

Копичко С.М., Коровай О.В. Аналіз методів агрегації кардинальних експертних оцінок для СППР . . . 158

Кравець І.О. Використання нейронних мереж для ідентифікації та керування слабоформалізованими об'єктами . . . 159

Ліщук К.І. Застосування метолу аналізу ієрархій для оцінки якості програмних продуктів 160

Ляхевич С.В., Тимошенко Ю.О. Архітектура системи моделювання поведінки коптера . 161

Олефір О.С., Савчук О.Б. Порівняння впливу різних факторів на очікувану тривалість життя населення України . . . 162

Павленко А.І. Пошук найкоротшого шляху у динамічному мультиграфі методами мурашиних колоній з одним поколінням . . . 163

Папуч Т.Ю. Аналіз кредитоспроможності юридичних осіб за методами інтелектуального аналізу даних . . . 165

Письменний І.О. Мультиагентні системи в моделюванні підприємств галузі . . . 167

Поворознюк А.І. Мінімізація ризиків лікарських помилок у системах підтримки прийняття рішень при проведенні діагностично-лікувальних заходів . . . 168

Попова М.В., Безносик О.Ю. Застосування методів кластерізації для слабкоструктурованих задач . . . 170

Розпутний М.В. Засоби здійснення логічних висновків над онтологіями . . . 171

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Сидорук І.А., Зайченко Ю.П. Формування оптимального інвестиційного портфеля в умовах невизначеності 172

Тавров Д.Ю., Булах О.Р. Меметичний метод побудови системи нечіткого виведення для визначення ступеня кредитоспроможності особи 173

Филатова А.Е. Программный модуль повышения качества визуализации рентгенологических изображений 175

Фіногенов О.Д., Штрик Р.В. Аналіз можливостей інтернет сервісів знаходження плагіату 177

Чегренець О.С. Острівна модифікація алгоритму оптимізації мурашиними колоніями . 179

Шульган І.В., Боровська Т.М. Узагальнення моделі оптимального агрегування систем «виробництво-розвиток» з урахуванням невизначеностей зовнішнього середовища виробництва 180

Section 3. High Performance Computing and Microsystems Engineering
Секция 3. Высокопроизводительные вычисления и разработка микросистем
Секція 3. Високопродуктивні обчислення та розробка мікросистем **182**

Бабенко О.А., Гіоргізова-Гай В.Ш. Балансування навантаження "хмарних" додатків . 184

Булах Б.В., Крамар О.В. Семантична інтеграція грід та хмарних ресурсів до складу комплексів інженерних обчислень 185

Вергун О.С. Розробка інтерфейсу для комп'ютерних кластерних систем 186

Ігнатченко Л.О. Оптимізація взаємодії веб-сервісів у складі системи GridALLTED . . 187

Капшук О.О. Системи розпізнавання людини за зображенням обличчя в IT-технологиях 188

Капшук О.О., Лугова О.Ю. Аналіз надійності технологій розпізнавання облич в системах аутентифікації користувачів комп'ютерних систем 190

Кирюша Б.А. Анализ модели идеальной линии на ABCD параметрах 192

Киселев Г.Д., Киселева А.Г. Управление сложными техническими системами с помощью контекстно-зависимых сервисов управления 193

Ковтун О.О. Організація хмарних інженерних сервісно-орієнтованих обчислень (SOC) на базі пакету ALLTED 194

Лавренюк А.М., Лавренюк С.І. Аспекти використання мови прогамування C++ AMP для обробки великих обє'мів сейсмічних даних високопродуктивними кластерами з GPGPU 195

Петренко А.І. Сервісно-орієнтоване дослідницьке середовище з складних обчислень . . 196

Петренко О.О. Наука про сервіси, управління та інжиніринг 198

Пикалов А.С. Использование методов вычислительного интеллекта для повышения точности распознования изображения 199

Ревуцька І.В. Задача розміщення елементів при проектуванні електронних схем в системах автоматизованого проектування та методи її вирішення 201

Романов В.В., Ткаченко В.А. О некоторых аспектах шумовых параметров полупроводниковых интегральных микросхем 203

Харченко К.В. Метод визначення відстані до об'єктів за допомогою характеристичних рис в стереозображеннях 205

Section 4. Progressive information technologies
Секция 4. Прогрессивные информационные технологии
Секція 4. Прогресивні інформаційні технології **206**

Belevets M.S., Bidyuk P.I. Systemic approach to estimation of operational risks 208

Boldak A.O., Yefremov K.V. WDC-Ukraine's distributed platform for supporting research data life cycle 210

Kachko N.A. Prime table generation 212

Kyrysha B.A., Kazachenko O.D. Authorization and Accounting for Distributed Systems . . 213

Litvinov V.A., Maistrenko S.Ya., Khurtsilava K.V. Regarding quality of the basic vocabulary of the subject-oriented spellchecking system and criteria of its evaluation 214

Piddubna Yu.I., Bidyuk P.I. Systemic approach to estimation of operational risks 215

Yaremenko V.S. Modelling of semantic Web applications based on annotated REST Web services 217

Акінфієва А.О. Створення бази знань підсистеми контролю успішності студентів вузу . 218

Артюхов В.Г., Бритов О.А. Розрахунок цифрових еліптичних фільтрів в Micro-Cap . 219

Архангельский К.Е. Алгоритмы компьютерного зрения в системах дополненной реальности 221

Арчвадзе Н.Н., Пховелишвили М.Г., Шецирули Л.Д. Шаблоны для Haskell функций с бесконечными рекурсивными типами аргументов 222

Білорус Д.С. Проблематика задачі пошуку і композиції веб сервісів для підтримки бізнес-процессів 226

Булах Б.В., Лашко О.В. Семантичні технології в неруйнівному контролі 227

Галатенко Д.В. Адаптивний фільтр для правильної локалізації підзображень 228

Джунковський В.О., Джунковський Ю.О. Задача вибору лідера в колективах однорідних агентів (роботів) 229

Діденко Д.Г. Нечітке відновлення довідника за кодом та назвою у базі даних 231

Діденко Д.Г., Попова А.В. Прискорення DataMining за допомогою Hadoop для пошуку схованих залежностей даних в БД 232

Дідик С.А., Бідюк П.І. Проблема відтоку абонентів телеком-операторів та можливі підходи до її вирішення 234

Драган А.В. Построение модели фреймворка с пакетным разбиением кода для внедрения в приложение как зависимость 236

Дубовой В.М., Дуда М.Е., Байас М.М. Інформаційна технологія координації багатолінійного технологічного процесу 238

Дубовой В.М., Пилипенко І.В., Стець Р.С. Моделювання процесу тестування програмного забезпечення як розгалужено-циклічного технологічного процесу 240

Душин О.С., Сирота С.В. Модель синхронізації роботи адаптивних систем регулювання дорожнього руху на суміжних перехрестях . . . 242

Забелин С.И. Метод раннего распознавания DDoS атак при помощи SDN . . . 243

Кислий Р.В. Моделювання профілю користувача для уникнення проблеми холодного старту в рекомендаційних системах . . . 244

Ковальчук-Химюк Л.О., Тавров Д.Ю. Система нечіткого виведення для оцінювання успішності учнів під час білінгвального навчання математиці . . . 245

Кондіус К.Ю. Наукові публікації та їх цитування в галузі інтелектуальних технологій 248

Копычко С.Н., Кулык Р.Ю., Дрозденко А.Н. Исследование XML парсеров языка программирования Java . . . 249

Королевський Г.Д. Візуалізація запитів до ресурсів Open Linked Data . . . 250

Котенко Д.А. Шляхи оптимізації роботи з деревоподібними структурами даних . . . 251

Круш І.В. Використання графових баз даних для оптимізація швидкості обробки складних запитів . . . 252

Куліш З.І. Забезпечення інформаційної безпеки в системі дистанційного навчання LON-CAPA . . . 254

Мазурик А.Ю. Создание интерактивной инфографики средствами библиотеки Highcharts.js 255

Малишев А.І. Документування RESTful API . . . 257

Малишев А.І. Розробка способів управління доступом в корпоративному Java-додатку 258

Маматов В.И. Электронная коммерция . . . 259

Медведський А.М. Індексація сторінок побудованих за допомогою JavaScript фреймворків 260

Мельничук М.М. Модель віддаленого керування мікроконтролерними системами . . . 262

Мирошникова І.Ю. Вибір провайдера хмарного сервісу за умови динамічного середовища 263

Михалько В.Г. Використання кешування і попередніх розрахунків для пришвидшення часу відгуку веб-сервісів . . . 264

Нарижна Н.С. Рівні тестування розподілених систем . . . 266

Науменко Т.О. Дослідження адаптивних технологій контролю знань . . . 267

Негода К.В., Ковальчук-Химюк Л.О. Дослідження динаміки ендогенного економічного зростання в Україні протягом 2007–2012 рр. за моделлю Ромера . . . 268

Перевертайло В.В., Майкут С.О., Чернацький А.Ю. Система автоматичного моніторингу параметрів приміщення на базі Arduino Mega 2560 . . . 269

Попеляев Д.П. Применение событийно-ориентированного подхода к разработке веб-сервисов . . . 271

Прогонов Д.А., Паничева Д.О., Кущ С.Н. Анализ точности статистических методов обнаружения стеганограмм с данными, встроенными в области преобразования контейнера . . . 273

Рева К.В., Родіонов А.М. Керування пулом ресурсів із застосуванням машинного навчання з підкріпленням . . . 275

Ротенберг А.В. Выбор модели данных СУБД: специфика примененинния SQL и NoSQL 278

Рубець М.Г., Бідюк П.І. Класифікація типів операційних ризиків та можливості їх математичного опису . . . 279

Савчук А.Ю. Створення сучасної системи навігації з елементами доповненної реальності для мобільних платформ . . . 281

Сарапулов А.С. Сравнение и анализ протоколов передачи мультимедиа в реальном времени 282

Саржинский В.В., Дидковская М.В. Методы машинного обучения в задаче идентификации диктора по голосу . . . 284

Семироз К.В. Побудова системи рекомендацій за допомогою графової бази даних . . . 285

Сливка Т.П. Розпізнавання нот при акапельному співі . . . 286

Смага М.В. Система передачі даних з використанням матричних штрихкодів . . . 287

Смаковський Д.С., Ляшенко М.М. Система пошуку кандидатів на вакансії роботодавців 288

Сутула О.В. Побудова архітектури Android додатку на основі патернів проектування 289

Тавров Д.Ю. Схема інформаційної технології забезпечення групової анонімності даних 290

Тимошенко Ю.А., Романко А.Р. Системный анализ и персональная медицина. Brain reading. . . . 292

Тригубишин І.Л. Виявлення ботів у соцмережах з використанням закону Бенфорда . . 294

Цурін О.П., Слухай Я.О. Web-сайти викладачів Університетів у світі . . . 295

Шевченко Я.Т. Нагрузочное тестирование web-приложений . . . 297

Юдін Є.М. Застосування методу навчання з підкріпленням для керування вивільненням опертативної пам'яті . . . 298

Яковлева Г.О. Автоматизація документообігу в роботі Compliance менеджера . . . 300

Authors · Авторы · Автори **301**

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Plenary talks

System analysis of complex systems of various nature

Intelligent systems for decision-making

High Performance Computing and Microsystems Engineering

Progressive information technologies

Plenary talks

Pankratova N.D.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Modelling and mitigation of social disasters of various nature

Currently disaster losses are increasing with grave consequences for the survival, decent living conditions and livelihoods of people. Disaster risk is increasingly global in nature, its impact and actions in one region can affect the risk in any other region, and vice versa. This, compounded by increasing vulnerabilities related to changing demographic, technological and socio-economic conditions, unplanned urbanization, development within high-risk zones, underdevelopment, environmental degradation, climate variability and change, competition for scarce resources, and the impact of epidemics It may lead to a future where disasters could increasingly threaten the world economy, the world's population and sustainable development of developing countries. In the past two decades, every year by disasters suffered an average of over 200 million people.

In this paper the methodology for scenario analysis regarding the field of social disasters is applied. The mathematical forecast models are useful while the process is monotonous, which is rarely the case for social disasters. During social disasters, certain events can happen that significantly change the attitudes and sentiments of the involved sides of conflict, causing the usual forecasts to be incorrect or inaccurate. Moreover, social disasters can lead to other accidents and catastrophes, e.g. explosions leading to oil spills and contamination of soils and water. To analyze these cases in situations of uncertainty and multi-factor risks, a certain methodology called foresight is developed. In such cases it is advisable to apply modified qualitative methods of the foresight methodology (Delphi method, morphological analysis method, analytical hierarchy method) to develop models suitable for description of the social disasters [1, 2]. These models will then be used to construct scenarios for the social disaster, allowing to assess the mitigation methods and provide decision-making support regarding the best actions to reduce the negative consequences of a social disaster in the context of sustainable development.

The application of the proposed above methodologies and techniques to solve problems associated with environmental, technical and social disasters as a software tool is developed [3]. On the basis of system methodology using the foresight methods the problems in direction of environmental, technical andsocial disasters were realized.

In this paper the integrated modeling tools of modeling and microbiological technologies for restoration of oil-polluted soils and water bodies are proposed. Application of this tools on the stage of modeling allows to determine the geometrical parameters of pollution, its density, concentration, rate of distribution, to study the forecast and foresight of the spread of contamination and then to efficiently solve the problem of calculating the required volume and concentration of the sorbent. As soon as the stage of modeling is fulfilled and the required volume of the sorbent is defined the stage of the sorbent production is carried out. The only feasible way of dealing with oil spills nowadays is a complex of measures, including modelling, new technology, safety, low cost and high effect. Proposed toolkit is based on the synthesis of modeling and technology of cleaning soil and water from oil products [4].

Analysis of accidents and catastrophes can identify the most important causes and weaknesses of control principles for survivability and safety of complex engineering objects (CEO). One of such reasons is the peculiarities of the functioning of the diagnostic systems aimed to identify failures and malfunctions. This approach to safety precludes a possibility of a priori prevention of abnormal modes and as a consequence, there is the possibility of its subsequent transition into an accident and catastrophe. Therefore, it is necessary to develop a new strategy to solve safety problems of modern CEO for various purposes. Here a strategy that is based on the conceptual foundations of systems analysis, multicriteria estimation and risk forecasting is proposed [5, 6]. The essence of the proposed concept is the replacement of a standard principle of identifying the transition from operational state of the object into inoperable one on the basis of detection of failures, malfunctions,

This research was sponsored in part by the NATO Science for Peace and Security Programme under grant G4877.

defects, and forecasting the reliability of an object by a qualitatively new principle. The essence of this principle is the timely detection and elimination of the causes of an eventual transition from operational state of the object into inoperable one on the basis of systems analysis of multifactorial risk of abnormal situations, a reliable estimation of margin of permissible risk of different modes of complex engineering objects operation, and forecast the key indicators of the object survivability in a given period of its operation.

The proposed strategy of system coordination of survivability and safety engineering objects operation, implemented as a tool of information platform of engineering diagnostics of the complex objects, ensures the prevention of inoperability and the danger of object's functioning. By force of systematic and continuous evaluation of critical parameters of object's functioning in the real time mode, the reasons, which could potentially cause the object' tolerance failure of the functioning in the normal mode, are timely revealed. For situations, development of which leads to possible deviations of parameters from the normal mode of the object's functioning, it is possible timely to make a decision about the change of the operation mode of the object, or an artificial correction of the parameters to prevent the transition from the normal mode into the abnormal one, accident and catastrophe. The used principles provide the flexible approach to timely detection, forecasting and system diagnosis of factors and risk situations, formation of sustainable solutions during the acceptable time within the fatal time limit.

The problem of society connected with drug addiction is considered and the threat of society is formulated: the growth in the structure of crimes committed fraud, drug trafficking, raids, corruption, kidnapping; increase in the number of crimes in the sphere of illegal drug trafficking and consumption, as well as committed on the grounds of ethnic and religious hostility.

Investigation of semi structured problems of society in the modern metropolis demanded implementation of a multidisciplinary approach, which is based on models and methods of cognitive modeling of complex systems, adapted to the particular problem. A study on cognitive maps developed modeling enabled to consider the various aspects of the problem and consider a variety of possible scenarios. All this together allows us to develop comprehensively analyze and validate appropriate management decisions.

References. **1.** Pankratova N.D, Savchenko I.A Morphological analysis. Problems, theory, application // Naukovadumka, Kyiv. – 2015.– 347 p. **2.** Pankratova N.D., Malafeeva L.Y. A system approach to solving foresight problems on the basis of Delphi method // International Journal. "Information technologies&knowledge". ITHEA. SOFIA. – 2013. V.7, – №4. – P.313–324. **3.** Zgurovsky M.Z., Pankratova N.D. System analysis: Theory and Applications //Springer.– 2007. – 475 p. **4.** Pankratova N, Khokhlova L. Integrated tools for restoration of oil polluted soils and water bodies // International Journal "Information theories&applications". ITHEA. SOFIA. – 2012. – V.12, №1. – P.39–49. **5.** Pankratova N.D., System strategy for guaranteed safety of complex engineering systems // Cybernetics and Systems Analysis. – 2010. – V.46, №2. – P.243–251. **6.** Pankratova N.D. Safety operations of the complex engineering objects // International Journal. "Information technologies&knowledge". ITHEA. SOFIA, – 2011. – V.5, №2. – P.152–167.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Валькман Ю.Р.
Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины, Киев, Украина

Когнитивная семиотика: определения понятия и проблемы

В работе определяется предмет когнитивной семиотики, исследуются проблемы и задачи этой науки.

Введение. Объектом данных исследований является моделирование когнитивных процессов в вычислительной среде. Здесь, предмет исследования – семиотические аспекты когнитивных процессов. Цель данного исследования – определение предмета и проблематики когнитивной семиотики.

О семиотике. Самым распространенным, классическим определением семиотики является определение по объекту: *"Семиотика - это наука о знаках и/или знаковых системах"*. Определение второго типа - определение по методу: *"Семиотика - это приложение лингвистических методов к объектам иным, чем естественный язык"*. Ч. Моррис полагал, что семиотика — наука широкая, интегративная, *"Логика, математика и лингвистика могут быть включены в семиотику полностью"*. Одно из самых широких определений ей даёт Умберто Эко, который заявляет, что *"семиотика интересуется всем, что может быть принято за знак"*, т.е., все сводится к знакам.

Широко известен семиотический треугольник Г. Фреге, К. Огдена, И. Ричардса, С. Ульриха, А. Черча и др., который связывает S (*знак*), C (*концепт, понятие*) и D (*денотат*). Вершине D соответствует трактовка знака S и денотата D (точнее отношения S-D для обеспечения перехода от знака к денотату или наоборот), которая производится в голове интерпретатора. Таким образом, знаком S можно считать все, что является обозначением чего-либо, т.е. денотата без знака не бывает (и знака без денотата не может быть). А концепт C – это наши мысли об этом знаке, которые позволяют, с одной стороны, синтезировать этот знак S, а с другой – распознать его денотат D. Заметим, в роли знака могут выступать не только слова, словосочетания, формулы, но и мимика, некий материальный объект (например, столб, флаг и т.п.), некоторая мысль. Также денотатом может быть значение интеграла, представление, например, русалки, а концептом может быть схема или фрейм. Ч. Пирс считал, что семиотика делится на три части – *синтактика* (правила, как корректно складывать знаки), *семантика* (как знаки и знаковые структуры связываются со своими денотатами) и *прагматика* - наука об отношении (например, использовании) создателей и получателей знаков к своим знаковым произведениям.

Эпистемология, когнитология и семиотика. Великий славянский философ восемнадцатого века Григорий Сковорода в своем сочинении «Потоп змиин» описал концепцию трех миров. Согласно его описанию, в первый мир входит окружающая нас реальность (по-английски Reality – R). Во второй мир входит внутренний мир человека (английский термин Mind или). Третий мир, согласно Г. Сковороде, есть мир знаков (signs – или S). Независимо от Сковороды к открытию трех миров пришли некоторые западные философы XX века – например, Карл Поппер.

Когнитологию понимают, как область междисциплинарных исследований познания, понимаемого как совокупность процессов приобретения, хранения, преобразования и использования знаний живыми и искусственными системами. Когнитивная наука поставила перед собой задачи чрезвычайной сложности: *понять, как человек воспринимает мир; в каких структурах знания отражает он результаты своего восприятия; как он приходит к знанию; в каком виде оказывается оно представленным в его голове; какими типами репрезентации владеет человек; как манипулирует ими в разного рода мыслительных процессах; какую роль играют в этих процессах память и воображение, фантазии и сенсомоторный опыт; на какие когнитивные способности и механизмы в голове человека указывают сами перечисленные феномены - рациональное мышление, разумное поведение и его планирование и т.д.*

Обратим внимание, что структура семиотического треугольника содержетельно полностью соответствует структурам миров Г. .Сковороды и К. Поппера. Так,

- денотат D – это окружающая нас реальность (R),
- концепт C - это внутренний мир человека (M),
- знак S, естественно, соответствует миру знаков (S).

Таким образом, очевидна тесная связь эпистемологии, когнитологии и семиотики.

Когнитивность семиотики. В настоящее время наблюдается бум появления когнитивных наук, дисциплин, технологий, приложений: *когнитивная лингвистика, когнитивная архитектура, когнитивная семантика, когнитивная социология, когнитивная география, когнитивная экономика, когнитивный маркетинг, когнитивная эргономика* и т.д.

Так же, можно говорить о семиотике почти всех предметных и проблемных областей, например, о *семиотике телешоу, семиотике костюма, семиотике математики, семиотке балета* и т.д.

Но, к когнитивной семиотике у нас особое отношение. Когнитивность изначально явно и неявно содержалась в определении и исследовании семиотических процессов. Широко известно высказывание одного из основоположников семиотики Ч. С. Пирса *"мы думаем только в знаках"*. Другой родоначальник семиотики Ч. Моррис, развивая учение Ч. Пирса, сформулировал (наряду с синтактикой, семантикой) проблематику прагматики, занимающейся исследованием процессов генерации и интерпретации, пониманием знаковых структур субъектом. Все эти процессы являются когнитивными. Согласно Ю. М. Лотману, под семиотикой следует понимать науку о коммуникативных системах и знаках, используемых в процессе общения. А коммуникации осуществляются посредством (и являются отражением) когнитивных процессов. Можно привести еще много доказательств когнитивности семиотики.

Семиотический квадрат. Для исследования когнитивных процессов целесообразно (вслед за Г. П. Щедровицким, Г. П. Мельниковым, и В. В. Рыковым) рассматривать *квадрат* (вместо треугольника). Схема его представлена на рис. 1. В этом четырехугольнике вершина «концепт (понятие)» заменяется отношением S''-D'', в котором S''- образ знака в голове интерпретатора (ментальная модель знака), а D''- ментальный образ денотата. Обозначение $''$ заимствовано из символики Г. П. Мельникова. Им помечаются все объекты, которые представлены в голове интерпретатора (ментальные модели).

Очевидно, важнейшим свойством знака является *его функция отображения* некоторого объекта, т.е. любой знак всегда знак некоторого объекта. Заметим, один и тот же объект может быть денотатом в одной структуре и знаком в другой. Известна коммутативная диаграмма Г. П. Мельникова (см. рис. 1):

$$S \to S'' \to D'' \to D \Leftrightarrow S$$

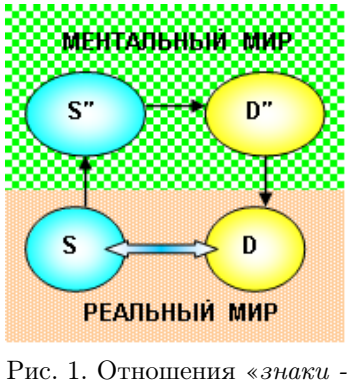

Рис. 1. Отношения «*знаки - денотаты*» в семиотике

По сути – это определение знака. Предположим, мы видим только объект S, далее активизируется его образ S'', затем образ D''. То есть мы не видим объекта D, но его образ D'' активизируется или возникает в нашем сознании! В этом случае мы говорим, что S - это знак, а D - это его денотат. Или, другими словами, знак (S) - это то, что замещает в нашем сознании другой объект (D). Заметим, что основные семиотические процессы $S'' \to D''$ или $D'' \to S''$, восприятия знаков $S \to S''$ или их порождения $S'' \to S$, восприятия денотатов $D \to D''$ проходят в ментальном мире (см. рис. 1) Далее надо говорить о семиозисах, т.е. о схемах порождения и интерпретации знаковых структур.

Тем самым показано, что *семиотика во многом имеет когнитивный характер.*

Когнитивная семиотика. Проблема *знак–значение–понимание-смысл* (главная проблема когнитивной семиотики) тесно связана, *с одной стороны*, с семиотикой, *с другой* – с когнитологией, т. к. речь идет, явно и неявно, о понимании, познании.

Когнитивная семиотика, как ее определил В.В. Рыков, представляет собой парадигму, *соединяющую знаковые и когнитивные структуры с предметной деятельностью* и является удобным инструментом для интеграции и совершенствования систем извлечения знаний из текста.

С нашей точки зрения – это весьма узкая трактовка проблемной области новой науки. В ее рамках целесообразно заниматься всем спектром проблем синтеза и анализа смыслов, значений знаковых структур, процессами понимания знаков и познанием мира посредством знаковых систем.

Когнитивная семиотика должна обеспечивать получение знаний на основе синтеза некоторой целостности путем интеграции структур трех миров Г. СКовороды. А такая интеграция осуществляется на базе *взаимодействия* (синергетического) реального, ментального и знакового миров.

Например, водитель крутит баранку руля, думает о своих предстоящих действиях для продолжения движения и в то же время наблюдает показания приборов на контрольной панели и дорожные знаки. Принимаемые им решения зависят от событий, происходящих в этих трех мирах –

- *окружающего его материального мира (R),*
- *знакового (семиотического) окружения (S),*
- *и его собственного знания о вождении, которое находится в его памяти (M).*

Так называемые когнитивные процессы, результатом которых являются управляющие воздействия на автомобиль, протекают в сознании водителя. Очевидно, что любые осознанные решения принимаются на основе взаимодействия этих трех миров. При этом «вклад» различных миров в синтез знаний существенно зависит от решаемой проблемы, от опыта лица, принимающего решение (его ментальных моделей), от уровня исследований данной проблемы (наличия эффективных формальных, знаковых моделей).

Таким образом, мы *когнитивную семиотику* рассматриваем, как *науку о методах получения знаний из знаковых структур и наоборот, синтеза знаковых конструкций, представляющих соответствующие структуры знаний.* При этом, в основе этих методов находится анализ и синтез взаимодействия трех миров.

Заключение. И когнитология, и семиотика являются междисплинарными науками. Введение понятия когнитивная семиотика, с нашей точки зрения, было обусловлено желанием, *с одной стороны*, вычленить из всей многообразной семиотики проблемы, ориентированные на моделирование мыследеятельности (термин Г. П. Щедровицкого), *с другой стороны*, наоборот, выделить из многих приложений когнитологии проблемы, связанные с синтезом и анализом знаковых структур для представления и использования наших знаний, моделирования смыслов, понимания и объяснения посредством знаковых конструкций. Полагаем, что это направление исследований должно обеспечить разработку формального аппарата для моделирования мышления в компьютерных технологиях. Здесь затронуты только некоторые темы дальнейших исследований.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Згуровский М.З.
Национальный технический университет Украины "КПИ", Киев, Украина

Большая Солнечная спираль "разогрева" системных мировых конфликтов

В докладе сформулированы гипотезы о наличии метрической связи между последовательностью 11-летних циклов солнечной активности Швабе-Вольфа и процессом эволюционного структурирования семейства C-волн системных мировых конфликтов, охватывающих большие и сверхбольшие временные интервалы и имеющие неустойчивую «временную конфигурацию». Указанную связь можно рассматривать, в частности, как еще одно подтверждение, достоверности F-закономерности цивилизационных процессов, опирающееся на глобальный внешний критерий.

В рамках сформулированных гипотез получена формула, связующая основные метрические характеристики трех глобальных периодических процессов, а именно: циклов Солнечной активности Швабе-Вольфа, C-волн системных мировых конфликтов и больших кондратьевских циклов развития мировой экономики.

Построены Большая и Гиперболическая "Солнечные спирали" процесса "разогрева" семействаволн системных мировых конфликтов последовательностью ансамблей циклов Швабе-Вольфа на временном интервале с 840 года до н.э. по 2097 год н.э., как визуальная иллюстрация выявленных закономерностей.

Рассмотрен вопрос формирования локальной конфигурации отдельно взятой C_k-волны системных мировых конфликтов, которую определяет внутренняя иерархия интенсивностей ее эволюционных фаз. Приведены текущие параметры нового, 24-го, наступившего цюриховского Солнечного цикла Швабе-Вольфа, "разогревающего" первую фазу ("Зарождение") прогнозируемой C_7-волны. Особо выделены 2013–2014, как годы активного Солнца.

Построены два возможных сценария "XXI – $2K$" и "XXI – $3K$" развития глобальных цивилизационных процессов в течение седьмого, заключительного системного мирового конфликта в XXI веке. На основе выявленных закономерностей сформулированы основные характерные особенности рассматриваемых сценариев, а также определены их метрические характеристики.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

This page was left blank intentionally

Plenary talks

System analysis of complex systems of various nature

Intelligent systems for decision-making

High Performance Computing and Microsystems Engineering

Progressive information technologies

1 System analysis of complex systems of various nature

Section 1

System analysis of complex systems of various nature

1. System analysis methods for complex systems of various nature in conditions of uncertainty and risks.
2. Mathematical methods, models and technologies for complex systems' research.
3. Technology foresight system methodology in problems of planning and strategic decisions' making.
4. Theory and methods of optimal decision-making.
5. Problem oriented methods of complex systems' analysis and designing in conditions of uncertainty and risks.
6. Nonlinear problems of system analysis.
7. System methodology of sustainable development.

Секция 1

Системный анализ сложных систем разной природы

1. Методы системного анализа сложных систем разной природы в условиях неопределенности и рисков.
2. Математические методы, модели и технологии исследования сложных систем.
3. Системная методология технологического предвидения в задачах планирования и принятия стратегических решений.
4. Теория и методы принятия оптимальных решений.
5. Проблемно-ориентированные методы анализа и проектирования сложных систем в условиях неопределенности и рисков.
6. Нелинейные задачи системного анализа.
7. Системная методология устойчивого развития.

Секція 1

Системний аналіз складних систем різної природи

1. Методи системного аналізу складних систем різної природи в умовах невизначеності та ризиків.
2. Математичні методи, моделі та технології дослідження складних систем.
3. Системна методологія технологічного передбачення в задачах планування та прийняття стратегічних рішень.
4. Теорія та методи прийняття оптимальних рішень.
5. Проблемно-орієнтовані методи аналізу та проектування складних систем за умов невизначеності та ризиків.
6. Нелінійні задачі системного аналізу.
7. Системна методологія сталого розвитку.

Bakhrushin V.E.
Classic Private University, Zaporizhia, Ukraine

Educational legislation from the standpoint of system analysis

Now the new Law of Ukraine on Education is under development. For the success of its implementation it is important that essential features of the education system must be taken into consideration in the Law. In particular:

- complex structure of the system;
- the large number of stakeholders that influence decision-making with the possibility of coalitions and hierarchies of interests;
- openness of the system - a large number of external systems and processes, which significantly affect its state and behavior;
- high inertia - the results of decisions will be visible through the long period of time; for different levels of education delay might be 2 - 10 years;
- large quantity of indicators, by which the system state and effectiveness of decisions are evaluated, complex relationships between these indicators;
- uncertainty of the consequences, the presence of unexpected side effects, fuzziness of available information about the system.

It is important to use reliable and complete information for decision-making. From the standpoint of the education Law preparation there are such important sources of errors:

- use of the accidental data, which do not represent the real state of the system; for example, it may be values that are far from the distribution center, or caused by short-term fluctuations;
- comparison of the data relating to different moments of time for indicators that change rapidly;
- the use of different processing methods of compared data;
- errors in the processing of the raw data;
- deliberate distortion of data.

Well-known examples of erroneous data are comparative information about the total number of higher education institutions; the proportion of high school graduates continuing their education in HEI; the proportion of citizens of appropriate age receiving secondary education, etc. To improve the reliability of data used in decision-making, it is appropriate to provide in the Law the need to develop a system of educational indicators comparable with the indicators of UNESCO and other authoritative international organizations. Also, it should be written some rules on the open data in education and mechanisms of their public control.

Decisions related to education, are usually multiobjective. Therefore, it is desirable to determine in the Law the mechanisms for the main types of decision-making; criteria for the assessment and / or ranking indicators of their effectiveness etc. If it is impossible or impractical to prescribe some of these rules directly in the Law, it should be clearly specified the authority responsible for its adoption. Current Ukrainian legislation on education contains many examples of wrong approach to multicriteria decision-making. In particular the method of calculating the competitive score of applicants for admission to universities is based on the linear convolution of particular indicators. This ignores the fact that used particular indicators by their nature are not the values with which arithmetic operations can be performed. Therefore, it would be appropriate to use other methods of multi-criteria ranking of applicants. Another well-known example is the use of educational institutions ratings for decision-making. At the same time the objectives of these ratings which determine the methods of their compiling, are often ignored.

The large number of stakeholders with different interests causes the usefulness of the game theory approaches in the development and evaluation of legislation. In particular, the Law should define a certain optimal state of the system, the deviation from which is not beneficial to anyone of stakeholders without agreement with others.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

***Bilushchak Yu.I.*[1], *Chernukha O.Yu.*[1], *Gera B.V.*[2], *Goncharuk V.Ye.*[3]**
[1]*Centre of Mathematical Modelling of Y.S.Pidstryhach Institute of Applied Problems of Mechanics and Mathematics of NAS of Ukraine, Lviv, Ukraine;* [2]*Lviv Branch of V. Lazaryan Dnipropetrovsk National University of Railway Transport, Lviv, Ukraine;* [3]*National University "Lviv polytechnic", Lviv, Ukraine*

Software for description of diffusion by two ways with cascade particle decay

The main soil pollutants include industry, such as radioactive waste, industrial waste water, industrial emissions, transport, agriculture, human habitation. Prediction of decaying admixture spread in soils and estimation of groundwater protectability from surface man-made pollution are based on certain model representations on their local spatial redistribution in subsurface layers of the Earth. In the work within the scope of the model of noninteracting flows it is formulated an initial-boundary value problem of diffusion in two ways with cascade particle decay when solution of the problem at one stage is the source on the next step. The processes of type of sorption-desorption are neglected.

Problem formulation. Let decaying particles of one chemical kind migrate in a body with two diffusion ways (fine-dispersed medium, monocrystals, ets.). Moreover, substances produced as result of decay can also decay. As an example of such a cascade decay can instance radionuclides with their radioactive decay

$$^{140}_{54}Xe \longrightarrow {}^{140}_{55}Cs \longrightarrow {}^{140}_{56}Ba \longrightarrow {}^{140}_{57}La \longrightarrow {}^{140}_{58}Ce,$$
$$^{135}_{52}Te \longrightarrow {}^{135}_{53}J \longrightarrow {}^{135}_{54}Xe \longrightarrow {}^{135}_{55}Cs \longrightarrow {}^{135}_{56}Ba$$

or various chemicals (pesticides, nitrates, nitrites, etc.) that decay due to chemical reactions.

The process of mass transfer in two ways with different diffusion coefficients in the absence of mass exchange between them, is described the mathematical model of non-interacting flows, which consists of assembly of two decoupled diffusion equations.

For one-dimensional case in dimensionless variables $\xi^{(\alpha)} = \left(k_2^{(0)}/D_1^{(0)}\right)^{1/2} x^{(\alpha)}$, $\tau = k_2^{(0)} t$; $\alpha = \overline{1;3}$, where $k_2^{(0)}$ and $D_1^{(0)}$ are the coefficient of intensity of total decay and diffusion coefficient on the way $j = 1$ for step $i = 0$; t is time, x^{α} is spatial cartesian coordinate, we have the following assemblies of equations

for the step of decay $i = 0$

$$\frac{\partial c_1^{(0)}}{\partial \tau} = d_1^{(0)} \frac{\partial^2 c_1^{(0)}}{\partial \xi^2} - a_1^{(0)} c_1^{(0)}, \quad \frac{\partial c_2^{(0)}}{\partial \tau} = d_2^{(0)} \frac{\partial^2 c_2^{(0)}}{\partial \xi^2} - a_2^{(0)} c_2^{(0)},$$

for the step of decay $i = \overline{1; N-1}$

$$\frac{\partial c_1^{(i)}}{\partial \tau} = d_1^{(i)} \frac{\partial^2 c_1^{(i)}}{\partial \xi^2} + a_{\lambda 1}^{(i-1)} c_1^{(i-1)} - a_1^{(i)} c_1^{(i)}, \quad \frac{\partial c_2^{(i)}}{\partial \tau} = d_2^{(i)} \frac{\partial^2 c_2^{(i)}}{\partial \xi^2} + a_{\lambda 2}^{(i-1)} c_2^{(i-1)} - a_2^{(i)} c_2^{(i)},$$

for the step of decay $i = N$

$$\frac{\partial c_1^{(N)}}{\partial \tau} = d_1^{(N)} \frac{\partial^2 c_1^{(N)}}{\partial \xi^2} + \sum_{i=0}^{N-1} a_{\lambda 1}^{(iN)} c_1^{(i)}, \quad \frac{\partial c_2^{(N)}}{\partial \tau} = d_2^{(N)} \frac{\partial^2 c_2^{(N)}}{\partial \xi^2} + \sum_{i=0}^{N-1} a_{\lambda 2}^{(iN)} c_2^{(i)},$$

where $c_j^{(i)}$ is concentration of admixture particles $\mathbf{K}_j^{(i)}$ in the jth migration way for the ith step of decay ($i = \overline{0; N}$), $d_j^{(i)}$ is the diffusion coefficient of substance $\mathbf{K}_j^{(i)}$ ($i = \overline{0; N}$, $j = 1; 2$), $a_{\lambda j}^{(i-1)}$ is coefficient of intensity of $\mathbf{K}_j^{(i-1)}$ decay ($i = \overline{1; N}$), $a_{\lambda 1}^{(iN)}$ is coefficient determining the part of non-decaying (or harmless) substance generated after decay $\mathbf{K}_j^{(i)}$ on the ith th step ($i = \overline{0; N-1}$, $j = 1; 2$). The following initial and boundary conditions are imposed

$$c_1^{(i)}(\xi, \tau)\Big|_{\tau=0} = 0, \qquad c_2^{(i)}(\xi, \tau)\Big|_{\tau=0} = 0, \qquad i = \overline{0; N};$$
$$c_1^{(0)}(\xi, \tau)\Big|_{\xi=0} = \alpha c_0, \qquad c_2^{(0)}(\xi, \tau)\Big|_{\xi=0} = (1-\alpha) c_0;$$

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

$$c_1^{(i)}(\xi,\tau)\Big|_{\xi=0} = c_2^{(i)}(\xi,\tau)\Big|_{\xi=0} = 0, \qquad i = \overline{1;N}.$$

Here α $(0 \leqslant \alpha \leqslant 1)$ is a parameter determining the admixture part coming from the ground surface into the quick diffusion way (state 1), i.e. $d_1^{(0)} > d_2^{(0)}$.

Scheme for modulus of software. Solutions of the initial-boundary value problems of cascade type are constructed by iterative procedure using Green functions. On the basis of the obtained formulae we design software for simulation of diffusion processes in two ways with cascade decay of migrating particles. Scheme for the algorithm of software modulus for calculation of decaying particle concentrations for all $i = \overline{0;N}$ is shown in fig. 1.

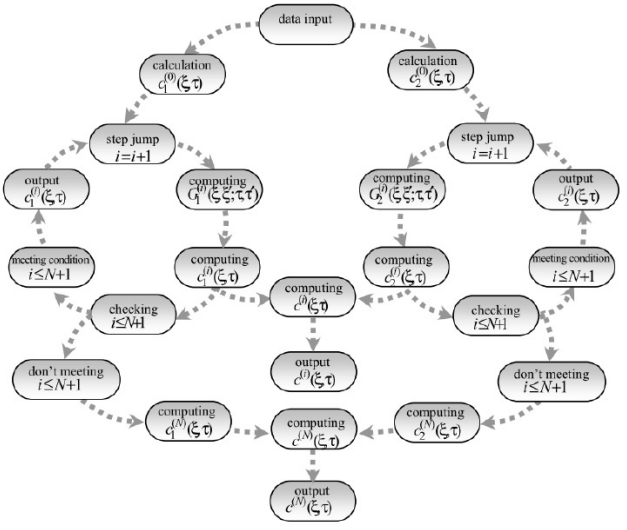

Figure 1. Scheme for the algorithm of software modulus for calculation of decaying particle concentrations within the model of non-interactive flows

In fig. 2 it is shown distributions of total concentration $c^{(1)}(\xi,\tau)/c_0$ on the step 1 of decay depending on the values of parameter $\alpha = 0{,}1;\ 0{,}3;\ 0{,}5;\ 0{,}7;\ 0{,}9$ (curves 1-5) for $a_\lambda^{(0)} = 0{,}8$; $d_1^{(1)} = 10$, $d_2^{(1)} = 0{,}125$, $a_1^{(1)} = a_2^{(1)} = 0{,}2$.

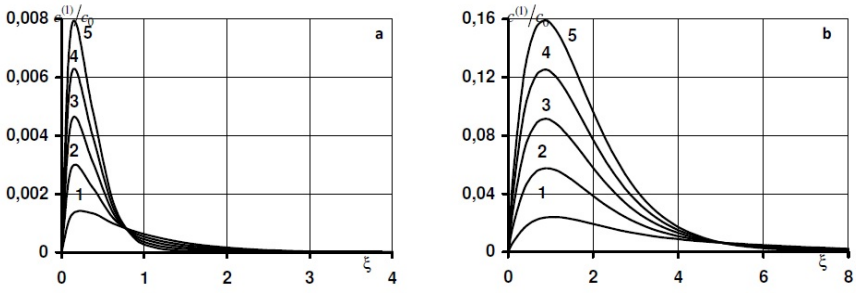

Figure 2. Concentration distributions $c^{(1)}/c_0$ for different values α in moments $\tau = 0{,}125$ (fig.a) and $\tau = 10$ (fig.b)

Note that behaviour of the concentration functions in both quick $c_1^{(1)}(\xi,\tau)$ and slow $c_2^{(1)}(\xi,\tau)$ migration ways in the stage $i = 1$ deffers substantial accumulation of substance $\mathbf{K}^{(1)}$ in the vicinity of the surface $\xi = 0$ in the slow way $j = 2$ and essential spread of this substance deep the layer by quick way $j = 1$. And both concentrations $c_1^{(1)}(\xi,\tau)$ and $c_2^{(1)}(\xi,\tau)$ are meaningful for the total concentration $c^{(1)}(\xi,\tau)$.

Chaplya Ye.Ya.[1,2] ***Chernukha O.Yu.***[1] ***Davydok A.Ye.***[1,3] ***Pabyrivskyi V.V.***[4]
[1]*Centre of Mathematical Modelling of Y.S.Pidstryhach Institute of Applied Problems of Mechanics and Mathematics of NAS of Ukraine, Lviv, Ukraine;* [2]*Institute of Mechanics and Applied Informatics of Kazimierz Wielki University in Bydgoszcz, Bydgoszcz, Poland;* [3]*Ya.S.Pidstryhach Institute of Applied Problems of Mechanics and Mathematics of NAS of Ukraine, Lviv, Ukraine;* [4]*National University "Lviv polytechnic", Lviv, Ukraine*

Diffusion equation for mass flux and problem of initial and boundary conditions

On modeling admixture mass fluxes in multiphase bodies of randomly nonhomogeneous structure there are significant difficulties during averaging over an ensemble of phase configurations because the correlation functions between gradient of stochastical field of the concentration and random diffusion coefficient are unknown. Therefore the original approach is proposed, according to which we construct the diffusion equation for the function of mass flux and formulate initial-boundary value problems directly to the flux [1]. However, within the scope of such approach it is necessary to formulate reasonably initial and boundary conditions, as in the case where values of the flux on the "top" body boundary are much greater than on the "bottom" one, unlimited amount of diffusing substance can enter into a limited body that is certain contradiction. Similarly, while maintaining a much larger flux through the "bottom" layer boundary we also come to certain collisions. In this regard, we propose to set the value of mass flux on one body surface, the value of substance concentration on another and further to determine the corresponding condition for the flux.

Let the process of admixture diffusion occurs in a medium. In the general case the equation of mass balance has the form

$$\frac{\partial c(\vec{r},t)}{\partial t} = -\vec{\nabla}\cdot\vec{J}(\vec{r},t), \tag{1}$$

where $c(\vec{r},t)$ is the concentration of admixture particles, $\vec{J}(\vec{r},t)$ is the mass flux of the diffusing substance, $\vec{\nabla}$ is Hamilton nabla-operator, $\vec{r}$ is a radius-vector of running point, t is time, operation of scalar multiplication is marked by a point.

Let us act left on equation (1) by the operator $(-\vec{\nabla})$ and multiply the result on the diffusion coefficient $D(\vec{r})$, which is accepted time-independent. Then we have

$$\frac{\partial}{\partial t}\left(-D(\vec{r})\vec{\nabla}c(\vec{r},t)\right) = D(\vec{r})\vec{\nabla}\otimes\vec{\nabla}\cdot\vec{J}(\vec{r},t). \tag{2}$$

Here $\ll\otimes\gg$ is the tensor multiplication and besides $\vec{\nabla}\otimes\vec{\nabla} = \nabla_i\nabla_j\,\vec{i}^{\,i}\otimes\vec{i}^{\,j}$ $(i,j=\overline{1,3})$, where ∇_i is symbol of the partial derivative, $\vec{i}^{\,i}$ is the base vector (in the case of Cartesian coordinate system $\nabla_1 = \partial/\partial x$, $\nabla_2 = \partial/\partial y$, $\nabla_3 = \partial/\partial z$; $\vec{i}^{\,1} = \vec{i}$, $\vec{i}^{\,2} = \vec{j}$, $\vec{i}^{\,3} = \vec{k}$).

Taking into account the first Fick low equation (2) is written as

$$\frac{\partial\vec{J}(\vec{r},t)}{\partial t} = D(\vec{r})\vec{\nabla}\otimes\vec{\nabla}\cdot\vec{J}(\vec{r},t). \tag{3}$$

So we obtain the equation of admixture diffusion represented by mass fluxes. In particular, in one-dimensional case equation (3) is reduced to the equation, which describes the flow function in a stochastically nonhomogeneous stratified medium

$$\frac{\partial J(z,t)}{\partial t} = D(z)\frac{\partial^2 J(z,t)}{\partial z^2} \tag{4}$$

Consider the process of admixture substance diffusion in a layer of thickness z_0 that contains sublayers, which location in area of the body, in general speaking, is unknown (fig. 1).

Accept that initial and boundary conditions of the first kind are satisfied for the flow function $J(z,t)$. In the initial moment of time there is no diffusion flow in the body. Admixture flow on the "upper" surface of the layer $z=0$ is constant, and particle concentration equals zero on the "lower" boundary of the strip $z=z_0$, namely

$$J(z,t)\big|_{t=0} = 0; \tag{5}$$

$$J(z,t)\big|_{z=0} = J_* \equiv const, \quad c(z,t)\big|_{z=z_0} = 0. \tag{6}$$

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

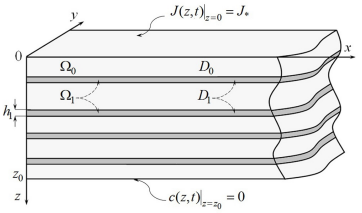

Figure 1. Possible realization of a multi-layered randomly nonhomogemeous strip

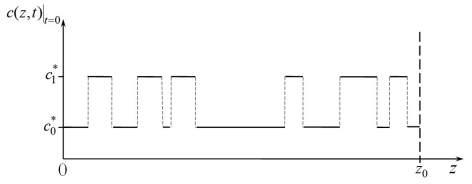

Figure 2. Schematic distribution of initial concentration in a two-phase stratified body

In this case diffusion flow on the "lower" boundary is a function of time $F(t)$ we need to define additionally from the corresponding problem for the concentration of migrating substance

$$J(z,t)\big|_{z=z_0} = F(t). \tag{7}$$

Let us set the initial condition for the admixture particle concentration, equivalent to the initial condition for flow of the substance (5). Taking into account the first Fick low for the chemical potential $\mu(z,t)$ we represent the condition on the flow (5) as

$$\mu(z,t)\big|_{t=0} = \mu_* \equiv const. \tag{8}$$

It follows from enough physical generals, that relation between chemical potential and concentration has logarithmic character

$$\mu(z,t) = \mu^0 + A\ln\gamma(z)c(z,t), \tag{9}$$

where μ^0 is the chemical potential for pure substance in the state determined by the values of absolute temperature T and pressure P; $A = RT/M$ is the coefficient, where R is the absolute gas constant, M is the atomic weight of admixture particles; $\gamma(z)$ is the activity coefficient, which can be presented for a two-phase body as $\gamma(z) = \begin{cases} \gamma_0,\ z \in \Omega_0; \\ \gamma_1,\ z \in \Omega_1, \end{cases}$ here Ω_j is the domain of phase j $(j = 0;1)$, $\sum \Omega_j = \Omega$, where Ω is the domain of whole body.

If we linearize the relation (9), find the expression for the function of admixture concentration and use the relation (8) and representation of the coefficient $\gamma(z)$ we obtain

$$c(z,t)\big|_{t=0} = \begin{cases} [1 + (\mu_* - \mu^0)/A]/\gamma_0, & z \in \Omega_0; \\ [1 + (\mu_* - \mu^0)/A]/\gamma_1, & z \in \Omega_1. \end{cases} \tag{10}$$

Denote $c_j^* = [1 + (\mu_* - \mu^0)/A)]/\gamma_j$ $(j = 0;1)$. Then $c(z,t)\big|_{t=0} = \{c_j^* \equiv const,\ z \in \Omega_j\}$. So we have obtained the piecewise-constant function of the initial concentration, schematic drawing of which is shown in fig. 2.

If the activity coefficients are close in different phases, i.e. we can assume that $\gamma_0 \approx \gamma_1 \equiv \gamma_*$, then $c_0^* \approx c_1^* \equiv c_*$ and the condition (10) is as follows

$$c(z,t)\big|_{t=0} = c_* \equiv const. \tag{11}$$

As a rule, the particular case of the initial condition on the function of admixture concentration in the form (11) is singled out, when admixture substance is absent in the initial moment of time

$$c(z,t)\big|_{t=0} = 0. \tag{12}$$

Remark that in the case of diffusion in stochastically nonhomogeneous bodies the initial condition (10) is random, and to construct solutions of initial-boundary value problems, in which the coefficients of differential equations and initial and boundary conditions are stochastic, it requires to develop the individual theory of mathematical physics using the theory of random fields.

Diffusion equations (3) or (4) accompanying by initial condition (5) in accordance with conditions (16) or (17) on concentration and boundary conditions (6), (7) are basis of formulation of initial-boundary value problems for fluxes, including processes in randomly nonhomogeneous structures.

References. **1.** Chaplya, Y., Chernukha, O., Davydok, A. (2012) Mathematical modeling diffusion flows in a randomly nonhomogeneous stratified strip. Reports of NAS of Ukraine, 11, 40-46.

Chereda H.O., Nikolaiev S.S.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Automatic annotation of digitized ECG signals with wavelets

Introduction. Modern people in the current times of fast pace of life don't have time to look after their health carefully. This problem also is applied to ECG diagnostic. The machine learning methods can be used in given portative diagnostic systems for automatic annotations and following analysis of ECG. In this paper features annotation techniques and parameter tuning are described.

Relevance. If people had a portative device which could analyze an ECG then there was no need to go to the doctors for an ECG recording and analysis. It would be enough to email the automatically analyzed data. In this case both doctor and patient would get the benefit. On the one hand the doctor could obtain more detailed ECG report, which includes ECG analysis during different human activities. It is very important because stress situations can show hidden heart pathologies. On the other hand - time of the patient would be saved and his health would be under control. This determines the relevance of the work.

The machine learning methods can automatically annotate and analyse ECG data. To apply these methods features sets should be extracted. One way of extracting features is to obtain an ECG with marked annotations. Term "marking" means indexes extraction of certain points of interest in time series of digitized ECG signals. A tool that marks ECG signals is called annotator.

Description of research. In monitoring device ECG analog signal from sensors goes through Analog-Digital-Converter (ADC). Obtained digitized signal has several parameters like sampling rate, bits per sample resolution and others - ADC gain, ADC baseline. Then this digitized signal is preprocessed and annotated. During this phase noise reduction techniques are applied to the signal and marking of P, T waves and QRS complexes [1] is performed. It is very important to make annotations on every heart beat and to know what annotation indexes corresponds to certain heart beat.

Two types of open source libraries were observed. The first one is Wave-Form-Database (WFDB) Toolbox. Tools of this library allow only marking the Q point in QRS complexes. But this library has powerful instruments for ECG signals visualization.

The second one is C++ ECG signal processing and annotation library [2]. The main math methods in this library provide filtering of the signal with continuous (CWT) and fast wavelet transforms (FWT).

$$T(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^*(\frac{t-\tau}{\sqrt{a}})dt$$

The CWT transform represents signal in time and frequency domains [3].

The consecutive application of CWT and FWT separates noise and P, T waves from the QRS complex. This provides better results compared to one stage filtering of the QRS complex with only FWT or CWT transforms. Results get even better if P and T waves are of high amplitude [2]. Different wavelets and wavelet filters were used to achieve better precision of the QRS complex location: Mexican hat, Gaussian derivatives, Daubechies, coiflets and biorthogonal. Inverse wavelet

$$\psi(x) = -xe^{-\frac{x^2}{2}}$$

and interpolation filters have provided better time localization of the QRS complex boundaries. CWT transforms at 3 Hz of the T wave provides inherent denoising.

First stage of detecting QRS complexes includes CWT transform with the Inverse wavelet at frequency 12 Hz. Detecting of P waves occurs with CWT which calculated with 9Hz. These frequencies are parameters that are defined and can be changed by user.

Taking into account the fact that all people have unique form of ECGs we may state question about parameter tuning to maximize annotations accuracy. For example some people are born with a prolonged end of T wave, but others have a prolonged QT interval induced by medicines [4]. All these factors make optimal parameters selection a quite challenging task.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Qualitative measurement occurs via annotations visualizer which was developed for these purposes. It allows checking of marked points on different parameters of annotator.

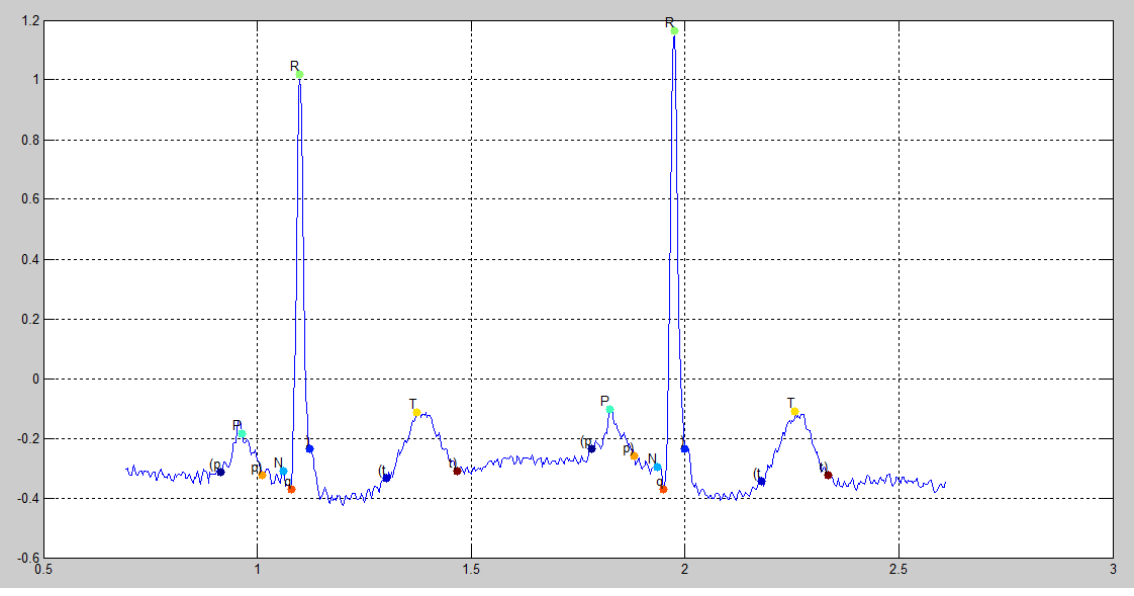

Figure 1. Annotations with CWT frequencies: P wave – 3 Hz, T wave – 2 Hz, QRS wave – 13 Hz

Quantitative measurement techniques based on quantity of passed annotated points allow to increase S points detection in QRS complex from 45% to 90% .

Conclusions. Importance of automatic ECG annotation quality improvement was explained. Wavelet digitalised ECG annotation principles and methods were observed. Quantitative measurement for Annotation evaluation was offered. The dependencies between frequencies of CWT and the quality of the annotations were experimentally established. Certain uniform changes of CWT frequencies for QRS complex and P, T waves detection implies monotonic accuracy increasing up to a certain limit.

References. **1.** Antoni Bayés de Luna, Velislav N. Batchvarov and Marek Malik, *The Morphology of cardiogram*, 2005, pp. 2-3.

2. Chesnokov Yuriy, Individually Adaptable Automatic QT Detector, *Computers in Cardiology* vol.4, 2006, pp. 337-341.

3. G.Kaiser, *A Friendly Guide to Wavelets*. New York:Birkhäuser, 1994, pp. 60-65.

4. Sala M. et al. Antidepressants: their effects on cardiac channels, QT prolongation and Torsades de Pointes in Curr *Opin Investig Drugs* vol.3, 2006, pp. 256-263.

Cojocaru S., Gaindric C., Secrieru Iu., Popcova O.
Institute of Mathematics and Computer Science of the Academy of Sciences of Moldova, Chişinău, Republic of Moldova

The SonaRes platform and its applications

SonaRes technological platform is designed to support diagnosis process using ultrasound examination methods. Diagnostics is one of the most difficult tasks clinicians may face every day. Determination of the correct diagnostics involves a multitude of factors, that characterize the patient's state of health. Omission or misinterpretation of them can lead to mistakes with unpredictable consequences. No less important is to reveal the relationship between these factors, as well as interaction between organs and its manifestation mode. In difficult cases, and not only, a second opinion, an advice, is useful to any physician, even to an experienced one. In such activities it is a natural tendency to ask for a help, to make sure we proceed in the right way or to find a solution when we do not know how to proceed.

The problem, associated with the medical diagnostics activities, acquires a special relevance in modern circumstances. First of all, it is connected with the fact that physicians have to work with poorly structured and weakly formalized information. Besides, the volume of information is in a continuous growth thanks to the appearance of new methods of medical examination.

The ultrasound investigation domain is not an exception. The appearance of new ultrasound devices or the improvement of old scanners doesn't simplify, but even complicates the physician's diagnostic thinking, because one has to analyze a number of diagnostic data, that increases the time of diagnostics determination and sometimes may reduce the accuracy.

The platform consists from four main components: expert shell, knowledge base, annotated images collection, images processing tools. Its scope is limited to hepatic-pancreatic-biliary zone.

Expert shell is designed to support effective communication with experts in the development of the knowledge base. The main stages of the development are: problem identification, knowledge acquisition, structure definition, formalization and implementation. The experience of experts and the medical specialty literature served as the main sources of knowledge.

The knowledge base contains information related to organ's localization, its pathologic states (for example: chronic cholecystitis, compressed gallbladder etc.), principal characteristics for organ description (for example: dimensions/volume, shape, tonicity, gallbladder contour etc.), description of organ's anomalies and pathologies.

SonaRes platform combines the two basic approaches of ultrasound diagnostic systems development by assisting the decision making process on the base of both rules and images. It contains a collection of model images attached to the corresponding rule, with ROIs (region of interest for a particular pathology) defined. The role of annotated images in decision making is to provide help by illustration, in cases when the medical specialist is not sure how to correctly interpret the images.

Basing on this platform a number of application were developed: SonaRes — a decision support system for ultrasound examination [1], Emergency SonaRes – a system for ultrasound examination in the case of mass casualty and disaster setting [2], Sonares.Edu – a system oriented to education process.

All these derivatives of SonaRes platform were evaluated both by experienced clinicians and novices. The applications demonstrated diagnosis accuracy at least similar to those formulated by specialists in free text and can be qualified as satisfactory for adequate diagnosis support.

References. **1.** L.Burtseva, S.Cojocaru, C.Gaindric, E.Jantuan, O.Popcova, I.Secrieru, D.Sologub, SonaRes - A decision support system in ultrasound investigations, Computer Science Journal of Moldova v.15 N2 (44), 2007, pp.153-177. **2.** C.Gaindric, S.Cojocaru, O.Popcova, S.Puiu, I.Secrieru. Emergency-SonaRes: A system for Ultrasound Diagnostics Support in Extreme Cases. In: Teodorescu,H.-N., Kirschenbaum,A., Cojocaru,S., Bruderlein,C.(Eds.) Improving Disaster Resilience and Mitigation - IT Means and Tools. Springer, 2014, XV, pp.283-292.

Ilchenko K.O., Pyshnograiev I.O.
National Technical University of Ukraine "Kyiv Polytechnic Institute", Kyiv, Ukraine

Knowledge Society Assessment as a Composite Index

Introduction. Classic economic theory presents three factors of production: land, labor and capital.This fundamental approach was divided into several complex theories that include additional factors of production, for example, technological progress, human capital and social capital. Those resources can be aggregated into one category – knowledge. According to modern science knowledge and information are the most important factors of production that form the basis for new technological mode.

This idea led to Knowledge society (K-society) concept where the mass production of knowledge changes the global economy. However, this process is dissimilar in different countries.

Taking into account the necessity of Knowledge society development estimations and complex character of knowledge, the hierarchy model can be used to describe this concept. Therefore, the main idea of the research is to draw out a composite indicator for measurement knowledge as a sophisticated category with a purpose of country development analysis.

Methodology. According to the UNO methodology, the index of K-society should be based on three dimensions: Assets, Advancement and Foresightedness [2]. The first one describes the level of education, especially, among young people, and the development of information streams. These two main directions include such indicators as: expected schooling, proportion of young people, the diffusion of newspapers, the Internet, main phone lines and cellular phones. The second dimension represents human and informational resources, which are indicated by public health expenditure, research and development expenditure, military expenditure, pupil/teacher ratios in primary education, and a proxy of the "freedom from corruption" indicator. The last dimension shows the external influence on K-society dynamics in the state. This dimension consists of low child mortality rates, equality in income distribution (GINI Index), protected areas as percentage of a country's surface, and CO2 emissions per capita indicators. This approach was officially accepted for approximately 45 countries in 2005.

Taking into account the existent basic specification of the main categories, it becomes possible to continue this research in terms of current informational mode. Thus, new hierarchical model for K-society measurement has been built with usage of UNO methodology and OECD approach for composite indicators [3]. As a result the index which aggregates 14 indicators was drown out. The changes in the number of indicators and using other categories were established in accordance to data limitations.

Results. The K-society Index was calculated for 87 countries that provide all necessary information to international organizations.

The results for 2013 year show that the Top 10 countries in terms of K-society consists of Switzerland, Denmark, Netherlands, Sweden, Slovenia, France, Austria, New Zealand, Japan and Finland.

The analysis of results shows that there is no direct dependence between K-society development and the country leadership in the world.

The situation for Ukraine was analyzed deeply. Firstly, Ukraine has lower meaning of index than it's neighbor countries Moldova, Poland, Hungary. Secondly, the less developed dimension is "Advancement". Thus, this strategy of country's development can be based on approaches described above.

References. **1.** UNESCO World Report: Towards Knowledge Societies. UNESCO, France (2005). **2.** Understanding Knowledge Societies. Department of Economic and Social Affairs, United Nations, New York (2005). **3.** Handbook on Constructing Composite Indicators: Methodology and User Guide. Organisation for Economic Co-Operation and Development, France (2008).

***Ivanchuk M.A.*[1], *Malyk I.V.*[2]**
[1]*Bukovinian State Medical University, Chernivtsi, Ukraine;* [2]*Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine*

Linear separation using ε-nets

ε-separability conception is defined in this paper. Necessary and sufficient condition of ε-separability is formulated. The problem of two sets ε-separation is reduced to the trivial problem of their non-intersecting ε-nets separating.

Consider two sets of points $A \subset R^d$ and $B \subset R^d$, $|A| = n_A$, $|B| = n_B$. Let $A \not\subset conv_B$, $B \not\subset conv_A$, but the sets are not separable: $conv(A) \bigcap conv(B) \neq \emptyset$. We will solve the task of separation of the sets A and B after rejecting the small quantity of the sets' points, for example $\varepsilon \in (0, 1)$ percent of the sets' size.

To solve this task we will use the ϵ-nets conception [1, 2].

Definition 1. The pair (X, R) is called a ***range space***, with X its points and the elements of R its ranges.

Definition 2. Let (X, R) be a range space. Given $A \subseteq X$ finite and $\varepsilon \in (0; 1)$. A subset N of A is called an ***ϵ-net*** of A if $\forall r \in \mathrm{R}$, $|r \bigcap A| \geqslant \varepsilon |A| \Rightarrow N \bigcap (r \bigcap A) \neq \emptyset$.

Definition 3. Two sets of points A and B are called ***ϵ-separability***, if there exist sets $A_1 \subset A$, $B_1 \subset B$, such that $conv(A \backslash A_1) \bigcap conv(B \backslash B_1) = \emptyset$ and $|A_1| + |B_1| < \varepsilon(n_A + n_B)$.

Definition 4. Hyperplane L is called ***separating*** for the sets A and B if $conv_A \subset L^+$, $conv_B \subset L^-$.

Definition 5. Hyperplane L_ε is called ***ϵ-separating*** for the sets A and B if $\frac{|A \bigcap L_\varepsilon^+| + |B \bigcap L_\varepsilon^-|}{n_A + n_B} \geqslant 1 - \varepsilon$.

Let (R^d, H^d) is an infinite range space with H^d the closed halfspaces in R^d bounded by hyperplanes.

Theorem 1. A necessary and sufficient condition that two sets of points A and B are ϵ-separability is there exist $\varepsilon_A, \varepsilon_B$ and corresponding ϵ-nets $N_{\varepsilon_A}^A$, $N_{\varepsilon_B}^B$ in (R^d, H^d) such that

$$\varepsilon_A n_A + \varepsilon_B n_B < \varepsilon (n_A + n_B)$$

and

$$conv N_{\varepsilon_A}^A \bigcap conv N_{\varepsilon_B}^B = \emptyset.$$

Let's determine

$$\eta_A = \frac{|A \bigcap conv_B|}{n_A}, \eta_B = \frac{|B \bigcap conv_A|}{n_B}$$

and $\varepsilon_A = \eta_A + \frac{1}{n_A}$, $\varepsilon_B = \eta_B + \frac{1}{n_B}$

Theorem 2. Let

$$\varepsilon > \frac{\eta_A n_A + \eta_B n_B}{n_A + n_B},$$

then for any ϵ-separating hyperplane L_ε for the sets of points A and B there exist ϵ-nets $N_{\varepsilon_A}^A$ and $N_{\varepsilon_B}^B$ in (R^d, H^d) such that L_ε is separating hyperplane for these ϵ-nets.

So, to solve the task of ϵ-separation of two sets A and B by hyperplane in R^d, it's enough to separate there ϵ-nets $N_{\varepsilon_A}^A$ and $N_{\varepsilon_B}^B$ in (R^d, H^d).

References. **1.** D. Haussler and Emo Welzl, Epsilon-nets and simplex range queries. Discrete Comput. Geom., 2, (1987), 127–151. **2.** Vapnik V.N. On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities /V. N. Vapnik and A. Ya. Chervonenkis // Theory Probab. Appl. – 1968. - 16(2).- p.p. 264–280.

Koval O.V., Volkova K.A.
National Technical University of Ukraine "Kyiv Polytechnic Institute", Kyiv, Ukraine

Computer modelling of management information systems with Actionable analytics

The effectiveness of strategic and tactical decision-making depends on the quality of information. Fixed rules and prepared strategy inferior to more informative and flexible solution using accurate information provided in the required time, or the system customer relationship management, or enterprise resource planning system or other *management information systems (MIS)*. Procurement of operational data in context of domain is one of *actionable analytics* elements in management information systems [1]. Actionable analytics is one of information technology trends for past years.

Actionable analytics is characterized by:

- Delivered to users at the point of action and in context.
- Performed in real-time for any business action.
- Optimization and simulation are enabled everywhere and in any time throughout big data repositories and cloud technologies.
- Data needs to be available at the right time and right place to be actionable.

Model development approach is suggested to development of MIS with actionable analytics. Model development approach can be used for assessment of system complexity through analyze of its computer model on completeness, integrity and correctness. The base principle of model development is system reflection as assembly of models.

Two software development methodologies use models:

1. Model-driven development;
2. Model-based development.

Both methodologies are used for design and development of physical objects and program systems. The main difference is in model-code transformation. In model-driven development model-code transformation performed automatically instead of model-based development, where model are used only in computer modeling stage. Two main requirements for software development are [2]:

1. Each model need to be formalized.
2. Each model need to interact with previous models.

In accordance to MDD MIS with actionable analytics are considered as sequence of models:

$$M^{MIS} : \left\{ Kn^M, D_i^M, Pr_i^M, Sc_i^M, C_i^M, Dv^M \right\} \tag{1}$$

, where Kn^M - knowledge model, that define two tipes of knowledge: technological knowledge (services, information technologies, program tools, DataBase, Big Data) and organizational knowledge (scenario executors, business rules);

D_i^M - domain object model , that contain description of entities and their meta-data;

Pr_i^M - use case model, which is a declarative description of the functional requirements of the system logic and user interaction with the GUI;

Sc_i^M - scenario model of analysis task, formalized the process of analyzing the information and is expressed in a logically-ordered sequence of events in the process of analyzing the data in order to obtain the necessary information to make informed decisions;

C_i^M - component model as a set of components necessary for functional and other requirements, which built and operates in accordance to certain rules and component interactions;

Dv^M - implementation component model, which also includes a description of information exchange objects and system elements.

References. **1.** Magic Quadrant for Business Intelligence and Analytics Platforms. — 2013, © Gartner Analytic. **2.** http://www.theenterprisearchitect.eu/blog/2009/06/03/a-framework-for-model-driven-soa/.

Makukha M.P., Oparina E.L.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Situational analytical center for monitoring, modeling, and mitigation of social disasters caused by catastrophes and terrorism

Natural disasters, technogenic catastrophes, and acts of terrorism can cause various *direct damage*: diseases and loss of life; food and housing shortage; infrastructure destruction; land, water and air pollution. In addition to direct humanitarian, environmental and economical damage, negative social consequences can arise. Social consequences can be more destructive, than the direct damage itself, and in some circumstances can lead to a *social disaster*.

Social disasters are usually characterized by processes and indicators, associated with the *indirect damage* caused by disasters, catastrophes and acts of terrorism: marauding, criminality growth, poverty, unemployment, narcomania, suicides, civil disorders, migration processes, etc. The situation is aggravated by the self-sustaining and mutually amplifying nature of negative social processes mentioned above.

Mitigation of social disasters require the system approach [1] and can have a significant positive impact on both humanitarian, social, psychosocial, demographical, ecnonmical and political development for the affected region. Monitoring and modeling provide necessary basis for both informed decision making and proactive disaster management measures.

Monitoring of catastrophes and social disasters indicators should be executed continuously and may be based on various regional and global information sources: social media (especially microblogging), news websites, discussion boards, ecommerce and other specific resources, statistical data sources. Data fusion from both structured and unstructured data sources provide the necessary level of information accuracy and completeness.

Modeling of social disasters require both quantitative and qualitative methods. Quantitative methods include descriptive and predictive modeling; qualitative methods include scenario analysis and foresight methodology [2,3].

Mitigation of social disasters may be based on qualitative methods (foresight and scenario analysis) accompanied with numeric optimization to determine parameters values where possible.

All three stages mentioned above require both powerful computational resources and proficient researchers and data scientists. The necessary level of integration of researchers and computational resources require development of a dedicated situational analytical center.

The Situational Analytical Center for monitoring, modeling, and mitigation of social disasters caused by catastrophes and terrorism is developed in the Institute for Applied System Analysis of NTUU "KPI" supported by the NATO Science for Peace and Security grant. The center should inculde several subsystems:

- Monitoring and data gathering subsystem.
- Analytical systems for both quantitative and qualitative methods.
- Data integration platform.
- Visualization subsystem.
- Communication subsystem for real-time collaboration.
- Human resources: researchers, data scientists, public relations specialists, management.
- Infrastructure: location, network infrastructure, video surveillance, intrusion protection, equipment for center mobility, etc.

References. **1.** Zgurovsky M.Z., Pankratova N.D. System analysis: Theory and Applications // Springer, 2007. 475 p. ISBN 978-3-540-48880-4. **2.** Pankratova N.D, Savchenko I.A Morphological analysis. Problems, theory, application // Naukovadumka, Kyiv, 2015. 347 p. **3.** Pankratova N.D., Malafeeva L.Y. A system approach to solving foresight problems on the basis of Delphi method // International Journal "Information technologies & knowledge". ITHEA, Sofia, 2013. V.7, №4. P.313–324.

This research was sponsored in part by the NATO Science for Peace and Security Programme under grant G4877.

Paliichuk L.S.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

On the limit cycles for controlled piezoelectric fields

Nowadays a lot of attention is paid to the accuracy of mathematical modeling and flexibility of control for contact piezoelectric processes and fields due to the accelerated development of equipment. For possible adequately modeling the actual physical effects and phenomena, taking into account the different kinds of impact on the investigated system, we need to weaken the conditions on the parameters of such models (in particular, to refuse the smoothness, continuity etc.). Since for contact piezoelectric processes the non-linearity is inherent (that can be discontinuous, multivalued), in such case, only the numerical analysis is not enough to adequately modeling the dynamics of the studied phenomenon, and sometimes it is not possible at all. So, it is necessary to use the methods of multivalued analysis (developed by Zgurovsky M.Z., Melnik V.S. Kasyanov P.O., Kapustyan O.V. etc. [1], [2], [3]). Thus the importance of qualitative analysis for mathematical models of piezoelectric contact processes and fields that submit to the laws with the nonlinear, discontinuous and multivalued interaction functions is obvious.

We consider the problem with multivalued "reaction-displacement" law:

$$u_{tt}(t) + Au_t(t) + Bu(t) + \partial J_1(u(t)) - \partial J_2(u(t)) \ni \overline{0} \text{ for a.a. } t > 0,$$

where $A : H \to H$, $B : V \to V^*$ are the linear symmetric operators, $J_i : H \in R$, $i = 1, 2$, are locally Lipschitz functionals, and ∂J_i are the Clarke subdifferentials for J_1, $i = 1, 2$. $(V; H; V^*)$ is an evolution triple. $X = V \times H$ is the phase space of investigated problem.

Taking into account some assumptions on the parameters of the studied problem [4], by using the principles and methods of the theory of global and trajectory attractors for multivalued semiflows it is established that the multivalued semiflow G generated by all weak solutions of considered problem satisfies the flattening property [5], i.e.:

for any bounded set $B \subset X$ and $\varepsilon > 0$ there exist moment of time $t_0(B, \varepsilon)$ and finitedimensional subspace E in X such that for some bounded proector $P : X \to E$ set $P(\bigcup_{t \geqslant t_0} G(t, B))$ is bounded in X and the next embedding holds:

$$(I - P)(\bigcup_{t \geqslant t_0} G(t, B)) \subset B_\varepsilon(\overline{0}),$$

where I is the identical mapping on X. This means that the dynamics of the investigated system is finitedimension within a small parameter.

Therefore, we can direct the solutions of the studied problem to the set of stationary states. Results of this investigation may be used for development of technical equipment based on the piezoeffect, for example for design of piezoelectric positioning controller etc.

Thesis based on the results of joint work with the Doctor of Physics and Mathematics P.O. Kasyanov.

References. **1.** Zgurovsky M.Z. Evolution Inclusions and Variation Inequalities for Earth Data Processing I. / Zgurovsky M.Z., Melnik V.S., Kasyanov P.O. – Berlin: Springer-Verlag, 2011. – 247 p. **2.** Zgurovsky M.Z. Evolution Inclusions and Variation Inequalities for Earth Data Processing II. / Zgurovsky M.Z., Melnik V.S., Kasyanov P.O. – Berlin: Springer-Verlag, 2011. – 274 p. **3.** Evolution Inclusions and Variation Inequalities for Earth Data Processing III. / Zgurovsky M.Z. and all. – Berlin: Springer-Verlag, 2012. – 330 p. **4.** Kasyanov P.O. Method of multivalued operator semigroup to investigate the long-term forecasts for controlled piezoelectric fields / P.O. Kasyanov, L.S. Paliichuk, A.M. Tkachuk // Chebyshevski Sbornik. — 2014. — Vol. 15, Iss. 2. — P. 21–32. **5.** Kalita P. Global attractors for multivalued semiflows with weak continuity properties / P. Kalita, G. Łukaszewicz // Nonlinear Analysis. — 2014. -– Vol. 101. — P. 124–143.

The investigation was partially supported by Grant of NAS of Ukraine 2273/14.

Pankratova N.D., Pyvovar V.V.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Big Data Overview

Introduction. There are some big things which have implications for everyone whether we want them to or not. They could be so very big. Big Data is one of those things, and is completely transforming the approach we do business and is influencing most other parts of our lives [1].

What is Big Data?. "Big Data" means different concepts to different people and there is not, and possibly never will be, a commonly agreed upon definition. But the phenomenon is not a fantasy and it is producing advantages in so many different areas, so it makes sense for all of us to have a working full understanding of the idea [1].

Big Data trends. Based on requests from clients - vendors of data processing platforms and products - as well as trends in popular blogs, job postings [2]. Trends are in: use data science in unusual fields such as astrophysics; rise of data plumbing; putting AI or deep learning into rudimentary big data applications for purpose of detecting and eliminating spam, fakes etc; increasing awareness against hackers; rise of mobile data exploitation in capturing and reacting to the right signal of a disease or other global risks; optimization of revenue from mobile apps; rise of automated, scalable, robust analytic solutions fit for black-box analytics used by non-experts; mathematical optimization and simulation become as an alternative to delicate and mysterious statistical models; rise of forecasting space weather, volcanoes, earthquakes; forecasting consequences of disasters; automated content generation, correction of student essays and plagiarism detection; digital health.

Open sources. There are both big and little data set resources have been published. Some of them associated with our data science interests [3]. Areas could be tagged with Agriculture, Biology, Climate/Weather, Complex Networks, Computer Networks, Data Challenges, Disaster statistics, Economics, Energy, Finance, GeoSpace/GIS, Government, Healthcare, Image Processing, Machine Learning, Museums, Blogger Corpus, Physics, Public Domains, Search Engines, Social Sciences, Social disasters, Sports, Time Series, Transportation. The majority such data sources are from Government, GeoSpace/GIS, Social sciences and Disaster statistics, which means mentioned data are most open for mining. Last one is especially good, because in our research we are interested in social disasters data.

Recent technologies. Regarding big data researches it is not possible not to mention blockchain-like architectures which rised very recent. One of most interesting is ethereum, a platform for decentralized applications [4]. "There will be no "killer app" for blockchain technology" [5], - said 19-years creator of innovation. Another example could be free search engine YaCy which can be use to build a search portal for their intranet or to help search the public internet. It is fully decentralized and it is not possible for anyone to censor the content of the shared index [6].

Conclusions. The biggest cause big data is important to everybody is that it is a trend that is only going to grow. Modern world will not let you escape from this. It was mentioned one of most important trend Big Data helps us to forecast consequences of disasters, therefore mitigate effects. With open data sources of history Big Data can help to manage effects on future of many people.

References. **1.** Jenn Webb and Tim O'Brien, *Big Data Now.* by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472., 2013 **2.** Vincent Granville, www.datasciencecentral.com/profiles/blogs/13-new-trends-in-big-data-and-data-science. **3.** Mirko Krivanek, www.datasciencecentral.com/profiles/blogs/great-github-list-of-public-data-sets. **4.** Website of project, www.ethereum.org. **5.** Vitalik Buterin, blog.ethereum.org. **6.** Website of project, yacy.net.

This research was sponsored by the NATO Science for Peace and Security Programme under grant SPS G4877

Parkhomchuk D.M., Tymoshenko Yu.O.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Analysis of second order dynamic method

Introduction. Inverse problems with approximation of continual differential and integral equations by finite discrete algebraic systems or local linearization of non-linear systems reduced to the solution of system of linear equations

$$Ax = b, A \in R^{m\times n}, x \in R^n, b \in R^m, \tag{1}$$

where matrix A is usually singular, so such systems usually do not have mathematically rigorous solutions, and it is only possible to find approximate solutions [1].

Dynamic method. This paper analyzes the dynamic method for solving ill-posed problems, according to which, the solution of the following Cauchy problem is accepted as an approximate solution of the system 1 [2]:

$$\xi_1 * x'' + \xi_2 * x' + A^T Ax = A^T b, x(0) = x_0, x'(0) = x_1 \tag{2}$$

In order to represent the solution of the problem 2 we will introduce the following notation: a_i - i-th column of matrix $A^T A$, $\delta(a_1, a_2, ..., a_n)$ - determinant of matrix, which consists of columns $a_1, a_2, ..., a_n$, λ_i - i-th eigenvalue of matrix $A^T A$, $c_i = \frac{-\xi_2+\sqrt{{\xi_2}^2-4\lambda_i\xi_1}}{2\xi_1}$, $d_i = \frac{-\xi_2-\sqrt{{\xi_2}^2-4\lambda_i\xi_1}}{2\xi_1}$,

$\triangle_i^j = \delta(a_1, ..., a_{i-1}, A^T b, a_{i+1}, ..., a_{j-1}, e_j, a_{j+1}, ..., a_n)$,

$\nabla_i^j = \delta(a_1, ..., a_{i-1}, \xi_1 x_1 + \xi_2 x_0, a_{i+1}, ..., a_{j-1}, e_j, a_{j+1}, ..., a_n)$,

$\aleph_i^j = \delta(a_1, ..., a_{i-1}, \xi_1 x_1 + \xi_2 x_0, a_{i+1}, ..., a_{j-1}, e_j, a_{j+1}, ..., a_n)$, $\triangle_i = \delta(a_1, ..., a_{i-1}, A^T b, a_{i+1}, ..., a_n)$,

$F_1^i(p) = \triangle_i + (\xi_1 p^2 + \xi_2 p) \sum_{j=1, j\neq i}^n \triangle_i^j + (\xi_1 p^2 + \xi_2 p)^2 \sum_{1<j_1<j_2<n}^{j_1, j_2 \neq i} \triangle_i^{j_1, j_2} + ... + p(\xi_1 p^2 + \xi_2 p)^{n-1} A^T b^i$

$F_2(p) = p \prod_{i=1}^n (p - c_i)(p - d_i)$,

$G_1^i(p) = \nabla_i + (\xi_1 p^2 + \xi_2 p) \sum_{j=1, j\neq i}^n \nabla_i^j + (\xi_1 p^2 + \xi_2 p)^2 \sum_{1<j_1<j_2<n}^{j_1, j_2 \neq i} \nabla_i^{j_1, j_2} + ... + p(\xi_1 p^2 + \xi_2 p)^{n-1} x_0^i$

$G_2(p) = \prod_{i=1}^n (p - c_i)(p - d_i)$,

$H_1^i(p) = p\xi_1(\nabla_i + (\xi_1 p^2 + \xi_2 p) \sum_{j=1, j\neq i}^n \nabla_i^j + (\xi_1 p^2 + \xi_2 p)^2 \sum_{1<j_1<j_2<n}^{j_1, j_2 \neq i} \nabla_i^{j_1, j_2} + ... + p(\xi_1 p^2 + \xi_2 p)^{n-1} x_0^i)$

so i - th coordinate of solution 2 will be:

$x_i(t) = \frac{\triangle_i}{\triangle} + \sum_{j=1}^n (e^{c_j t}(\frac{F_1(a_j)}{F_2'(a_j)} + \frac{G_1(a_j)+H_1(a_j)}{G_2'(a_j)}) + e^{d_j t}(\frac{F_1(a_j)}{F_2'(b_j)} + \frac{G_1(a_j)+H_1(a_j)*\frac{b_j}{a_j}}{G_2'(b_j)}))$

These solutions have the following properties:

1. If ξ_1, ξ_2 is greater than 0 the solutions are stable.
2. In case of determinacy solutions converge to the normal solutions according to the initial conditions.
3. There is a method for selection of the stop criteria, in which the method of DM 2 becomes regularizable.

Using these properties we can offer an algorithm for solving inverse problems:

1. Transform the inverse problem to the problem 1
2. Identify the largest eigenvalue λ_m of the matrix $A^T A$
3. Select values ξ_1, ξ_2 according to $\xi_2 > 2 * \xi_1 * \lambda_m$
4. Define the functional form $H(\alpha, \beta, x(\theta^*))$
5. Determine the level for stop e_d
6. Start the numerical solution of system 2, continue solving while $e_d > H(\alpha, \beta, x(\theta^*))$
7. Accept $x = x(\theta^*)$ as the solution of 1

References. **1.** Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений: Пер. с англ. Икрамова Х.Д. -М. : Мир, 1980. -277 с. **2.** Гутенмахер Л. И.,Тимошенко Ю. А,Тихончук С. Т. О динамическом методе решения некорректных задач: Докл. АН СССР. — 1977. — Т. 237. — № 4. — С. 776—778.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

***Petic M.*,[1,2] *Cojocaru S.*[1]**
[1]*Institute of Mathematics and Computer Science of the Academy of Sciences of Moldova, Chişinău, Republic of Moldova;* [2]*Alecu Russo Balti State University, Bălţi, Republic of Moldova*

Vocabulary enriching for text analysis

To develop various applications in the domain of natural language processing is not enough to use vocabularies which contain only lemma words, usually there is necessary to have all their inflected and derived forms. An example of this type are applications designed for sentiment analysis, which are operating not only with the stand alone dictionary entry, but also with a lot of the corresponding words derivatives. One of the objectives of the project "Modeling and Mitigation of Social Disasters Caused by Catastrophes and Terrorism", developed jointly by teams from Ukraine, Romania and Republic of Moldova, is to analyze information related to social disasters in order to identify sentiments containing in corresponding texts.

Ukrainian, Russian and Romanian languages belong to the class of highly inflexional ones. The tools we use for text monitoring and analysis operate with vocabularies which must contain lemma-words and all their words-forms. Therefore we elaborated a tool to inflect and derivate the selected keywords and applied it for Romanian language. The idea is based on the developing of special grammar formalising word-forms production.

Let us consider a scattered context grammar rule:

$$[/] * [N_1]a_1 \neg b_1 a_2 ... a_{n-1} \neg b_{n-1} a_n \to a\prime_1 \neg b_1 a\prime_2 ... a\prime_{n-1} \neg b_{n-1} a\prime_n [N_2]$$

where a_i, $a\prime_i$ are arbitrary words and either b_i is a nonempty word or the special symbol * stands instead of b_i. N_j are the endings set numbers. The interpretation of this rule is as follows.

Let w be a lemma. Every sign / indicates cutting the last letter from w. Word v (obtained after the deletions) is considered as a root (if N_1 is not empty) and N_1 is its index in the endings set list L. In any case the word v should have the form

$$f_0 a_1 f_1 a_2 f_2 ... a_{n-1} f_{n-1} a_n f_n,$$

where every f_i is an arbitrary (possible empty) word, not containing (for $i = 1, 2, ..., n-1$) the veto subword b_i. If there exists more than one representation of this kind, the first (scanning v from the left to the right or vice versa if the sign ♯ is present) should be selected. The special character * instead of b_i admits arbitrary f_i.

In this context the parallel substitution

$$a_1, a_2, ..., a_n \to a\prime_1, a\prime_2, ..., a\prime_n,$$

is produced, generating a new root $v\prime = f\prime_0 a\prime_1 f\prime_1 a\prime_2 f\prime_2 ... a\prime_{n-1} f\prime_{n-1} a\prime_n f\prime_n$ and N_2 is its endings set number.

So, in order to generate word-forms it is sufficient to interpret the corresponding grammar rules [1].

The above method of inflexion is based on the knowledge about the morphological group of a given word. An algorithm for morphological group of an arbitrary word was developed [2].

The experiment showed that 866 grammar rules and 320 ending sets were sufficient to formalize the inflexion process of productive parts of speech for the Romanian language.

Obviously, for each ending we can establish a correspondence with morphological characteristics of the word-form, thus obtaining a possibility of morphological annotation.

The particularities of the derivational morphology mechanisms help in lexical resources extension without any semantic information. The existent approaches and mechanisms we studied on the examples from Romanian, but, in most of the of cases, they can be, more or less, applicable to other languages [3].

NATO.NUKR.SFPP 984877 "Modeling and Mitigation of Social Disasters Caused by Catastrophes and Terrorism".

In general case the process of derivation can be written in the form of set of rules of the form:

$$\alpha_1\beta_1\alpha_2\beta_2...\alpha_n\beta_n \rightarrow [p]\alpha_1\beta\prime_1\alpha_2\beta\prime_2...\alpha_n\beta\prime_n[s],$$

where $|\alpha_i| \geqslant 0$, $|\beta_i| \geqslant 0$, $|\beta\prime_i| \geqslant 0$ and in the process of alternation there are the following three possibilities: $|\beta_i| = |\beta\prime_i|$, $|\beta_i| > |\beta\prime_i|$, $|\beta_i| < |\beta\prime_i|$.

For example: dărima → dărimătură, the alternation a-ă, corresponds to the case $|\beta_i| = |\beta\prime_i|$, ceaţă → ceţos, the alternation ea-e, corresponds to the case $|\beta_i| > |\beta\prime_i|$, contagios → contagioasă, the alternation e-ea, corresponds to the case $|\beta_i| < |\beta\prime_i|$.

Studies on derivation process allow us to conclude that we cannot propose an effective algorithm for automatic derivation in general, but we can highlight some models of derivation, for which construction of such algorithms is possible. Below we will underline four the most important models of derivation.

Affixes substitution Let be x_1 a word of the form $x_1 = \alpha_1\omega$, where α_1 is a prefix. After the substitution $\alpha_1 \rightarrow \alpha_2$ we obtain the word $x_2 = \alpha_2\omega$, where x_2 is the obtained derivative, for example, *În*chide-*des*chide. The same for the derivatives $x_1 = \omega\beta_1$, with the suffixe β_1. After the substitution $\beta_1 \rightarrow \beta_2$ we obtain the word $x_2 = \omega\beta_2$, for example, corig*enţă*-corig*ent*.

Derivatives projection Let be ω a word, α - a prefix, β - a suffix, then the following relations are valuable:

$(\omega \rightarrow \alpha\omega)\bigwedge(\omega \rightarrow \omega\beta) \Rightarrow (\omega \rightarrow \alpha\omega\beta)$, for example, (a lucra → lucr(a)ă*tor*)$\bigwedge$ (a lucra → a *pre*lucra) ⇒ (a lucra → *pre*lucr(a)ă*tor*);

$(\omega \rightarrow \alpha\omega)\bigwedge(\omega \rightarrow \alpha\omega\beta) \Rightarrow (\omega \rightarrow \omega\beta)$, for example, (a capitula → capitula*ţie*) $\bigwedge$(a capitula → *re*capitula*ţie*) ⇒ (a capitula → capitula*ţie*);

$(\omega \rightarrow \alpha\omega\beta)\bigwedge(\omega \rightarrow \omega\beta) \Rightarrow (\omega \rightarrow \alpha\omega)$, for example, (a centraliza → *des*centraliza*tor*)$\bigwedge$(a centraliza → centraliza*tor*) ⇒ (a centraliza → *des*centraliza);

Formal derivational rules For example, let be an adjective of the form $\omega\prime = \omega\beta$, where $\beta \in$ {-tor, -bil, -os, -at, -it, -ut,-ind, -Înd}, as a result we will obtain the derivatives of the form $\omega\prime\prime = ne\omega\beta$, for example, invidios → *ne*invidios.

Derivational constraints These can be functions of the form:

$$f : \{wrd, pos, mod, sla, fgw, mvca \dots\} \rightarrow derivative$$

where *wrd* is a word to derivate, *pos* − part of speech of *wrd*, *mod* − model of derivation, *sla* − the set of letters to which the affix is attached, *fgw* − flection group of *wrd*, *mvca* − modifications and vocalic or consonant alternations. For example, f : { a spinteca, verb, des< *verb* >, ...s..., V14, double consonant avoiding } → de(s)spinteca.

As derivatives generation represents an overgenerating mechanism, these derivational models implementation needs a level of validation [3], in order to enrich existent digital lexicons.

Generation of derivatives is not a problem trivial, because the process does not have a regular mechanism. Solution to store all derivatives of a dictionary is a reasonable one because, these derivatives still will not cover the full diversity of language, being in continous evolution [4].

References. **1.** Alhazov A., Boian E., Cojocaru S. Modelling Inflections in Romanian Language by P Systems with String Replication. In: Proc. of the 10th Workshop on Membrane Computing, WMC10, Curtea de Arges (Romania), August 24 – 27, 2009, p.116 -128. **2.** Cojocaru, S., Boian, E. Determination of inflexional group using P systems. Computer Science Journal of Moldova. 2010, 18(1), 70-81. ISSN 1561-4042. **3.** Cojocaru S., Boian E., Petic P. Stages in automatic derivational morphology processing, In: KEPT2009, Knowledge Engineering, Principles and Techniques, Selected Papers, Cluj-Napoca (Romania), July 2 - 4, 2009, pp. 97-104. **4.** Petic M. Generative mechanisms of Romanian derivational morphology. In: Memoirs of the Scientific Section of the Romanian Academy, Series IV, Tome XXXIV, Bucureşti: Editura Academiei, 2011, p. 21-30.

***Ptuha Yu.A.*[1], *Slashcheva T.G.*[2], *Slyota M.R.*[1], *Tarasiuk T.S.*[1]**
[1]*Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine;* [2]*Clinical hospital "Feofaniya", Kyiv, Ukraine*

Modern problems in prevention and treatment of cardiovascular disease

Introduction. A great problem faced by modern medicine is the difficulty of chronic illness treatment. A major part of these issues constitutes cardiovascular diseases(CVD). Most cardiovascular diseases can be prevented by addressing behavioral risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity and harmful use of alcohol using population-wide strategies. But people who are at a high cardiovascular disease risk (due to the presence of one or more risk factors such as hypertension, diabetes or hyperlipidaemia) need proper monitoring and heart malfunction early detection personal systems.

In this paper the official statistics concerning the lethality caused by CVD are described and the role of information technology in solution of this problem is studied to the great extent.

Relevance. The common problem of modern medicine is that many heart diseases do not show up their symptoms during the long period. Advanced stage of heart disease often lead to chronic ailment, that significantly reduces human life and doesn't allow to lead a normal life.

Cardiovascular diseases remain the most prevalent illness in the world. An estimated 17.5 million people died from CVDs in 2012, representing 31% of all global deaths(see fig.1). Of these deaths, an estimated 7.4 million were due to coronary heart disease and 6.7 million were due to stroke [1,2].

In the process of detailed study of the reports, it was found that the level of GDP is correlated with the mortality rate of CVD. At least three quarters of the world's deaths from CVDs occur in low- and middle-income countries.

At macro-economic level, CVDs place a heavy burden on the economies of low- and middle-income countries. Noncommunicable disease including cardiovascular disease and diabetes are estimated to reduce GDP by up to 6.77% in low- and middle-income countries experiencing rapid economic growth, as many people die prematurely [2].

It is necessary to pay people's attention to the existing problem in our country. We need to adopt the experience of European specialists, where the mortality rate of CVD is significantly lower. But even in developed countries, doctors are still unable to counteract untimely deaths caused by disorders of the cardiovascular system. Even with the aid of good equipment. So, it is necessary to apply novelties of modern technology, which are becoming more accessible to the great majority of population every year.

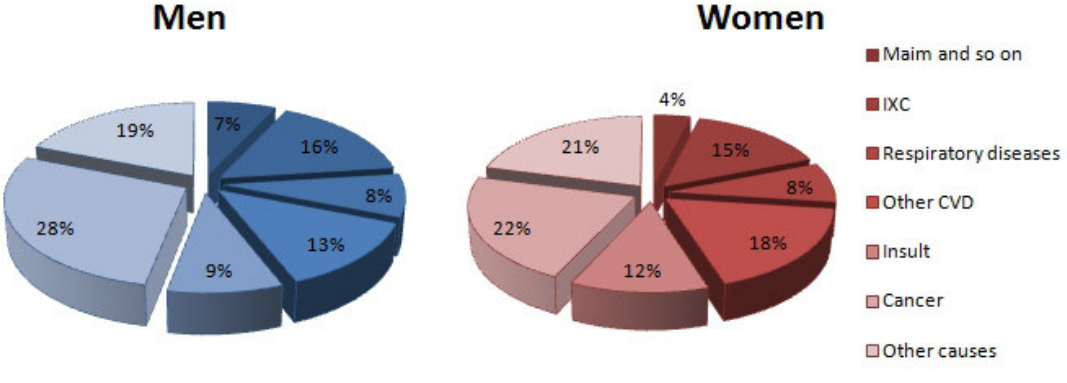

Figure 1. The death causes distribution for men and women of all ages in the European Union.

In Ukraine, this problem is far from easy solution. In 2003, the main noncommunicable diseases

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

accounted for about 80% of all deaths in Ukraine. In total 60% of all deaths were caused by diseases of the circulatory system and 12% by cancer. The mortality rate among males attributed to cardiovascular diseases is the third highest in European countries. Ischemic heart disease is the single biggest killer in Ukraine, being responsible for almost 40% of all male deaths in 2003. In 2014, mortality rate of CVD are 68% of all diseases in the Ukraine [2].

The need of IT. In spite of rapid development of science and medical technology the number of working-age population in the US and Europe is steadily decreasing. We have a problem of finding new methods that can provide every person a healthy and long life. But the solution of this problem is closer than we think, we just need to answer one question: do we actually care about our health? Most of us do it only when health issues became visible. It means that disease has been successfully thrived and probably became chronic. We need to avoid it.

In modern cardiology automatic prediction and detection of heart pathologies is becoming more popular. These approaches are based on special methods that analyze electrocardiogram (ECG) recordings. Automatic system for heart pathologies detection can be build using machine learning techniques applied to ECG signals taken from portable sensors.

Not long ago people could easily live without mobile phones, digital cameras, small laptops and other devices which have become integral part of our daily life. These gadgets have changed our life style significantly. Nowadays information technologies are rapidly developing and it is possible to make devices that can perform twenty-four-hour monitoring of our health. These devices using machine learning methods can't provide absolutely 100% correct diagnosis and cannot replace doctors, but with their help it is possible to warn us in time about any tiny problems with our health that are absolutely invisible for our eyes and feelings. Model can warn about certain abnormalities in the cardiovascular system at the time when we cannot even see or feel any discomfort. It will give us a possibility to seek help from our doctor to avoid future complications. This approach can help us to detect specific deviations or pathologies in time and with the early stage detection we can prevent diseases from progression or becoming chronic.

Information and communications technologies are able to provide a day-and-night connection between patient and doctor. That allows to transmit correct data from one to another anytime and anywhere. So, the physician will be notified of patients' health issues when they've just appeared. And this issue is very important in CVD disease treatment. For example, smart clothes take indicators of biometric parameters of the body, share them to the laptop or other devices for analysing and in a moment give an advice. The program also could correct your timetable to minimize the risk of attenuation of the organism. Even now world-wide organizations create databases which contain deviances from physiology norms during illness period. This information can be successfully used for the "personal doctor" learning.

Conclusions. In this paper we described the importance of early detection and treatment of CVD and it's impact on mortality rates and GDP. It is shown that modern information technologies could help to create personal automatic health monitoring systems that will aid doctors to make the right diagnosis in time and detect specific deviations or pathologies which will prevent diseases from progression or becoming chronic. Also the usage of portable sensors for timely identification of disorders in the cardiovascular system may reduce the mortality rate of CVD. The IT innovations can propel the quality of modern medical services and improve health care system.

References. **1.** European Cardiovascular Disease Statistics, 2008. - London: British Heart Foundation, http://medbe.ru/materials/obshchee/statistika-serdechno-sosudistykh-zabolevaniy-v-evrope/. **2.** World Health Organization, http://who.int/mediacentre/factsheets/fs317/en/.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Savastiyanov V. V.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Discovering of potential positive and negative factors of social disaster using sentiment analysis

Mitigation of social disaster factors caused by catastrophes, terrorism, war or natural disaster is one of the vital topics in foresight of sustainable development of society [1, 2]. Social disaster factors are the most dangerous in the long term period and cause consequences that threaten society member's life due to crimes, radical ideas dissemination and spreading of dangerous habits (drugs, alcohol, etc.).

The first weak signal of social disaster appear under topics and discussions in social media. And under disaster pressure neutral objects in social media become carriers of negative meaning. In that way we would call as product any vector of emotions connected by context to any social disaster dimension, and as feature - any new object associated with negative emotion. I.e. after some situational consequences in society word "reforms" become associated with "unclear, fear, protest, distrust" product and any new mention of "reforms" in media could worsen public mood. The aim is to discover migration of objects in emotional field for particular time frame for media corpus classified as having social disaster factors.

There is a known technique to extract product features, which imply opinions [3] [Liu et al, 2011]. It allows automatically extract possible features from subject domain, which are declaring desirable or undesirable facts. In this paper modified method of determining the impact of the emotional context on potential features candidates is used. Additionally to the distance weight coefficient ω^s in the aggregation formula for total score of emotional context is used:

$$score_k(f) = \sum_{w_i : w_i \in S \cap w_i \in L_k} \frac{w_i.SO}{dis(w_i, f)} . \omega_k^s(w_i, t_j) \tag{1}$$

$$\omega_k^s(w_i, d, D, t_j) = SF(w_i, d) * SIDF(w_i, D, t_j) \tag{2}$$

where w_i is an opinion word, L_k - is the set of all opinion words, $w_i \in L_k$, k – dimension of emotional space – some of them are depicted in [4] [He, 2012] (i.e. Plutchik space with $k_{Plutchik}$=8, k_{Arnold}=11, k_{Ekman}=6, etc.), S - sentence that contains the feature f, dis(w_i,f) - the distance between opinion word w_i and feature f, w_i.SO - is the semantic orientation of word, $t_j \in [t_{start};t_{end}]$ – time boundaries (time horizon) which limits the existence of chosen emotional space and under which significant trends [5] [Pillkahn, 2008] are relevant in researched system. Time horizon is a key parameter for emotion calculation on fact based corpus enlightening temporal events (i.e. economical or political crisis time, war time, social crisis, terrorists act or natural cataclysm, etc.) and in which emotions about events or features are skewed under pressure of stress, authorities or government agencies (see example in Table 1). Weight coefficient ω is calculated similar to TF-IDF [6] [Wu, 2008] with only difference that text corpus origin dates are limited with time horizon and all calculations are in chosen dimension k of emotional space:

$$SF(w_i, d) = \frac{n_{w_i}}{\sum_k n_{w_k}} \tag{3}$$

$$SIDF(w_i, D, t_j) = \log \frac{|D_{t_j}|}{|(w_i \supset d_i)|}, t_j \in [t_{start}, t_{end}] \tag{4}$$

where n_{w_i} – frequency of opinion colored term, n_{w_k} – frequency of other opinion colored terms from chosen dimension k, $|D_{t_i}|$ - length of all corpus limited by time horizon, $|w_i \supset d_i|$ - length of all corpus limited by time horizon which contain opinion colored term w_i.

One more modification to the scoring of the importance of feature in task class of high, low, increased and decreased quantity of a positive or negative potential item were to include additional

This research was sponsored by the NATO Science for Peace and Security Programme under grant SPS G4877.

Table 1. Example of term roaming in emotional space

Term (Product or Feature)	Time frame (week)	Dominant emotion from emotional space
Donetzk region	23	anxiety
Donetzk region	25	anxiety
Donetzk region	29	intended act
Donetzk region	30	desire
Donetzk region	32	sorrow
Donetzk region	40	aversion

weight in dimension of features with time boundaries:

$$score_k(f_l) = \sum_{w_i : w_i \in S \cap w_i \in L_k} \frac{w_i.SO}{dis(w_i, f_l)} \omega_k^s(w_i, t_j) \omega_l^f(f_l, D, t_j) \quad (5)$$

where ω_l^f is calculated similar to ω_k^s.

The formula could also be expanded on other sentiment/opinion classes. The main sense is to lower the significance of the most frequent features met in text to be aware of opinion spamming and avoid of "popularity" in communities or message boards due to "Internet meme" [7] or similar effect.

That approach is promising to be relevant under time limited conditions where in news corpus some features (and sentiments) are implicitly be referred more frequently under pressure of political situation, propaganda, etc. Mined object could be user in rule building approaches for sentiment analysis [2].

Also that approach is possibly to use under any other subject domain where user opinion is changing in time frame under the influence of external forces, by emotional impact of new subjects or objects on personal meaning, by improved quality of reviewed objects and subjects in new product line/generation/situation.

References. **1.** Zgurovsky, Mikhail Z., Pankratova, N.D., System Analysis: Theory and Applications, Springer, 2007, ISBN 978-3-540-48880-4. **2.** Pankratova N., Savastiyanov V. Foresight Process Based on Text Analytics // Inter- national Journal «Information Content and Processing». — 2014. — 1, No 1, ITHEA. — P. 54–65. **3.** Zhang, Lei and Bing Liu. Identifying noun product features that imply opinions. in Proceedings of the Annual Meeting of the Association for Computational Linguistics (short paper) (ACL-2011). 2011b. **4.** Yulan He, A Bayesian Modeling Approach to Multi- Dimensional Sentiment Distributions Prediction, Knowledge Media Institute The Open University, UK, 2012. **5.** Ulf Pillkahn, 'Using Trends and Scenarios as Tools for Strategy Development', Siemens, 2008. **6.** Wu HC, Luk RWP, Wong KF, Kwok KL (2008). "Interpreting tf–idf term weights as making relevance decisions". ACM Transactions on Information Systems 26 (3): 1–37. doi:10.1145/1361684.1361686. **7.** "Meme". Wikipedia. Retrieved 2015-05-15 **8.** McNamara, Adam (2011), "Can we measure memes?", Frontiers in Evolutionary Neuroscience 3, doi:10.3389/fnevo.2011.00001.

Savchenko I.O.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Modeling complex socio-technical systems using networks of morphological tables

One of the highly productive ways of dealing with complex, unstructured problems is applying the modified morphological analysis method [1]. The essence of this method is in analyzing objects, processes or phenomena by picking several their characteristics which can have alternate values. By combining different values for each characteristic, a very large multitude of object configurations is produced and can be easily analyzed using expert estimations for the constituent parts.

The procedure of morphological analysis, described in [1], is suitable for dealing with relatively small or poorly detailed problems, up to 8–10 parameters. Adding more parameters inflates the cross-consistency matrix to excessively large sizes, which makes impossible to evaluate and operate with such matrices. Moreover, large parts of this matrix are empty, as groups of parameters are often independent.

To deal with this problem, a modular approach is proposed, which assigns separate morphological tables to different aspects of the problem, or different objects in a system. The connections between the tables are described by a consistency matrix, and the dependence can be either bidirectional, or unidirectional (causal). A group of morphological tables, linked by bidirectional dependences, technically create a single combined morphological table, which is processed by a system of equations [1]. However, a modular approach has many advantages. As each table in a group has its own cross-consistency matrix, other tables can be added or removed from the group without having to re-evaluate the whole virtual cross-consistency matrix uniting all tables in a group. Single morphological tables in a group can be assessed by different groups of experts, that have knowledge exactly in the field of the corresponding morphological table. A repository of these tables with evaluated cross-consistency matrices can be created if the problem is going to be analyzed several times. As for the consistency matrices that connect tables in a group, they are mostly sparse and easier to estimate.

The unidirectional type of connection corresponds to the two-stage morphological analysis procedure, where the independent parameters are calculated at the first stage, and then the expected efficiencies of the second morphological table are calculated at the second stage. However, the number of stages is actually not limited to two. Modeling complex socio-technical systems can require three and more steps of calculations. The dependences between morphological tables can be presented as a graph. A single table or a group of morphological tables connected by bidirectional dependencies constitutes a node of the graph, and a unidirectional link represents an arc. First the morphological analysis procedure is conducted for the source(s) of the graph. Then this procedure is conducted for the nodes that are directly connected with the already calculated nodes, etc.

In the NATO supported project NATO.NUKR.SFPP 984877, which is carried out in IASA, a cascade of morphological tables is used to describe the decision to be made for the social disaster. A first group of tables consists of the table with common disaster parameters (which are applicable for all disasters), and a table with specific disaster parameters (this table contains parameters which are relevant for the specific type of disasters). A second table describes the consequences of the disaster, and the third table contains the measures of preventing and mitigating the considered disasters. The operation of the method can be in two modes: evaluating preparedness (in this case the probabilities of a certain disaster parameters are evaluated, and efficiency of measures is estimated for a hypothetical variety of disaster situations), and reaction (in this case the disaster already happened, and its parameter alternatives are known, so the measures are evaluated according to a known configuration of tables for non-controlled parameters).

References. **1.** Панкратова Н.Д., Савченко І.О. Морфологічний аналіз. Теорія, проблеми, застосування. Навчальний посібник. – Київ: Вид-во «Наукова думка». – 2015. – 244 с.

This research was sponsored by the NATO Science for Peace and Security Programme under grant 984877.

Schuka R.V., Ivanov S.S.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Analysis of crime in Ukraine with using methods of regression analysis in SAS Enterprise Guide system

Introduction. Nowadays information technologies are widely used in many areas of human's activity, for gathering, accumulation and forward analyzing of this data. World leader in data analyzing is SAS, which offers large spectrum of analytical methods and instruments. Since 2010 Ministry of Internal Affairs (MIA) of Ukraine is using specialized SAS's software for statistical data analysis.

Problem statement. Information MIA's system of citizenship's treatment registration consists millions of commited crimes records. This statistical information was represented with using of ETL's procedures[2], on SAS Base language, as more 50000 time series in monthly basis, for all country's regions (25 regions) and crime's types (16 categories and 123 subcategories), according to Criminal Procedure Code. Each time series should be analyzed for the presence of trends, seasonality and anomalies to create forecasting model.

Solution. SAS Enterprise Guide 6.1 is used for automatically time series handling, which is a powerful instrument for creating of mathematical models based of time series, specifically regression models, non-linear regression, exponential smoothing, auto regression, auto regression with integrated trend and moving average. For example, mathematical model of crime's number in the scope of drug's trafficking and production in Kyiv is next:

$$y(t) = 30 + 0,3 * y(t-1) + 0,2 * y(t-2) - 0.9 * PCA_1(t-1) + 7 * PCA_2(t-1) \quad (1)$$

$$PCA_1(t) = -0,2 * d(t) - 0,36 * w(t) + 0,42 * u_1(t) + 0,42 * u_2(t) + 0,38 * u_3(t) \quad (2)$$

$$PCA_2(t) = 0,85 * d(t) + 0,15 * w(t) \quad (3)$$

where y – number of crimes, t – temporal dimension (month), $PCA_1(t)$ and $PCA_2(t)$ – first and second principal components, which were received with principal components method , wrote as linear combination of inputted regressors, d – wage's debt in Kyiv (mln. grn.), w – average wage (grn.), $u_1(t)$ – common percent of unemployed , $u_2(t)$ – percent of unemployed women, $u_3(t)$ - percent of unemployed men [1].

Conclusions. SAS analytical tolls allows create models automatically or with using of experts [3]. Standard deviation of created model is 19, the coefficient of determination is 0,4. Sample of retrospective data from January 2005 to December 2014 used to train the model, and the period from January to June 2015 for forecasting. Further on the base of SAS technology planned to realize automatic rapid prediction system of the number of crimes for all crime subtypes. This work made within scientific research project NUKR.SFPP G4877 "Modeling and reducing impact of social disasters, which occurred because of catastrophes and terrorism" (NATO scientific program "Science for peace and security"), whereby MIA of Ukraine acts as end-user of received results.

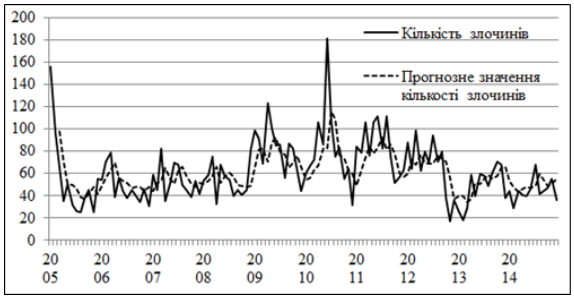

Figure 1. Chart of real and forecasted model of crime's number.

References. **1.** Бідюк П.І. Прикладна статистика (навчальний посібник) / П.І. Бідюк, О.М. Терентьєв, Т.І. Просянкіна-Жарова. – Вінниця : ПП "ТД Едельвейс і К", 2013. – 288 с. – ISBN 978-966-2462-21-0. **2.** Терентьев А.Н., Домрачев В.Н., Костецкий Р.И. SAS BASE: Основы программирования (научное издание). – К: Эдельвейс, 2014. – 304 с. – ISBN 978-966-2748-49-9. **3.** SAS Enterprise Guide tutorial. [Електронний ресурс]. – Режим доступу: URL : https://support.sas.com/documentation/onlinedoc/guide/tut71/en/menu.htm.

Shcherban M.M., Bidyuk P.I.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Modeling and forecasting macroeconomic processes

A new approach to modeling and short term forecasting of macroeconomic processes is proposed based on automatic analysis of statistical data. The system proposed contains a set of data processing and forecasting techniques that provide a possibility for computing short-term forecasts of acceptable quality.

To improve substantially quality of forecasts for macroeconomic processes it is useful to construct adaptive computing scheme for a model and parameter estimation (Fig.1). Each step of data processing is controlled by appropriate set of statistical parameters each of which characterizes specific features of data, model as a whole, model parameters and quality of forecasts estimates. We propose new adaptive computing scheme that is distinguished with several possibilities for adaptation using complex quality criterion. The data collected should be correctly prepared for model structure and parameter estimation. The model structure and parameter estimation is a key issue for reaching necessary quality of forecasts. It is proposed to define model structure as follows: $S = \{r, p, m, n, d, z, l\}$, where r is model dimensionality (number of equations); p is model order (maximum order of differential or difference equation in a model); m is a number of independent variables in the right hand side; n is a type of nonlinearity; d is a lag or output reaction delay time; z is external disturbance and its type; l are possible restrictions for variables. For automatic search of the "best" model it is proposed to use the following criteria:

$$V_N(\theta, D_N) = e^{|1-R^2|} + ln(1 + \frac{SSE}{N}) + e^{|2-DW|} + ln(1 + MSE) + ln(MAPE) + e^U$$

where θ is a vector of model parameters; N is a power of time series used; R^2 is a determination coefficient; DW is Durbin-Watson statistic; MSE is mean square error; $MAPE$ is mean absolute percentage error; U is Theil coefficient. The power of the criterion was tested experimentally with a wide set of models.

There are several possibilities for adaptive model structure estimation: (1) automatic analysis of partial autocorrelation for determining autoregression order; (2) automatic search for the exogeneous variable lag estimate (detection of leading indicators); (3) automatic analysis of residual properties; (4) analysis of data distribution type and its use for selecting correct model parameters estimation method; (5) adaptive model parameter estimation with hiring extra data; (7) optimal selection of weighting coefficients for exponential smoothing, nearest neighbor and some other techniques; (6) the use of adaptive approach to model type selection. The use of specific adaptation scheme depends on volume and quality of data, specific problem statement, requirements to forecast estimates, etc. In some cases we used successfully logistic regression together with linear regression to describe the data with nonlinearity. These models as well as classification trees and Bayesian networks have been used to forecast the direction of stock price movement. Application of the concept described provides the following advantages: (1) automatic search for the "best" model reduces the search time for many times; (2) it is possible to analyze much wider set of candidate models than manually; (3) the search is optimized thanks to the use of complex quality criterion; (4) in the frames of computer system developed it is possible to integrate ideologically different methods of modeling and forecasting and compute combined forecasts estimates that are distinguished with better quality. Testing of the system with wide set of macroeconomic and stock price data showed that it is possible easy to reach a value of absolute percentage error of about 3–4% for short term forecasting. The use of dynamic and static forecasts allows for generating necessary forecasts estimates depending on specific problem statement.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

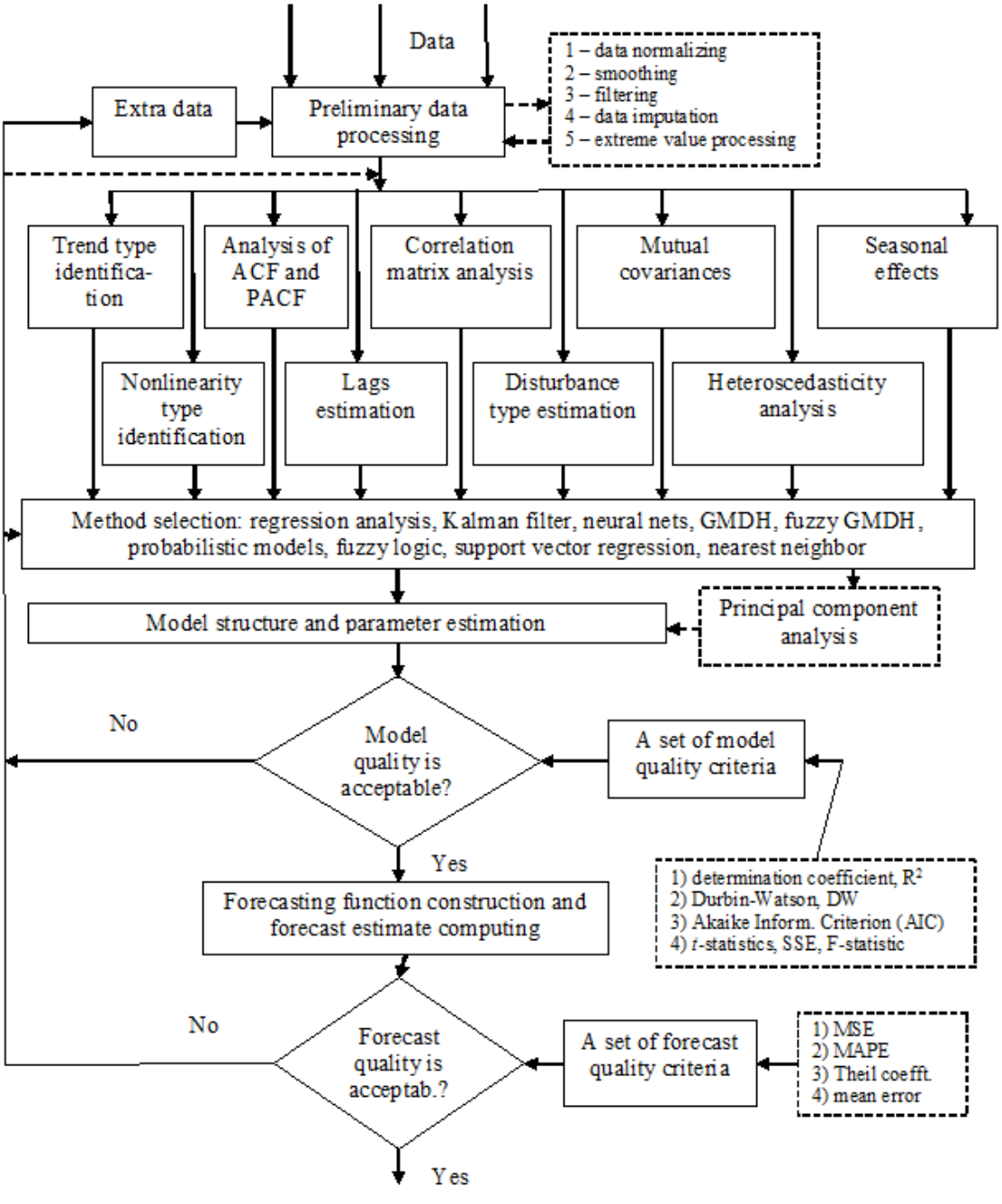

Figure 1. Flowchart for adaptive model estimation and forecasting

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Tripolsky O. V.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Probabilistic methods for studying of optimal gambling systems

The fundamental problem in gambling is to find positive expectation betting opportunities. The analogous problem in investing is to find investments with excess risk-adjusted expected rates of return. Once these favorable opportunities have been identified, the gambler or investor must decide how much of his capital to bet. This is the problem which considered in this paper. In the stock market the problem is similar but more complex [1].

These settings can be solved by using the Kelly system, which maximizes the expected value of the logarithm of wealth period by period [2].The main advantage of the Kelly criterion, is that it maximizes the limiting exponential growth rate of wealth.

Wong demonstrated that Kelly betting has two disadvantages over flat betting. These disadvantages are related, although they are not quite the same. The first disadvantage is that, for the case when the number of games won is equal to the expected value, the net gain of the Kelly bettor is about half of that of the flat bettor, for the same initial capital, the same number of games and the same edge. The second disadvantage that Wong considered is that the win to action ratio for Kelly betting with the optimal fraction is about one-half of that for flat betting.But for long term compounders, the good properties dominate the bad properties of the Kelly criterion [3, 4].

In this research we examine what happens when we bet a fraction of Kelly in terms of the risk of losing specified proportions of one's bank.The main result presented gives the probability that one will win a specified multiple of one's bankroll before losing to specified fraction as a function of the fraction of Kelly bet [5].

In red and black, a player bets, at even stakes, on a sequence of independent games with success probability p, until he either reaches a fixed goal or is ruined. In this paper were explored two strategies: timid play in which the gambler makes the minimum bet on each game, and bold play in which she bets, on each game, her entire fortune or the amount needed to reach the target (whichever is smaller). We study the success probability (the probability of reaching the target) and the expected number of games played, as functions of the initial fortune. The main interest is success probability, the probability of reaching target. A secondary interest is the expected number of games played. Two very different strategies were compared: with timid play, the gambler makes a small constant bet on each game until he is ruined or reaches the target. This turns out to be a very bad strategy in unfair games, but does have the advantage of a relatively large expected number of games. If you play this strategy, you will almost certainly be ruined, but at least you get to gamble for a while. With bold play, the gambler bets her entire fortune or what he needs to reach the target, whichever is smaller. This is an optimal strategy in the unfair case; no strategy can do better. If you play this strategy, there is a good chance that your gambling will be over after just a few games. Amazingly, bold play is not uniquely optimal in the unfair case. Bold play can be rescaled to produce an infinite sequence of optimal strategies. These higher order bold strategies can have large expected numbers of games, depending on the initial fortune (and assuming that the casino is nice enough to allow you to make the weird fractional bets that will be required for the strategy) [6].

References. **1.** E.Thorp, *The Kelly criterion in blackjack, sports betting, and the stock market*, 2006, pp.10-13. **2.** J.Kelly, *A new interpretation of information rate*, Bell System Tech.J. 35, 1956, pp.917-926. **3.** R.Phatarfod, *Some aspects of gambling with the Kelly Criterion*, Math.Scientist, pp. 23-31. **4.** S.Wong, *What proportional betting does to your win rate. Blackjack World*, vol.4, 1981, pp.162-168. **5.** W.Chen, M.Ingenoso, *Risk formulas for proportional Betting . In Optimal play: Mathematical studies of games and gambling*, Institute for the study of gambling and commercial gaming, Reno, 2007, pp.541-548. **6.** Kyle Siegrist, *How to Gamble If You Must*, University of Alabama in Huntsville, www.unc.edu/~hannig/STOR089/handouts/Casino.pdf.

Белошицкая О.К.
Национальный технический университет Украины "КПИ", факультет биомедицинской инженерии, Киев, Украина; Государственное научное учреждение "Научно-практический центр профилактической и клинической медицины" Государственного управления делами, Киев, Украина

Исследование поведения нейронной сети головного мозга при различных патологиях с помощью методов нелинейной динамики

Функционирующий головной мозг представляет собой чрезвычайно сложную нелинейную систему, т.е. любое малейшее изменение внешнего условия и любое минимальное воздействие может привести к значительному изменению в деятельности системы.

Электроэнцефалограмма (ЭЭГ) представляет собой сложный колебательный электрический процесс, который может быть зарегистрирован при расположении электродов на мозге или на поверхности скальпа, и является результатом электрической суммы и фильтрации элементарных процессов, которые происходят в нейронах головного мозга 1.

Электрическая активность отдельных нервных клеток отражает их функциональную активность по переработке и передаче информации. Суммарная ЭЭГ отражает функциональную активность, но уже не в отдельных нервных клетках, а их больших популяций, то есть функциональную активность мозга. Это очень важно для анализа ЭЭГ, поскольку дает понимание того, какие системы мозга определяют внешний вид и внутреннюю организацию ЭЭГ.

Характер ЭЭГ определяется функциональным состоянием нервной ткани, а также обменными процессами, которые происходят в ней. Нарушение кровоснабжения приводит к угнетению биоэлектрической активности коры больших полушарий. Важной особенностью ЭЭГ является ее спонтанный характер и автономность. В электроэнцефалографии различают четыре основных диапазона: альфа-, бета-, гамма- и тета-волны.

Существование характерных ритмических процессов определяется спонтанной электрической активностью мозга, обусловленная суммарной активностью отдельных нейронов. Ритмы электроэнцефалограммы отличаются друг от друга по продолжительности, амплитуде и форме.

Реальные данные ЭЭГ получено во время электроэнцефалографических исследований на базе Государственного научного учреждения «Научно-практический центр профилактической и клинической медицины» Государственного управления делами. Запись ЭЭГ проводилась за международною схемой отведений «10-20» (Jasper H., 1957) с использованием 16-канального телеметрического электроэнцефалографа Expert компании Tredex с видеомониторингом с верхним граничным фильтром 30 Гц, постоянной времени 0,1 с и с частотой дискретизации 400 Гц. Визуальный анализ ЭЭГ осуществлялся с помощью программы ExpertNet. В программе ExpertNet применяется разделение экрана: на одной половине можно контролировать и просматривать ЭЭГ, на второй – видеомониторинг записи.

База данных ЭЭГ (156 клинических случаев) была разделена на 3 группы: очаговые поражения (например, эпилепсия), системные поражения (например, цереброваскулярные болезни) и норма. Данные по трем группам были проанализированы вместе с врачом-энцефалографистом с медицинской точки зрения, заполнены протоколы исследований и сделаны заключения о функциональных и морфологических особенностях динамики обследуемого, о характере и выраженности церебральных патологий (если такие есть). Далее каждая ЭЭГ была проанализирована и сопоставлена с плоскостью стохастического гомеостаза согласно теории Р. Юсупова и Р. Полонникова 3. Плоскость состоит с пяти фаз, каждая из которых описывает состояние головного мозга. Среди этих фаз выделена норма и четыре фазы, которые в целом описывают суперкритический и субкритический состояния.

Нулевая фаза называется «норма» и характеризует нормальное состояние организма человека. Фаза I характеризует «продромальный состояние» (состояние, предшествующее заболеванию). В этой фазе организм человека в дисбалансе, что может привести к возвращению его в норму или перехода в фазу II в. Фаза II характеризует острое состояние течения заболевания. Результатом пребывания организма человека в этом состоянии может стать возвращение в

фазу I и из нее к норме или переход в фазу III, которая характеризуется хроническим течением заболевания. IV фаза - худший случай состояния регуляторных систем организма. Иерархическая сложность ЭЭГ в этой фазе, близкой к нулю, характеризует истощения регуляторных резервов, что может привести к смерти.

Собственно говоря, эта плоскость отражает такие свойства мозговой активности, как сложность и вариабельность (амплитуда).

Так как мозг работает в нелинейном, хаотичном режиме (Albert et al., 1994), ЭЭГ можно рассматривать как результат динамической системы мозга, в ответ на воздействие внешней и внутренней среды. Система содержит много нелинейных компонентов (нейроны) и имеет определенные потенциальные идентифицирующие параметры системы.

Деятельность мозга может быть описана в рамках диссипативных систем. Наличие детерминированного хаоса в таких системах можно установить, определив любую из динамических переменных и при этом, в частности 4 можно оценить:

- размерность реконструкции (размерность вложения аттрактора) – ее величина позволяет говорить о том, как много компонентов формируют изучаемую нейродинамическими систему;
- корреляционная размерность (нижняя граница размерности Хаусдорфа), которая является характеристикой «сложности» динамической системы мозга;
- старший показатель Ляпунова (максимальная экспонента Ляпунова), который является характеристикой «хаотичности» поведения динамической системы мозга;
- нижняя граница энтропии Колмогорова-Синая (корреляционная энтропия), которая показывает степень «предсказуемости» процессов, происходящих в мозгу;
- энтропия Колмогорова-Синая, характеризующая скорость потери динамической информации системой, которая позволяет оценить промежуток времени, на которых можно сделать прогноз поведения нейродинамической системы;
- логистическое отображение – это полиномиальное отображение, которое описывает, как меняется численность популяции нейронов с течением времени;
- показатель Херста – используется для анализа временных рядов сложных систем, содержит минимум предположений об изучаемой системе и позволяет ввести классификацию временных рядов в зависимости от своего значения.

Приступы генерируются в детерминированных нелинейных хаотических системах, возникновение эпилептических припадков можно представить в виде прерывистых фазовых переходов системы.

Начало приступа представляет собой пространственно-временной переход от сложных к более упорядоченным состояниям.

Внимание к проблеме заблаговременного определения эпилептических приступов обосновано существенным числом пациентов, страдающих неконтролируемой формой заболевания, нередко сопровождающейся моторными судорогами и потерей сознания.

С помощью методов нелинейной динамики можно спрогнозировать начало приступа, его продолжительность и окончание. В дальнейшем эти методы будут также использованы при моделировании патологий головного мозга с помощью клеточных автоматов.

Литература. **1.** Меклер А.А. Программный комплекс для анализа электроэнцефалограмм методами теории динамического хаоса: Дисс. канд. техн. наук: 05.13.18 – СПб. 2006 г. – 168 с. **2.** Аристов А.В. Фрактальный анализ электрической активности головного мозга при генерализованной эпилепсии: Дисс. канд. биол. наук: 03.03.01 – М., 2010 г. – 156 с. **3.** Юсупов Р. М., Полонников Р. И. Телемедицина. Новые информационные технологии на пороге XXI века. – СПб.: Изд-во«Анатолия», 1998. – 488 с. **4.** Майоров О.Ю., Фенченко В.Н. Исследование биоэлектрической активности мозга с позиций многоразмерного линейного и нелинейного анализа ЭЭГ// Клиническая информатика и телемедицина. – Харьков, 2008. – Т.4, вып. 5. – С. 12-20.

***Бояринова Ю.Е.*[1], *Калиновский Я.А.*[2], *Туренко А.С.*[2]**
[1] *Национальный технический университет Украины "КПИ", Киев, Украина;* [2] *Институт проблем регистрации информации НАН Украины, Киев, Украина*

Обобщение кватернионов вращения и связь с процедурой удвоения Грассмана-Клиффорда

Исследованы связи между обобщенными кватернионами и классом некоммутативных гиперкомплексных числовых систем четвертой размерности, построенных с помощью некоммутативной процедуры удвоения Грассманна-Клиффорда систем второй размерности.

Гиперкомплексные числовые системы находят немало применений в науке и технике. Впервые обобщенные кватернионы были применены при изображении пространственно-временных групп Куртом Геделем в 1949 году, с помощью которых он представил решение уравнений Эйнштейна гравитационного поля. Обобщенные кватернионы также исследованы в работах многих других авторов [1, 2].

Обобщенный кватернион имеет вид:

$$q = a_1 e_1 + a_2 e_2 + a_3 e_3 + a_4 e_4 \tag{1}$$

где a_i - действительные числа, а e_i при $i = 2, 3, 4$ - мнимые единицы, удовлетворяющие следующей таблице Кэли, где $\alpha, \beta \in R$:

$H_{\alpha,\beta}$	e_1	e_2	e_3	e_4
e_1	e_1	e_2	e_3	e_4
e_2	e_2	$-\alpha e_1$	e_4	$-\alpha e_3$
e_3	e_3	$-e_4$	$-\beta e_1$	βe_2
e_4	e_4	αe_3	$-\beta e_2$	$-\alpha\beta e_1$

Также, в работе [2] рассмотрены конкретные ГЧС в зависимости от того, какой знак имеют $alpha$ и $beta$: $alpha = 1, beta = 1$, тогда H - система кватернионов; $alpha = 1, beta = -1$,тогда H - система антикватернионов; $alpha = 1, beta = 0$, тогда H - система псевдо-кватернионов;$alpha = -1, beta = 0$, тогда H - система псевдо-антикватернионов;$alpha = 0, beta = 0$, тогда H - система 1/4-кватернионов.

Как показано в работе [3], существует связь между обобщенными кватернионами и классом ГЧС, полученным с помощью некоммутативной процедуры удвоения Грассманна-Клиффорда (ГК-процедура), который состоит из шести представителей классов изоморфизмов: Д(C,C,4)=H- система кватернионов, Д(C,W,4)=AH система антикватернионов, Д(C,D,4)=Д(D,C,4),Д(W,W,4), Д(D,D,4), Д(W,D,4)=Д(D,W,4).

Если анализировать полученные результаты, то можно увидеть, что система антикватернионов AH соответствует случаю обобщенных кватернионов при $alpha = 1, beta = -1$; системе Д(C,D,4) соответствует $alpha = 1, beta = 0$; системе Д(W,D,4) соответствует $alpha = -1, beta = 0$; и, наконец, пятому случаю 1/4-кватернионов соответствует система Д(D,D,4). В случае обобщенных кватернионов рассматриваются 5 отдельных ГЧС, а с помощью ГК-процедуры получено 6 систем. То есть, система Д(W,W,4) для обобщенных кватернионов не рассматривается. Проанализировав таблицу Кэли данной системы можно определить, что она соответствует таблице умножения базисных элементов обобщенных кватернионов при $alpha = -1, beta = -1$.

Литература. **1.** Szeto G. On generalized quaternion algebras / G. Szeto // Internat. I. Math. And Math. Sci. – 1980. – Vol. 3, No. 2. – P. 237-245. **2.** Jafari M. Generalized quaternion and rotation in 3-space/ M. Jafari, Y. Yayli // Department of Mathematics, Faculty of Science Ankara University, 06100 Ankara, Turkey. - 11 p. **3.** Kalinovsky Y. O. Computing Characteristics of One Class of Non-commutative Hypercomplex Number Systems of 4-dimension. [Електронний ресурс] / Y. O. Kalinovsky, D. V. Lande, Y. E. Boyarinova, A.S. Turenko. ttp://arxiv.org/ftp/arxiv/papers/1409/1409.3193.pdf.

Васильев В.И., Виштальд Д.М., Любашенко Н.Д.
Национальный технический университет Украины "КПИ", Киев, Украина

Оценка стационарных характеристик надежности коммуникационных сетей

В качестве модели сети рассматривается связный стохастический граф без петель большой размерности с произвольной структурной организацией. Допускаются параллельные ребра. Заданы весовые значения параметров элементов графа - средние стационарные временные интервалы пребывания элементов в состояниях доступности и недоступности, стационарные вероятности доступности элементов (коэффициенты готовности). Требуется для любого заданного подмножества вершин графа оценить точно или приближенно вероятность связности, среднее стационарное время пребывания графа в заданном подмножестве состояний.

Пусть $\{1, 2, ..., n\}$ - пронумерованные элементы стохастического графа. Предполагается, что эволюция элементов графа во времени моделируется независимыми альтернирующими процессами восстановления (θ_i, ξ_i) с конечными математическимми ожиданиями $M\theta_i < \infty, M\xi_i < \infty$, где θ_i - случайные интервалы доступности i-го элемента, а ξ_i - случайные интервалы недоступности i-го элемента.

Эволюция графа, относительно рассматриваемого свойства, моделируется соответствующим альтернирующим процессом восстановления $(\theta_\varphi, \xi_\varphi)$, где θ_φ - случайный интервал времени, когда граф обладает заданным свойством, а ξ_φ - случайный интервал времени, когда граф не обладает заданным свойством, φ - это структурная функция, соответствующая рассматриваемому свойству.

Распределения всех случайных величин, характеризующих поведение стохастическго графа, не являются арифметическими.

Распределение F называют арифметическим (решетчатым) [1], если существует $\lambda > 0$, такое, что любая точка роста F кратна λ, то есть имеет вид $n \cdot \lambda$, где n - целое число.

Если такого числа λ не существует, то распределение F называется неарифметическим (нерешетчатым).

Точка x называется точкой роста функции распределения F, если для любых чисел a и b таких, что $a < x < b$: $F(b) - F(a) > 0$.

Граф рассматривается как носитель определенного набора свойств (предикатов), например, граф обладает свойством $(s\text{-}t)$-связности, если существует, по крайней мере, один простой путь, соединяющий вершины s и t; граф обладает свойством полносвязности, если существует, по крайней мере, одно остовное дерево; граф обладает свойством полносвязности, если две любые пары вершин связны.

Если свойство имеет структурную интерпретацию, то его можно заменить соответствующей структурной функцией. Основанием для такой замены служат законы Буля [2]. Первый из них отвечает на вопрос, когда свойства образуют булеву алгебру?

I закон Буля: Пусть свойства обозначены буквами, слова "и", "или", "не" символами "$\wedge$","$\vee$", "",а высказывание "из x следует y" - как $x \leqslant y$. Тогда свойства образуют булеву алгебру.

II закон Буля: Соответствие $x \to [x]$ между свойством и классом объектов, обладающих данным свойством, является изоморфизмом между булевой алгеброй свойств и булевой алгеброй классов: $x \wedge y = [x] \cap [y]$; $x \vee y = [x] \cup [y]$; $x' = [x]'$.

В силу этих законов многие задачи комбинаторной логики свойств и классов могут быть получены алгебраическим путем.

Для решения таких задач требуется находить семейство минимальных подграфов графа, обладающих заданным свойством Q. Подграф $\tilde{G}$ графа G называется минимальным подграфом [3] по отношению к некоторому свойству Q, если $\tilde{G}$ обладает свойством Q и не является собственным подграфом никакого другого подграфа графа G, обладающего свойством Q.

Если свойство Q соответствует высказыванию "граф G связен", то в качестве минимальных носителей свойства Q выступают остовные деревья графа.

Если свойство Q' соответствует высказыванию "граф G не связен", то в качестве минимальных носителей свойства Q' выступают минимальные разрезающие множества графа G.

Если свойство P соответствует высказыванию "граф G $(s-t)$-связен", то в качестве минимальных носителей свойства P выступают минимальные $(s-t)$-пути.

Если свойство P' соответствует высказыванию "неверно, что граф G $(s-t)$-связен", то в качестве минимальных носителей свойства P' выступают минимальные $(s-t)$-сечения графа.

Основная проблема, с которой приходится сталкиваться при оценке вероятности связности произвольного подмножества вершин стохастического графа - трудновычислимость.

Основные положения, которые используются для снижения вычислительной трудоемкости метода:

- изоморфизм алгебры свойств и алгебры классов, что позволяет работать не с отдельными состояниями, а с классами состояний;
- двойственность и минимизация сложности представления структурной функции, соответствующей рассматриваемому отношению связности;
- ортогонализация структурной функции;
- алгоритмы, представляющие структурную функцию по заданному набору свойств.

В результате получаем вычислительную схему с минимальной вычислительной трудоемкостью, позволяющую проводить оценку и анализ вероятности связности произвольного подмножества вершин стохастического графа.

В случае графа чрезмерно большой размерности на основе детерминированных и вероятностных свойств детерминированных структур [4], [5] конструктивным образом формируются оценки "снизу" и "сверху" с минимальной вычислительной трудоемкостью.

Знание вычислительной схемы для нахождения вероятности связности и знание вероятностей существования элементов позволяет осуществлять мониторинг доступности сетевой услуги в процессе эксплуатации, что дает возможность своевременно принимать меры по развитию сети, оценивать важность как отдельных элементов так и групп элементов при решении задач оптимального распределения ресурсов с целью максимизации вероятности доступности сетевой услуги.

Литература.

. **1.** Климов Г.П. Теория вероятности и математическая статистика.- Москва: Изд-во МГУ, 1983. - 328 с.

2. Зайченко Ю.П., Васильев В.И., Вишталь Д.М., Гвоздев В.С. Булев метод исследования свойств сети связи в классе бинарных стохастических моделей //Матеріали Четвертої Міжнародної науково-технічної конференції і Другої студентської науково-технічної конференції "Проблеми телекомунікацій" (20-23 квітня 2010 р., м. Київ, Україна). - 2010, с. 55.

3. М. Свами, К. Тхуласираман. Графы, сети и алгоритмы. - Москва: Изд-во "Мир", 1984.- 444 с.

4. Васильев В.И., Вишталь Д.М., Гвоздев В.С., Галушко М.Е. Детерминированные и вероятностные свойства монотонных бинарных структур в задачах анализа надежности сетей связи //Матеріали XII Міжнародної науково-технічної конференції "Системний аналіз та інформаційні технології" (25-29 травня 2010 р., м. Київ, Україна). - 2010, с. 57.

5. Васильев В.И., Вишталь Д.М., Гвоздев В.С. Точные и интервальные оценки вероятности связности сетевых структур // Матеріали Міжнародної науково-технічної конференції "Системний аналіз та інформаційні технології" (23-28 травня 2011 р., м. Київ, Україна). - 2011, с. 66.

Винокуров Д.С.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Деякі підходи до модифікації методу Динамічного програмування Беллмана

Метод динамічного програмування Беллмана - один з підходів до розв'язання задачі оптимального керування, що базується на принципі оптимальності Беллмана. В доповіді розглянуто деякі існуючі підходи до модифікації методу Динамічного програмування, запропоновано власну модифікацію методу, що дозволяє розв'язувати стохастичну задачу оптимізації.

Принцип оптимальності Беллмана. Принцип оптимальності формулюється наступним чином: "Будь-який оптимальний розв'язок складної задачі має наступну властивість: незалежно від початкового стану та першого кроку всі наступні кроки мають бути оптимальними".

Метод Динамічного програмування. Необхідною умовою для застосування методу Динамічного програмування для розв'язу складної задачі оптимального керування є можливість розбиття задачі на набір незалежних між собою підзадач. Кожна з таких підзадач описується можливими початковими і кінцевими станами та можливістю і складністю переходів між ними. Таким чином, єдиним початковим станом першої підзадачі є вихідні умови всієї задачі, а єдиним кінцевим станом останньої підзадачі відповідно є кінцевий стан всієї задачі. Задача Динамічного програмування є задачею оптимізації, а отже, еквівалентна задачі максимізації деякого функціоналу. Зазвичай такий функціонал - зростаюча функція від складності переходів з початкового стану в кінцевий.

Базова версія методу Динамічного програмування. Для більшості рельних задач метод Динамічного програмування застосовують у найпростішому вигляді: складність будь-якого переходу між початковим та кінцевим станами є детермінованими величинами та кількість проміжних станів є скінченою величиною.

Стохастична модифікація методу. Реальні системи в абсолютній більшості є недетермінованими, а отже моделі таких систем можуть бути покращені завдяки використанню стохастичних величин в якості параметрів. Для стохастичного моделювання систем за допомогою методу Динамічного програмування зазвичай в якості складності переходу між двома станами використовують випадкову величину. Таким чином, задача оптимізації зводиться до задачі максимізації функціоналу від випадкової величини.

Модифікація з обмеженнями в проміжних станах. Іноді трапляються задачі динамічного програмування, в яких кількість проміжних станів є достатньо високою, при цьому є достатньо багато можливостей переходу з початкових станів у кінцеві на більшості кроків. Прикладом такої задачі є задача прокладання маршруту в повітрі. Такі задачі є достатньо складними для обчислення і іноді виникає потреба спростити складність розв'язку таких задач. Тоді для спрощення використовують ітераційний підхід: спочатку для кожної підзадачі обирають підмножину кінцевих та початкових станів таким чином, щоб для послідовних підзадач відповідні множини мали велику кількість переходів між собою. Далі задача Динамічного програмування розв'язується для підмножин проміжкових станів. У випадку якщо оптимальний розв'язок для якогось з проміжкових станів знаходиться на межі обраної підмножини проміжкових станів, підмножина проміжкових станів зсувається в сторону обраного розв'язку з попередньої ітерації. Таким чином, данна модифікація методу динамічного програмування дозволяє значно скоротити обсяг обчислень для знаходження оптимального розв'язку задачі. Слід зазначити, що розв'язок, знайдений за допомогою такої модифікації, може бути локальним, а не глобальним максимумом досліджуваного функціоналу.

Віжевський П.В., Чертов О.Р.
Національний технічний університет України "КПІ", Київ, Україна

Порівняльний аналіз моделей швидкості осадження частинок бруду під час очистки води

Гравітаційне осадження широко визнається як ключовий механізм видалення часток бруду з товщі води у водних середовищах, таких як озера, дельти річок, установки очистки води [1,2]. Під час даного процесу малі частинки бруду агрегують з утворенням більших *флоків*. Цей механізм агрегації змінює процес осадження, об'єднуючи маленькі, повільно тонучі частинки у великі, швидко тонучі флоки [3].

Швидкість осадження визначається як функція від розміру, форми та щільності флока [4]:

$$V_f = \sqrt{\frac{4}{3}\theta g C_d^{-1}\frac{\Delta\rho}{\rho_w}D_f}, \tag{1}$$

де θ – безрозмірний фактор форми частинки, g – прискорення вільного падіння (м/с^2), C_d – безрозмірний коефіцієнт опору, $\Delta\rho = \rho_f - \rho_w$ – ефективна (надлишкова) густина флока, D_f – еквівалентний сферичний діаметр флока (м), ρ_f і ρ_w – густина флока і води відповідно (кг/м^3). Коефіцієнт опору може бути отриманий з [5]:

$$C_d = \frac{24}{Re}(1 + 0,15Re^{0,687}), \tag{2}$$

$$Re = V_s D_f/\nu, \tag{3}$$

де Re – *число Рейнольдса* флока, V_s – швидкість осадження первинних частинок, ν – кінематична в'язкість рідини. Швидкість осадження первинних частинок описується законом Стокса:

$$V_s = \frac{2}{9}\frac{(\rho_p - \rho_w)}{\mu} g\, D_f^2, \tag{4}$$

де ρ_p – густина частинки, μ – динамічна в'язкість рідини.

У роботі були розглянуті наступні моделі ефективної густини:

- модель Лау та Крішнаппана [6],
- модель Краненбурга [7],
- модель Кхеліфа [8],
- модель МакКейва [9].

Модель Лау та Крішнаппана:

$$\Delta\rho = \rho_s exp(-0,02D_f^{1,85}), \tag{5}$$

де ρ_s – густина первинних частинок флока (г/см^3).

Модель Краненбурга:

$$\Delta\rho \propto (\rho_s - \rho_w)(\frac{D_f}{d})^{F-3} \tag{6}$$

де ρ_s – густина первинних частинок флока (г/см^3), F – фрактальна розмірність флока, d – діаметр первинних частинок, що формують флок. Припускається, що відмінностями первинних частинок можна знехтувати.

Модель МакКейва:

$$\Delta\rho = \begin{cases} 1, & D_f \leqslant 1\mu\text{м} \\ \propto D_f^{-0,42}, & 1\mu\text{м} \leqslant D_f \leqslant 50\mu\text{м} \\ \propto D_f^{-1,3}, & 50\mu\text{м} \leqslant D_f \leqslant 1200\mu\text{м} \\ 0,003, & D_f \geqslant 1200\mu\text{м} \end{cases}. \tag{7}$$

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Модель Кхеліфа:

$$\Delta\rho = (\rho_s - \rho_w)(\frac{D_f}{d_{50}})^{F-3}\phi, \tag{8}$$

де ρ_s – густина первинних частинок флока (г/см^3), F – фрактальна розмірність флока, d – діаметр первинних частинок, що формуюють флок, ϕ – ефект розподілу розмірів первинних частинок, що формуюють флок:

$$\phi = \frac{\sum_{i=1}^{k} d_i^3}{\left(\sum_{i=1}^{k} d_i^F\right)^{3/F}}, \tag{9}$$

де d_i – діаметр первинної частинки флока i, k – кількість первинних частинок, що формуюють флок.

Наведені моделі ефективної густини є функціями регресії спостережень і, зазвичай, покривають вузький діапазон розмірів флоків та специфічні умови їх утворення. Лише модель Кхеліфа та модель МакКейва оцінюють швидкість осадження на широкому діапазоні значень розмірів (модель МакКейва ефективна при розмірах до 1000 μм).

Різну фрактальну розмірність враховують лише моделі Кхеліфа та Краненбурга. Цей фактор є дуже важливим, адже есперементально була підтверджена наявність флоків з різною фрактальною розмірністю [10], яка, в свою чергу, напряму залежить від кількості первинних часток, що формуюють флок. Проте усі моделі прогнозують швидкість вищою, ніж насправді.

Таким чином, для побудови моделі швидкості осадження частинок найбільш адекватною наразі є модель Кхеліфа. Варто також зазначити, що ця модель є чутливою до параметра d і навпаки – параметр ϕ демонструє слабкий вплив на результати моделювання.

Література. **1.** W. H. McAnally and A. J. Mehta, "Aggregation rate of fine sediment", *J. Hydraul. Eng.*, vol. 126, no. 12, pp. 883–892, Dec. 2000. **2.** J.M. Foxa, P.S. Hilla, T.G. Milliganb, A.S. Ogstonc, A. Boldrin, "Floc fraction in the waters of the Po River prodelta", *Continental Shelf Research*, vol. 24, pp. 1699–1715, Oct. 2004. **3.** W.R. Geyer, P.S. Hill, G.C. Kineke, "The Transport and Dispersal of Sediment by Buoyant Coastal Flows", *Continental Shelf Research*, vol. 24, pp. 927-949, 2004. **4.** J. C. Winterwerp, "A simple model for turbulence induced flocculation of cohesive sediment", *J. Hydraul. Res.*, vol. 36, pp. 309-326, 1998. **5.** A. J. Raudkivi, *Loose Boundary Hydraulics* 2nd ed. New York: Pergamon, 1976. **6.** Y.L. Lau and B.G. Krishnappan, "Measurement of Size Distribution of Settling Flocs", *NWRI Publication*,No. 97-223, 1997. **7.** C. Kranenburg, "Effects of Floc Strength on Vis- cosity and Deposition of Cohesive Sediment Suspensions", *Continental Shelf Res.*, vol. 19, pp. 1665–1680, 1999. **8.** A. Khelifa and P. S. Hill, "Models for effective density and settling velocity of flocs", *J. Hydraul. Res.*, vol. 44, no. 3, pp. 390–401, 2006. **9.** I.N. McCave and P. S. Hill, "Vertical Flux of Particles in the Ocean", *Deep Sea Research and Oceanographic Abstracts*, vol. 22, no. 7, pp. 491–502, 1975. **10.** X. Li, B.E. Logan, "Size distributions and fractal properties of particles during a simulated phytoplankton bloom in a mesocosm", *Deep Sea Research Part II: Topical Studies in Oceanography*, vol. 42, no. 1, pp. 125–138, 1995.

Візнюк О.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Ідентифікація динамічних систем прямими методами оптимізації в умовах пасивних експериментів

Вступ. Одним з підходів до розв'язку задач ідентифікації є використання теорії лінійної алгебри, але у випадку багатовимірних систем число обумовленості матриць, що обертаються, зростає неприйнятно швидко, що ускладнює процедуру розв'язку та не дозволяє одержати необхідну точність. Інший підхід базується на теорії математичної статистики. Однак неточність вхідних даних призводить до того, що задача ідентифікації виявляється некоректно поставленою. Тому на практиці алгоритми ідентифікації містять процедури регуляризації, які встановлюють обмеження розмірності апроксимуючої моделі порівняно з реальною моделлю [1].Проте поява високопродуктивних комп'ютерів та нових чисельних методів дозволяє повернутися до основ теорії та не зводити некоректну задачу до мінімізації регуляризуючих функціоналів.

Попередні дослідження. В роботі [2] було досліджено метод диференційної еволюції для ідентифікації лінійних динамічних систем у просторі стану на основі даних активних експериментів. Вхідний сигнал, вихідний сигнал та шуми було змодельовано програмними засобами. Результати ідентифікації даним методом засвідчили його спроможність знаходити коректні розв'язки задач ідентифікації та редукції порядку моделі. Пропонується дослідити роботу методу диференційної еволюції для ідентифікації системи «Гнучка рука робота» [3], тобто в умовах пасивних експериментів.

Опис досліджуваної системи. Механізм «Руки» встановлено на електричному моторі. Авторами [3] було змодельовано функцію передачі від виміряного реактивного обертового моменту бази до прискорення гнучкої руки. На вхід системи було подано періодичний синусоїдальний сигнал.

Підхід до розв'язання задачі. Вхідний сигнал, заданий у вигляді послідовності Nдискретних вимірів, апроксимується ступінчатою функцією $u(t)$

$$\forall t \in [t_i, t_{i+1}], i = \overrightarrow{0, N-1}$$
$$u(t) = u_i$$

Ідентифікація системи полягає у визначенні параметрів A, B, C шляхом мінімізації за θ квадрата норми різниці між теоретичним $\vec{y}(\theta)$ та експериментальним $\vec{y}^*$ векторами вихідного сигналу:

$$\|\vec{y}(\theta) - \vec{y}^*\|^2 = \frac{1}{N}\sum_{i=0}^{N-1}(C\int_{t_0}^{t_i} e^{A(t_i-\tau)}Bu(\tau)d\tau - y^*(t_i))^2, \quad (1)$$

де $\theta \in D$ визначає параметри моделі A, B, C з допустимої множини D. Апроксимація вхідного сигналу ступінчатою функцією дозволяє обчислити інтеграл у формулі 1 аналітично. Процедура мінімізації реалізована за допомогою методу диференційної еволюції [4], придатного для знаходження глобального екстремуму мультимодальних функцій багатьох змінних.

Література. **1.** Губарев В.Ф., Гуммель А.В., Кришталь А.А., Олесь В.Ю. О структурно-параметрической идентификации в условиях ограниченной неопределенности и аппроксимирующих моделях многомерных систем // Проблемы управления и информатики. – 2011. - №3. – С. 42-56. **2.** Губарев В.Ф., Визнюк Е.В., Пархомчук Д.М. Прямой вариационный метод идентификации систем и редукции порядка модели // Проблемы управления и информатики. - 2014. - №2. - С. 14-27. **3.** Data from a flexible robot arm [Електронний ресурс]/ Wouter Favoreel. - Режим доступу: http://www.esat.kuleuven.ac.be/ tokka/daisydata.html **4.** Rainer S., Price K. Differential Evolution — A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces // Journal Of Global Optimization. – 1997. - №11. – P. 341-359.

Гоблик А.В.
Київський національний університет будівництва і архітектури, Київ, Україна

Комп'ютерне моделювання поля градієнта інтенсивності тяжіння населення до об'єктів громадського обслуговування

Задача про раціональну територіально-планувальну організацію сфери громадського обслуговування населення відноситься до найважливіших задач містобудування. Її вирішення потребує вивчення закономірностей формування зон обслуговування населення на основі аналізу даних про інтенсивність культурно-побутових зв'язків, співвіднесених з певною зоною обслуговування. Виокремлення зони з підвищеною активністю культурно-побутових зв'язків має важливе науково – практичне значення, оскільки саме тут формується контингент населення, що істотно впливає на розрахункову місткість установи обслуговування та умови її розміщення в планувальній структурі міста [1].

Вивчення закономірностей формування зон обслуговування населення пов'язано з трудомістким збором та обробкою величезного статистичного матеріалу. У 1990 - х рр. групою фахівців під керівництвом проф. Дьоміна М.М. була розроблена методика та проведено унікальне за своїми масштабами експериментальне дослідження закономірностей формування зон обслуговування підприємств торгівлі у ряді великих міст України. У відповідності з методикою вивчалася інтенсивність потоків покупців з віддалених населених пунктів області у наперед визначені центральні міські універмаги, які є центрами масового тяжіння іногородніх покупців [1]. На основі таких експериментальних даних були встановлені функціональні залежності інтенсивності тяжіння населення I(r), які зв'язують, з одного боку, інтенсивність потоків покупців в торгові центри різних рівнів і, з іншого боку, дальність поїздки (r) з місця постійного проживання (рис.1).

По суті, в даній роботі М.Дьоміна, було відкрито закон інтенсивності тяжіння населення (ІТН) до об'єктів громадського обслуговування I(r).

Подальший аналіз та узагальнення експериментальних закономірностей інтенсивності потоків покупців з віддалених населених пунктів у наперед визначені центральні підприємства торгівлі обласного центру привів до висновку [2, 3], що отримані закономірності з високою точністю описуються функцією Гауса (рис.1,2). Таке узагальнення має важливе практичне значення, оскільки з'явилися нові можливості для створення ефективних методів, математичних та комп'ютерних моделей містобудівних систем та процесів, каліброваних на експериментальних результатах. Зокрема, значний інформативний потенціал для вирішення задачі про раціональну територіально-планувальну організацію сфери громадського обслуговування прихований в польових моделях інтенсивності тяжіння населення до об'єктів громадського обслуговування.

Тому метою даної роботи є побудова польових моделей ІТН до об'єктів громадського обслуговування засобами комп'ютерного моделювання з використанням ГІС- технологій та MATLAB.

Для досягнення даної мети в роботі розв'язані наступні задачі:

1) на основі функціональної залежності скалярної величини I(r) (рис.1) побудовано в MATLAB польову модель ІТН до об'єктів громадського обслуговування (рис.2), яка описує випадок однакової радіальної (по прямій лінії) просторово-часової доступності, тобто не враховує в даній задачі вплив конфігурації доріг;

2) для урахуванням реальної конфігурації дорожніх мереж та транспортної доступності населення на наступному кроці побудовано на основі методики [4] в середовищах ArcGIS та MATLAB відповідну польову модель (рис.4), кожна точка поверхні якої відображає скалярну величину ІТН до об'єктів громадського обслуговування;

3) на наступному етапі для підвищення інформативності отриманих моделей було побудоване векторне поле градієнта скалярної функції I (x,y,z), що значно покращує якість візуалізації результатів моделювання швидкості зміни величин інтенсивності тяжіння населення в певній

точці містобудівного простору до об'єкта торгівлі (рис.3).

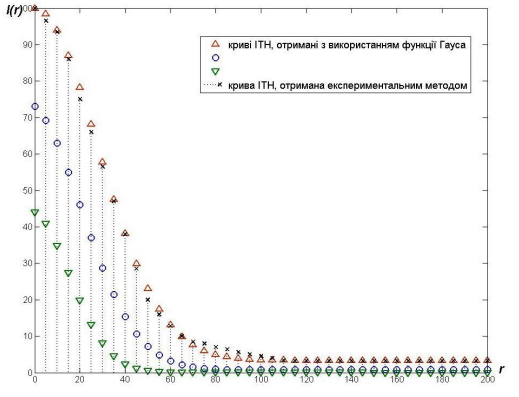

Рис. 1. Графіки функції Гауса, що описують величину I(r) – ІТН (кількість поїздок на 1 тис. жит.)

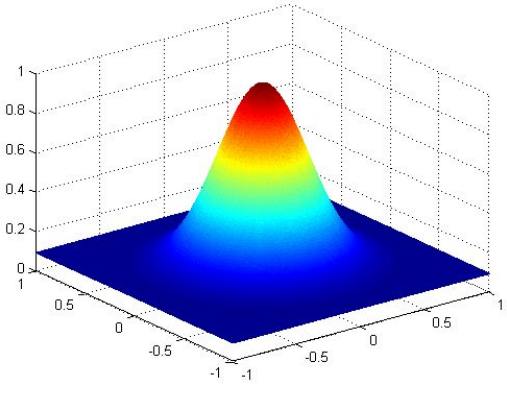

Рис. 2. Просторовий розподіл поля ІТН до центрів громадського обслуговування

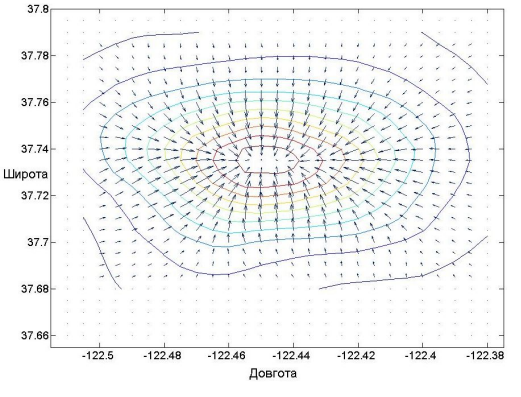

Рис. 3. Векторне поле градієнта скалярної функції I (x,y,z)

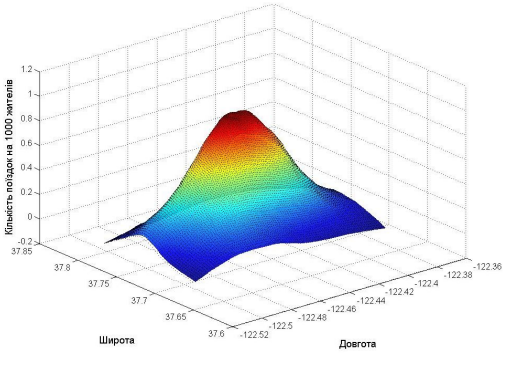

Рис. 4. Польова модель ІТН з урахуванням конфігурації дорожніх мереж

Новизна роботи полягає у тому, що вперше розроблено методику створення комп'ютерних моделей розподілу поля ІТН в середовищі ArcGIS та MATLAB на основі експериментально установленого закону інтенсивності тяжіння населення до установ громадського обслуговування.

Література. **1.** Дёмин Н.М. Управление развитием градостроительных систем / Дёмин Николай Мефодиевич. – К.: Будивэльнык, 1991. – 184 с. – Библиогр.: с. 184 – 185. – ISBN 5-7705-0310-6. **2.** Гоблик А.В. О применении методов теории потенциала в градостроительных задачах / А.В. Гоблик, Н.М. Дёмин // Містобудування та територіальне планування. – 2013. – № 50. – С. 140-154. – Бібліогр.: с.154. **3.** Hoblyk A., Gaussian law and its application in resolving urban planning problems, Space and Form, vol.22, no.1, pp. 123-128, 2014. **4.** Гоблик А.В. О моделировании распределения поля интенсивности тяготения населения к объекту общественного обслуживания / А.В. Гоблик, Н.М. Дёмин // Сучасні проблеми архітектури та містобудування. –2014. – №. 36. – С. 303-313. – Бібліогр.: с.313.

***Горбань І.М.*[1], *Хоменко О.В.*[2]**
[1]*Інститут гідромеханіки НАН України, Київ, Україна;* [2]*Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна*

Керування пристінковим потоком за допомогою інтерцептора і відбору рідини

Однією з важливих задач, які виникають при експлуатації гідродинамічних систем є керування пристінковим потоком рідини з метою зменшення гідродинамічного опору. Таке керування полягає у перетворенні турбулентної пристінної течії в систему регулярних вихрових структур. Утворені вихори називають "спійманими" або стоячими. Їх можна генерувати за допомогою створення штучних нерівностей на поверхні, що обтікається. Дослідження показують [1, 2] істотну нестійкість стоячих вихорів за наявності збурень течії, що значно ускладнює їх практичне застосування. Тому реалізація алгоритмів управління такого типу потребує розвитку схем стабілізації стоячих вихорів, які працювали б в широкому діапазоні чисел Рейнольдса. При цьому, значно більші можливості дають активні (з підводом енергії) алгоритми керування стоячими вихорами.

В роботі розвинена комплексна схема керування пристінковим потоком рідини за допомогою стоячого вихору, що генерується інтерцептором, стабілізація якого забезпечується відбором рідини. Для її обгрунтування використовується спрощена динамічна модель з одним ступенем свободи, коли циркуляційна течія в канавці замінюється точковим вихором, розташованим в центрі вихрової зони. Задача полягає у визначенні параметрів відбору рідини (потужності та місця розташування точки відбору рідини), при яких виконується умова рівноваги вихора (1) та забезпечується безвідривне обтікання кінця інтерцептора (2), тобто виконується теорема Кутта-Жуковського :

$$\frac{d\overrightarrow{X}(t)}{dt} = 0 \tag{1}$$

$$\overrightarrow{V}(z_1) = C_1, \tag{2}$$

де $\overrightarrow{X}(t) \in R^2$ – вектор координат вихора, $\overrightarrow{V}$ – швидкість течії, z_1 – координата кінця інтерцептора, C_1 – довільна константа.

Задача розв'язувалась чисельно. Одержано криві стаціонарних вихорів та розраховано відповідні параметри відбору рідини. Показано, що відбір рідини змінює топологію потоку. Аналіз поведінки вихора в околі стаціонарної точки свідчить, що у випадку, коли відбір рідини відсутній, ця точка є центром. При відборі рідини стаціонарна точка є стійким або нестійким фокусом, що залежить від положення точки на стаціонарній кривій. Одержано, що кути нахилу інтерцептора менші 20^0 є більш перспективними для керування. Розраховано діапазон значень потужності відбору рідини, що забезпечує існування стійких вихорів та оптимальне місце розташування точок стоку. Отримані результати показують, що область стійкості стоячих вихорів звужується при зсуві точки відбору рідини за потоком від закріпленого кінця інтерцептора.

Література. **1.** Gorban I.M. Dynamics of Vortices in Near-wall Flows with Irregular Boundaries// Continuous and Distributed Systems: Theory and Applications. Solid Mechanics and Its Applications / I.M. Gorban, O.V. Khomenko / M.Z. Zgurovsky, V.A. Sadovnichiy (Eds.). – 2014. – P. 115-129. **2.** Горбань І.М. Теоретичні моделі керування пристінковими потоками в гідродинамічних системах / І.М. Горбань, О.В. Хоменко// Системні дослідження та інформаційні технології.– 2014. – №4. С.87-99.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Горбань Н.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Сильний глобальний атрактор тривимірної модифікованої системи Нав'є-Стокса в необмеженій каналоподібній області

Однією з ключових задач, що виникають при дослідженні якісної поведінки розв'язків еволюційних дисипативних систем, є вивчення глобального атрактора (мінімальної компактної множини, що притягує всі траєкторії системи)(див. [1, 2]). При цьому використання сучасних методів нелінійного та багатозначного аналізу дає можливість узагальнення результатів теорії глобальних атракторів на випадок дисипативних некоректно поставлених задач (див. [3, 4]). Особливе місце серед таких задач посідає тривимірна система ріванянь Навье-Стокса (див. [5]). В даному напрямку проводилися дослідження як у випадку обмеженої області ([1]), так і в необмежених областях ([6]). Однак на сьогоднішній день для тривимірної системи рівнянь Нав'є-Стокса відомі лише вельми умовні результати в слабкій топології фазового простору (см. [7–9]). Один з методів вивчення даної проблеми полягає в аналізі різноманітних модифікацій системи з подальшим переходом до границі. (див. [10]). D ході даного дослідження була розглянута одна з таких модифікацій (вперше запропонована в [11]), що співпадає з немодифікованою тривимірною системою Нав'є-Стокса при обмежених градієнтах швидкостей. В [11]) ця задача була розглянута в обмеженій області, для досліджуваної системи було доведено існування глобального атрактора відповідної напівгрупи в сильній топології, а також показана його збіжність до множини обмежених повних траєкторій немодифікованої системи. В ході даного дослідження ці результати були узагальнені на випадок необмежених областей, що задовольняють нерівність Пуанкаре.

Тези підготовлені за результатами спільних досліджень з доктором фізико-математичних наук, професором Капустяном О.В. та Хоменко О.В.

Література. **1.** Temam R. Infinite-dimensional dynamical systems in mechanics and physics. — N.Y.: Springer, 1988. — 645 p. **2.** Chepyzhov V.V. Attractors for equations of mathematical physics / V.V. Chepyzhov, M.I. Vishik. — Rhode Island: American Mathematical Society. — 2002. — 324 p. **3.** Kapustyan O.V. Global attractors of multi-valued dynamical systems and evolution equations without uniqueness / O.V. Kapustyan, V.S. Mel'nik, J. Valero, V.V. Yasinsky. — Kyiv: Naukova Dumka, 2008. — 215 p. **4.** Evolution inclusions and variation inequalities for Earth data processing III. Long-time behavior of evolution inclusions solutions in Earth data analysis / [Zgurovsky M.Z., Kasyanov P.O., Kapustyan O.V., etc.]. — Berlin: Springer, 2012. — 340 p. **5.** Темам Р. Уравнения Навье-Стокса. Теория и численный анализ. — М.: Мир, 1981. — 408 с. **6.** Rosa R. The global attractor for the 2D Navier-Stokes flow on some unbounded domains // Nonlinear analysis. — 1998. — 32, № 1. — P. 71–85. **7.** Kapustyan O.V., Mel'nik V. S., Valero J. A weak attractor and properies of solutions for the three-dimensional Benard problem // Discrete and continuous dynamical systems. — 2007. — 18, № 2. — P. 449–481. **8.** Kapustyan O.V., Valero J. Weak and strong attractors for the 3D Navier-Stokes system // Journal of Differential Equations. — 2007. — 240, № 2. — P. 249–278. **9.** Cheskidov A., Foias C. On global attractors of the 3D Navier-Stokes equations // J. Dif. Eqs. — 2006. — 231. — P. 714–754. **10.** Constantin P. Near identity transforms for the Navier-Stokes equations // Handbook of mathematical fluid dynamics. — 2003. — II. — P. 117–141. **11.** Caraballo T., Kloeden P.E., Real J. Unique strong solution and V-attractor of a three-dimensional system of globally modified Navier-Stokes equation // Advanced Nonlinear Studies. — 2006. — 6. — P. 411–436.

Дослідження частково підтримано грантами Президента України GP/F50/049 та грантом НАН України 2273/14.

Горелова Г.В.
Институт управления в экономических, экологических и социальных системах,
Инженерно-технологическая Академия Южного федерального университета, Таганрог, Россия

Предпроектные когнитивные исследования для разработки интеллектуализированных систем поддержки принятия решений

Предложено использовать методологию когнитивного моделирования [1] на этапе предпроектных исследований для разработки интеллектуализированных систем поддержки принятия решений в различных предметных областях.

Проектирование интеллектуальных систем (ИС) требует проведения предварительных исследований как предметной области, так и соответствующих им средств исследования и принятия решений. Т.е. при проектировании ИС необходим предпроектный этап разработки и на этом этапе исследователю/эксперту также необходима система поддержки принятия решений. Проведенный анализ проблем предпроектных исследований и проектирования ИС привел к заключению о возможности применения в этих исследованиях когнитивного подхода [1, 2] и последующего построения баз знаний ИС на основе применения методологии когнитивного моделирования сложных систем, таких, как социально-экономические, экологические, политические, социотехнические и т.п.

Заметим, что в технологиях принятия решений интеллектуальная система – это информационно-вычислительная система с интеллектуальной поддержкой, решающая задачи без участия человека – лица, принимающего решение (ЛПР), в отличие от *интеллектуализированной системы*, в которой он присутствует. В нашем случае речь идет именно об интеллектуализированной системе, поскольку в процессе исследований оператор исследователь (ЛПР) обязательно присутствует, используя когнитивный инструментарий для разработки, оценки и обоснования своих решений. Итогом его работы является когнитивная модель сложной системы, которая определяет структуру базы знаний ИС.

В работах [2, 3] представлены варианты схемы модулей базы знаний ИС, которая взаимодействует с объектом управления (ОУ). Основными модулями являются: модуль диагностики состояния ОУ (разрабатывается и/или корректируется когнитивная модель объекта), модуль прогнозирования состояния ОУ (операции когнитивного анализа, позволяющие исследовать устойчивость системы, ее связность, сложность, проводить сценарный анализ на основе импульсного моделирования возможного развития ситуаций в системе и др.), модуль формирования управляющих воздействий (операции когнитивного синтеза, генерация альтернатив), модуль принятия решений (моделирование задач принятия решений, выбор решений). Через обратную связь с ЛПР управляющие воздействия направляются на объект.

Практически идея использования возможностей когнитивного моделирования сложных систем была реализована в разработке интеллектуальной системы поддержки принятия решений управления транзитной торговлей [4].

Литература. **1.** Горелова Г.В., Захарова Е.Н., Радченко С.А. Исследование слабоструктурированных проблем социально-экономических систем: когнитивный подход. – Ростов н/Д: Изд-во РГУ, 2006. – 332 с. **2.** Горелова Г.В. О разработке интеллектуализированных систем поддержки принятия решений на основе когнитивного моделирования / В сб. трудов 10-й Всероссийской научно-практической конференции “Перспективные системы и задачи управления”, Т.II – Ростов-на-Дону: изд-во ЮФУ, 2015. – С.255–265. **3.** Горелова Г.В., Мельник Э.В., Коровин Я.С. Проектирование интеллектуальных распределенных информационно-управляющих систем / Тр. Междун. научно-техн. мультиконф. “Актуальные проблемы информационно-компьютерных технологий, мехатроники и робототехники; Многопроцессорные вычислительные и управляющие системы”. – Таганрог: Изд. ТТИ ЮФУ, 2009. – Т.2. – С.28–31. **4.** Горелова Г.В., Хлебникова А.И. Когнитивное моделирование для интеллектуальной системы поддержки принятия решений управления транзитной торговлей // Научно-теоретический журнал “Искусственный интеллект” 3’2010 / Национальная академия наук Украины, Институт проблем искусственного интеллекта. Донецк, Украина: издательство “Наука и образование”, 2010г. – С.473–482.

Работа выполнена в рамках совместного с ННК “Інститут прикладного системного аналізу” Національного технічного університету України “КПІ” научного проекта РФФИ №301*213.01-12/2014-79 от 01.06.2014 г.: “Синтез методологий предвидения и когнитивного моделирования для разработки стратегии инновационного развития региона”, по гранту РФФИ № НК 14-01-90401\14

***Грайворонська А.М.*[1], *Ланде Д.В.*[2]**
[1] *Національний технічний університет України "КПІ", Київ, Україна;* [2] *Інститут проблем реєстрації інформації НАН України, Київ, Україна*

Дослідження інформаційних потоків, як динамічних мультиагентних систем

Сучасний інформацій простір є динамічною системою, яка складається з інформаційних елементів (документів) пов'язаних між собою за тематикою [1]. Багато інформаційних ресурсів, які публікують інформацію он-лайн, мають потокову структуру [2] (соціальні мережі, блоги, публікація наукових статей).Інтерес представляють шаблони поведінки, виявлення аномалій та трендів [3].

Документ, опублікований в інформаційному просторі, має певний вплив. Він може викликати різні види реакції: позитивні та негативні коментарі, копіювання, посилання. Отже, документ у інформаційному просторі можна розглядати як агента з певними властивостями та правилами поведінки. Тому одним з підходів до дослідження динаміки інформаційного потоку є мультиагентний підхід.

Мультиагентні моделі. В мультиагентних моделях агент – це абстрактна сутність, для якої характерна автономна поведінка на основі правил. Принциповою особливістю моделей на основі агентів є децентралізованість. В таких моделях визначаються правила поведінки окремих агентів, їх локальна взаємодія і взаємодія із середовищем [4]. Такий підхід є зручним в імітаційних моделях, коли глобальна динаміка системи дуже складна або недостатньо вивчена для безпосереднього моделювання.

Модель інформаційного потоку на основі агентів. В роботі представлена модель, у якій агенти відповідають інформаційним повідомлення (документам) з певною тематикою. Будемо розглядати еволюцію в дискретному часі. При $(t = 0)$ в системі існує лише один агент. Нові агенти можуть з'являтися в системі двома способами. По-перше в кожен момент часу існує ймовірність p_s самонародження агента. Це відповідає появі повідомлення із зовнішнього світу, наприклад публікації людиною нової статті. У випадку самонародження агент може дати посилання на існуючого агента. Другий спосіб появи агента - це копіювання існуючого, тобто операція repost.

Кожен агент народжується із початковою енергією E_0. З кожним часовим відліком енергія агента зменшується на 1. Якщо значення енергії стало меньше за 0, то агент помирає. Події, які відбуваються з агентом, змінюють значення енергії за наступними правилами:

- якщо поставили like, то енергія збільшується на 1;
- якщо поставили dislike, то енергія зменшується на 1;
- якщо зробили repost, то енергія збільшується на 2;
- якщо інший агент зробив посилання на даного агента, то енергія збільшується на 1.

Ймовірність того, що відбудеться певна подія, залежить від енергії агента. Нехай в n-ий момент часу енергія рівна E_n, тоді:

$$\mathbf{P}\{\text{при t=n поставили like}\} \equiv P^n_{like} = P_{like_0} \cdot \varphi(E_n),$$

$$\text{де } \varphi(E) = \begin{cases} \frac{E}{2(2E_0-E)}, & E \in \left[0;\ \frac{4E_0}{3}\right]; \\ 1, & E > \frac{4E_0}{3} \end{cases}$$

Аналогічно для операцій repost і dislike. Величини P_{like_0}, $P_{dislike_0}$, P_{repost_0} є параметрами моделі. Функція φ - монотонна зростаюча, тому ймовірність будь-якої події є більшою для агента з більшим значенням енергії.

Результати моделювання. Для візуалізації роботи моделі створюється спеціальний граф. На ньому зображені лінії життя агентів, а також червоними стрілками позначено посилання одного агента на іншого, зелені стрілки позначають операцію копіювання (repost). Приклад графу, як результату роботи моделі з параметрами $P_{like_0} = 0.4$, $P_{dislike_0} = 0.05$, $P_{repost_0} = 0.1$ і

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

$E_0 = 35$ зображений на Рис. 1.

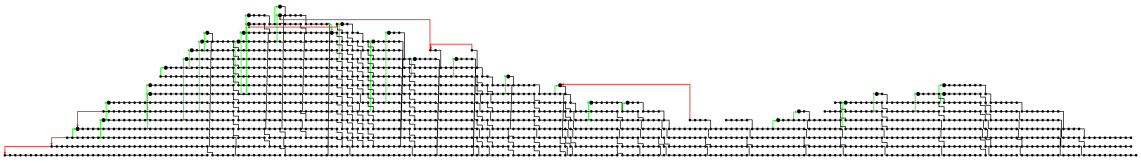

Рис. 1. Візуалізація евлюції системи агентів у вигляді графу

У вступі було зауважено, що основним об'єктом дослідження є часовий ряд. У роботі [3] обґрунтовується, що для таких часових рядів характерні декілька локальних екстремуми, тобто декілька піків популярності тематики. За допомогою даною мультиагентної моделі вдається отримати часові ряди, які відповідають описаним характеристикам (Рис. 2). Також для часових рядів, які були отримані при моделюванні, показник Херста має значення, що відбовідають високому ступеня автокореляцій в ряді.

Теоретичні характеристика одного агента. Позначимо μ_n - математичне очікування енергії агента при $t = n$. Отримані наступні рекурентні оцінки на середні значення енергії:

$$\mu_n - 1 + k_0 \frac{\mu_n}{2(2E_0 - \mu_n)} \leqslant \mu_{n+1} \leqslant \mu_n - 1 + \frac{3k_0}{4E_0}\mu_n,$$

де $k_0 = 2P_{repost_0} + P_{like_0} - P_{dislike_0}$. На рис. 3 зображені теоретичні оцінки значення енергії, та практичне середнє значення з межами відхилення. Також знайдена теоретична оцінка зверху на середній час життя агента (його середній "вік"):

$$\mathbf{M}\tau \leqslant -\frac{1}{\ln\left(1 + \frac{3k_0}{4E_0}\right)} \cdot \ln\left(1 - \frac{3k_0}{4}\right).$$

На рис. 4 зображений розподіл часу життя агентів в результаті моделювання.

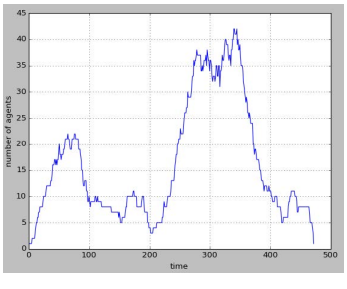

Рис. 2. Зміна кількості агентів

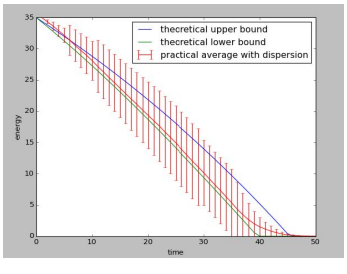

Рис. 3. Середнє значення енергії агента

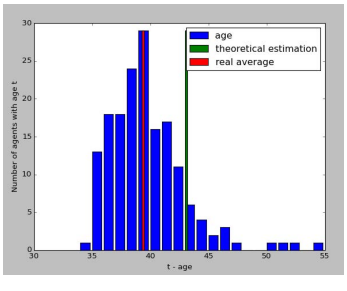

Рис. 4. Розподіл часу життя або "віку" агентів

В роботі розглянена модель тематичного інформаційного потоку на основі агентів, які відповідають опублікованим документам. В результаті моделювання отримані часові ряди, які відображають зміни кількості агентів в системі і дослідженні їх властивості. Також встановлені теоретичні оцінки для параметрів агента.

Література. **1.** В.П. Горбулін, *Інформаційні операції та безпека суспільства: загрози, протидія, моделювання*, О.Г. Додонов і Д.В. Ланде, К.: Інтертехнологія, 2009. **2.** J. Kleinberg, Temporal Dynamics of On-Line Information Streams. Garofalakis, M., Gehrke, J. and Rastogi, R., eds. Data Stream Management: Processing High-Speed Data Streams, Springer, 2006. **3.** О.Г. Додонов і Д.В. Ланде, "Мультиагентная модель поведения тематических информационных потоков," в Материалах 6-ой Всероссийской мультиконференции, с. Дивноморское, Россия, 2013, сс. 102-107. **4.** J. M. Epstein, "Remarks on the foundations of agent-based generative social science," in Handbook on Computational Economics, Volume II, K. Judd and Tesfatsion, eds. North Holland Press, 2005.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Громова Ю.А.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Особливості інноваційного розвитку України

Актуальність *інноваційної моделі* розвитку зумовлюється стрімким зростанням впливу науки та нових технологій на соціально-економічний розвиток, що відбувся протягом останніх 20-30 років. Нові технології докорінно і швидко змінили структуру світової економіки. Виявилося, що неспроможність країни здійснити структурну перебудову економіки відповідно до нового технологічного укладу (чи зволікання з цим) не просто гальмує її розвиток, але й призводить до економічної деградації і відсуває її на периферію світових економічних процесів.

Формування *інноваційної економіки* потребує значних фінансових ресурсів, тому переваги від застосування інноваційної моделі отримали перш за все багаті країни. Для економік, що розвиваються, досягнення цих переваг залежить від ефективності їх спеціальної інноваційної *політики проривного типу*, що полягає в дієвому державному стимулюванні прогресивної структурної перебудови економіки та реформуванні сфер освіти, науки, інноваційної діяльності на основі наявного науково/технічного потенціалу та з урахуванням світових тенденцій науково/технологічного розвитку.

Оскільки Україна має намір інтегруватися до європейського співтовариства, забезпечити суспільний добробут європейського рівня, то запровадження такої політики виглядає безальтернативним і на цей час має реальні передумови (інноваційний потенціал).

Відмова від інноваційної політики може призвести до консервації накопичених у країні передумов *структурної деградації* — оскільки стабілізаційне зростання останніх років (відновлення потенційного ВВП) не є достатнім свідченням позитивних довгострокових перспектив української економіки і підвищення її конкурентоспроможності у світовому економічному просторі.

Характер взаємовідносин між державою та наукою, що склалися в сучасному українському суспільстві, справедливо можна назвати суто *«українською моделлю»*. Річ у тому, що, на противагу світовій тенденції, коли країни, постійно нарощуючи свій науковий та інноваційний потенціали, покращують у результаті й інвестиційну привабливість, Україна, навпаки, — його втрачає, знижуючи, відповідно, свою інвестиційну привабливість. Як засвідчують дослідження компанії Battelle1, світові валові видатки на наукові дослідження та розробки (R&D) упродовж 2012–2013 рр. зросли на 2,7% (з 1517 млрд. дол. за паритетом купівельної спроможності, до 1558 млрд. дол.), а в 2014 р. передбачалося їх зростання, порівняно з 2013 р., ще на 3,9% (до 1618 млрд. дол.). Порівняно з докризовим 2007 р., вони зросли у 2013 р. на 38,6% (світові валові R&D видатки у 2007 р. становили 1123,9 млрд. дол.). [1]

Чинником, який додатково ускладнює означену проблему, є масштабні зміни *демографічної ситуації* в країні. Так, за дослідженнями Світового банку (Доповідь про світовий розвиток, 2013), трудові ресурси України зменшуються щорічно приблизно на 160 тис. осіб. Водночас, відзначається в Доповіді, швидке зростання міст у світі змінює структуру зайнятості населення. За оцінками її авторів, очікується, що до 2020 р. більше половини населення країн, що розвиваються, проживатиме у великих та малих містах.

В результаті цього темпи приросту робочої сили не сільськогосподарських сфер діяльності перевищуватимуть темпи приросту трудових ресурсів сільського господарства. Заодно наголошується, що в індустріально розвинених країнах означені структурні зміни відбувалися досить тривалий час (упродовж десятиліть). В сучасних же умовах світового розвитку у країнах, що розвиваються, вони стають реальністю впродовж одного покоління. Це означає, як свідчать дослідження, що структурні зміни спроможні забезпечити небувале зростання ефективності економіки, внаслідок чого деякі країни, що розвиваються, змогли цим скористатися й швидко зменшити розрив у рівнях продуктивності з індустріально розвиненими країнами. [2]

Глобалізованість світової економіки змінює й характер праці. В індустріально розвинених країнах відбувається посилення сфери послуг і наукоємних сфер діяльності та послаблення

добувної і традиційної обробної промисловості на продуктивність господарського комплексу.

Питома вага сукупних валових R&D видатків названих країн у 2012 р. становила понад 78,6% зазначеного світового показника та 79,0% у 2013 р. Така тенденція збереглася і в 2014 р. Питома вага валових R&D видатків десяти країн-лідерів у світовому показнику зросла і становила понад 79,2%. Втім, порівняно з 2007 р., питома вага десяти країн-лідерів за R&D інвестиціями у 2013 р. зменшилася. У 2007 р. вона становила 81,5%. Змінився й список таких країн. До нього увійшла Бразилія. У 2007 р. Бразилія перебувала на 12 місці за обсягами R&D фінансування. Характерно, що вона не класифікується високорозвиненою економікою, а вважається країною, що розвивається. Стосовно України позивним є те, що за показником R&D інвестицій вона увійшла у 2013 р. до 40 країн, сумарні R&D видатки яких становили 97,4% світового показника.

Також варто зазначити, що найбільшу питому вагу складали підприємства, які здійснювали придбання машин обладнання та програмного забезпечення, а найменшу – що здійснювали придбання інших зовнішніх знань.

Найбільша питома вага серед досліджень припадала на розробки, проте надання науково-технічних послуг займало найменше місце в структурі наукових та науково-технічних робіт. Також частка виконаних науково-технічних робіт у ВВП до 2002 р. зменшувалась; у 2003 р. відбулось її значне зростання на 13 в.п., проте дана динаміка не була стабільною. Так, починаючи з 2004 р., показник зменшувався, і лише 2009, а також 2012-2013 рр. показали зростання частки.

Удосконалення технологій і перенесення виробництв до країн, що розвиваються, приводить до скорочення робочих місць середньої кваліфікації. Відбувається подрібнення виробничих завдань, що дає можливість виконувати їх у різних місцях (зокрема й у малих містах). Транснаціональні корпорації створили інтегровані вартісні ланцюжки, щоб ефективно використовувати кадрові резерви різних країн. Аутсорсинг (субпідряд) впроваджується як у промисловості, так і у сфері послуг.

Зайняти українській економіці гідне місце в міжнародній інноваційній сфері стає дедалі складніше. Так, упродовж 2012—2013 рр. першу десятку країн-лідерів за валовими R&D видатками утримували одні й ті ж країни. При цьому не змінювався навіть їхній рейтинг. Окрім уже вище названих трьох країн, до зазначеного списку входили Німеччина (4), Південна Корея (5), Франція (6), Великобританія (7), Індія (8), Росія (9) та Бразилія (10). Ця ж десятка країн була лідером за показником валових R&D видатків і у 2014 р., і їхній рейтинг у цьому списку теж не змінився.

Отже, незважаючи на очевидні позитивні тенденції розвитку інноваційної сфери у світі, українська економіка демонструє протилежну закономірність. Феномен швидкого наукоємного економічного зростання поки що не спрацьовує. Україна невпинно втрачає свій інноваційний та науковий потенціали. У розрахунку на 10 тис. осіб економічно активного населення чисельність дослідників в Україні впродовж 2005–2013 рр. зменшилась у 1,2 рази.

Таким чином, нова якість світової економіки полягає в тому, що високий рівень розвитку інноваційної сфери може бути притаманним не лише високорозвиненим економікам. І, завдяки цьому феномену, країни, які не входять до клубу високорозвинених, мають можливість увійти до нього саме завдяки нарощуванню потенціалу інноваційної сфери та ефективності її функціонування.

Отож Україна, маючи відповідні інноваційну сферу і науковий потенціал, повинна постійно їх нарощувати та активніше долучатися до процесу перерозподілу світових інвестиційних потоків на свою користь. Інакше залишиться на цивілізаційних задвірках.

Література. **1.** Global R&D funding forecast, 2014 www.battelle.org/docs/tpp/2014_global_rd_funding_forecast **2.** Головатюк В. Інноваційний розвиток української економіки drozdenko.at.ua/news/2014-06-26-8982

Довгаль К.І., Чертов О.Р.
Національний технічний університет України "КПІ", Київ, Україна

Байесові мережі для моделей оцінки якості програмного забезпечення

У роботі [1] виконано огляд значної частини опублікованих емпіричних робіт в області прогнозування дефектів в індустрії розробки програмного забезпечення (ПЗ). Він показав наявність великої кількості методологічних та теоретичних помилок у цих роботах. Так, багато останніх досліджень страждали від різних недоліків, починаючи від помилкок у специфікації до використання невідповідних даних.

Для оцінки якості вихідного коду програмного забезпечення у роботах [2, 3] запропоновано використовувати байесові мережі. Встановлено, що даний спосіб моделювання варто використовувати у випадках, коли необхідно приймати рішення в умовах невизначеності. На математичній основі байесівської ймовірності можна сформулювати експертні уявлення про взаємозалежність між різними змінними та послідовно, через мережу, поширити вплив вихідних метрик на результат та рівень його вірогідності.

Для побудови байесової мережі при створення математичної моделі оцінки якості вихідного коду програмного забезпечення запропонований спосіб [2], який походить від моделі якості на основі активностей (Activity-based quality model, ABQM). Результуюча байесова мережа складається з трьох типів вузлів:

- вузли активностей, які представляють активності на ABQM;
- вузли фактів, що відповідають фактам на ABQM;
- вузли індикаторів, що дають метрики для активності чи факту.

Для побудови вузлів, що містять інформацію щодо моделі якості на основі активностей, необхідно виконати три кроки.

Крок 1. Визначення активностей, виходячи з оцінки та прогнозування кінцевої мети.

Спочатку встановлюється мета моделювання - планування потреб у ресурсах на основі оцінки якості вихідного коду програмного забезпечення. Така мета вимагає відповідної активності (діяльності) - супроводження вихідного коду ПЗ. Ця активність уточнюється та деталізується шляхом постановки запитань та отримання відповідей на них задля досягнення мети.

Крок 2. Ідентифікація впливових підактивностей та фактів.

Тут можливі до розгляду два варіанти:

1. підактивність вже ідентифікованої активності;
2. вплив фактів на ідентифіковану активність.

Цей крок необхідно повторювати рекурсивно для усіх підактивностей до тих пір, поки усі факти, що мають вплив на піддерево кожної з ідентифікованих підактивностей, не будуть визначені. Відразу ж стає видимим вплив на кожну активність, а, отже, і відповідні факти. Всі активності і факти, виявлені таким чином, представляються у вигляді вузлів у байесовій мережі. Ребра додаються від підактивностей до відповідних активностей, що знаходяться вище у ієрархії, та від фактів до активностей, на які вони мають вплив. Факти, що отримуються у результаті, в подальшому моделюються разом з кількісною величиною їхнього впливу.

Крок 3. Додати додаткові вузли до фактів у вигляді індикаторів для кожного факту та вузла активності, що потребують вимірювань.

Незалежно від того, чи визначені індикатори для активностей на першому чи на другому кроці, при необхідності, можуть застосовуватися додаткові індикатори для підактивностей. У будь-якому випадку, для кожного факту повинен бути принаймні один індикатор, що використовується при моделюванні. Це може бути вичерпна характеристика, яка лежить у межах самої моделі оцінки якості, самостійно побудована або запозичена з літератури вже існуюча метрика. Індикатор не повинен обов'язково вимірюватися автоматично, ручні викладки також можуть бути включеними в модель. Ребра направляються від активностей і фактів до індикаторів, тобто індикатори залежать від фактів та активностей, оскільки індикатори є лише вираженням фактору, що лежить в основі вершини, яка описується. Головна перевага

використання ABQM, як способу побудови байесової мережі, - це наявність топологічного опису. Додатковою перевагою таких моделей є якісний опис зв'язків між різними факторами, що мають відношення до якості програмного забезпечення. Припускається, що всі залежності змодельовані, а усі інші фактори є незалежними.

На рис. 1 зображена топологія байесової мережі для симуляції моделі оцінки якості ПЗ, що побудована слідуючи цим крокам. При визначенні індикаторів для вузлів фактів використовуються метрики (цикломатична складність, LCOM4 та ін.) [3–5] вихідного коду програмного забезпечення, розрахунок яких відбувається статично, тобто без запуску програми.

На даному етапі байесова мережа заповнюється кількісною інформацією. Стан кожного вузла визначається таблицею вірогідності на основі поширеного в індустрії розподілу чи результатами внутрішньокорпоративного вимірювання. Подальші дослідження варто спрямувати на оцінку та агрегацію шкал вузлів. Зазвичай, вони ранжуються за порядковою шкалою. Найпоширеніша шкала складається зі значень *малий*, *середній* та *великий*. Хоча ця шкала і має перевагу у вигляді простоти, але перспективними є інші, наприклад, неперервні шкали.

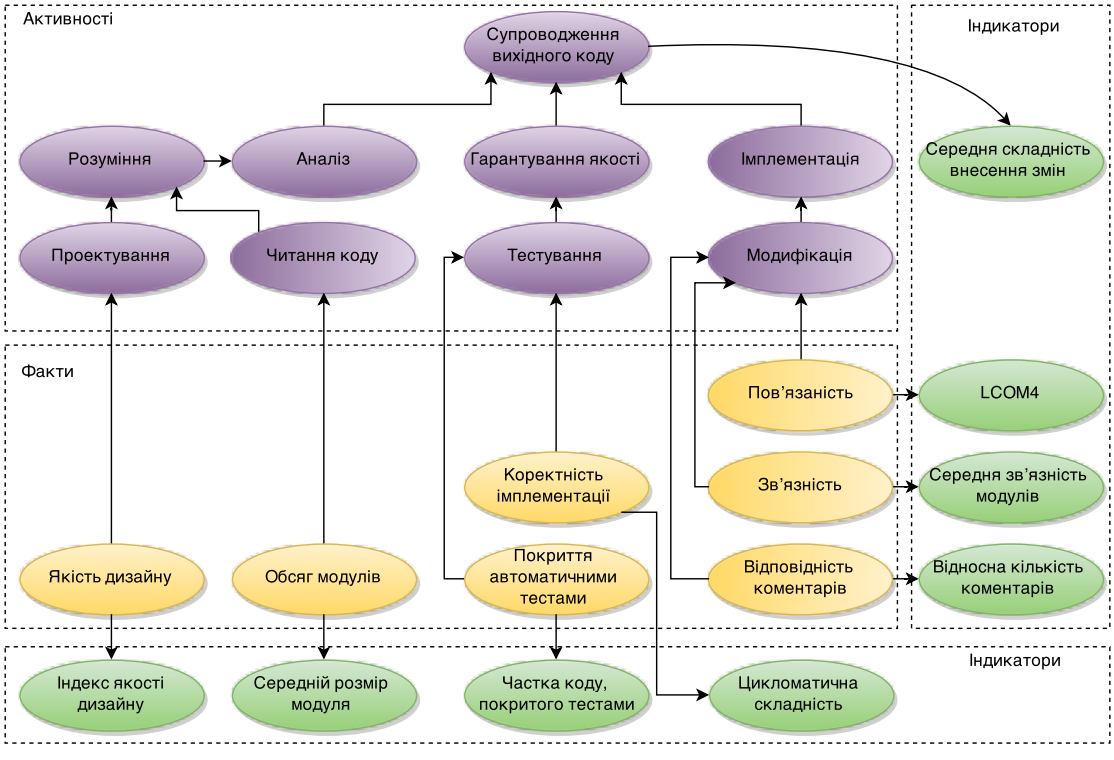

Рис. 1. Топологія байесової мережі для оцінки якості ПЗ

Література. **1.** N. E. Fenton and M. Neil, "A Critique of Software Defect Prediction Models" in IEEE Transactions on software engineering, vol. 25, № 3, 1999, pp. 675-689. **2.** S. Wagner, "A Bayesian network approach to assess and predict software quality using activity-based quality models" in Information and Software Technology, vol. 52, № 11, 2010, pp. 1230–1241. **3.** N. E. Fenton and M. Neil, "Software metrics: roadmap" in ICSE '00 Proceedings of the Conference on The Future of Software Engineering, 2000, pp. 357-370. **4.** F. Fioravanti et al., "Metrics for Controlling Effort During Adaptive Maintencance of Object Oriented Systems" in Proceedings IEEE International Conference on Software Maintenance, 1999, pp. 483-492 **5.** Cohesion metrics. [Online]. Available: http://www.aivosto.com/project/help/pm-oo-cohesion.html

Донченко В.С., Назарага І.М., Тарасова О.В.
Київський національний університет ім. Тараса Шевченка, Київ, Україна

Регресії з матричними представниками : теорія та приклади застосування

Задачі групування інформації (ГіП): класифікація - кластеризація та розпізнавання образів, а також відновлення залежностей, представлених своїми значеннями, - представляють собою фундаментальний клас прикладних задач. Задача відновлення залежностей може розглядатися у детермінованому та статистичному варіантах, але в обох базовим методом розв'язку - оцінювання є метод найменших квадратів за векторних аргументів та скалярних відгуків. В той же час в багатьох прикладних задачах природним є використання ширших за типом складових спостережень: матричних. Таке використання дозволяє зберегти, а отже, - і використати, - структурну інформативність об'єктів. Таке розширення можливостей розв'язання ГіП - задач вимагає створення відповідного апарату оперування із матричними об'єктами як із векторами евклідового простору числових векторів R^n. Нагадаємо, що основою апарату оперування із базовими лінійними та нелінійними структурами в R^n є апарат SVD та ПдО (псевдообернення) за Муром-Пенроузом (див., наприклад, [1]). В роботі [2] висунута та реалізована концепція кортежних операторів, яка дозволила перенести на матричні евклідові простори всі переваги використання базових структур в R^nна основі SVD та ПдО.

Приклади застосування для розпізнавання див., наприклад в [3]. В цій роботі представлені "матричні" можливості використання в ГіП - відновленні залежностей для прогнозу на прикладі макроекономічних та медіа показників. Обґрунтування та необхідні теоретичні результати можна знайти в [4] (див. також [5]).

У макроекономічному варіанті запропонований підхід застосовано до показників внутрішнього валового доходу (GDP), заробітної плати працівників (WE), витрат на кінцеве споживання (FCE), експорту товарів та послуг (E) та імпорту товарів та послуг (I). У досліджені використано дані 2007 року – 3 кварталу 2013 року Держкомстату та Міністерства розвитку економіки і торгівлі України. Для оцінки точності отриманого прогнозу використано критерій APE. У відповідності до даних матриці APE максимальне значення похибки для GDP становить 6,98%, для заробітної плати WE 12,02%, для FCE 7,74%, для експорту 13,2%, для імпорту 15,66%. Перевищення 10% значення похибки спричинено використанням при прогнозуванні даних кризових років 2008-2009.

У телемедійному прикладі використано дані спостережень 2003-2013 років з помісячною дискретизацією. Прогноз будується для 5 показників: рейтинг загального теледивлення (T), частка аудиторії каналу (S), рейтинг каналу (R), рейтинг реклами (Ra) та коефіцієнт усадки рекламних блоків каналу (K); використано метод попереднього лінійного та нелінійного масштабування даних у вигляді шести моделей. всі показники мають точність прогнозування в межах 10%-інтервалу.

Література. **1.** Ben-Israel A., Greville T.N.E. Generalized Inverses: Theory and Applications. New York: Wiley, 1977. **2.** Donchenko V., Zinko T., Skotarenko F .'Feature vectors' in grouping information problem in applied mathematics: vectors and matrixws /Problems of Computer Intellectualization. - Editors Vitali Velichko, Aleksey Voloshin O, Krassimir Markov. - Kyiv,Ukrain – Sofia, Bulgaria: 2012 - p.111-124. **3.** Volodymyr Donchenko, Andrew Golik Matrix Feature Vectors in Speech and Gesture recognition //International Journal "Information Theories & Applications": ITHEA: Volume 20, Number 2. - 2013 . - P.143-155. **4.** Donchenko V. S., Nazaraga I.M., Tarasova O.V. Matrixes least squares method: examples of its application in macroeconomics and tv-media business // Eastern-European Journal of Enterprise Technologies. – 2014. – 4/4(70). – P. 42-46. **5.** Donchenko V. S., Nazaraga I.M., Tarasova O.V. Vectors and matrixes least square method: foundation and application examples // International Journal "Information Theories & Applications". – 2013. – Vol. 20, No 4. – P. 311–322.

Заец Р.В.
Центр исследований научно-технического потенциала и истории науки им. Г.М.Доброва НАН Украины, Киев, Украина

Концептуализация системных представлений и анализа научно-технической политики

Обсуждается концептуализация – начальный методологический этап – системных описаний и анализа научно-технической политики как сложного многопланового когнитивного образования, которое не имеет четких границ компетенции и социальной ответственности за принимаемые решения, направляющие развитие науки, техники и технологий. Намечаются задачи конституирования и проектирования новой политики производства знаний и инноваций, необходимых для построения экологически устойчивого общества.

1. Концептуализация изучения современного состояния и перспектив научно-технической политики предполагает рассмотрение методологических оснований и социальных условий её формирования и реализации. Прежде всего, это касается мировоззренческих позиций, с которых обозреваются прошлый опыт организации и поддержки науки, Status Quo и возможные альтернативы научно-технического развития, которые обусловят существенно разные варианты социально-экономического и экологического будущего общества. Обоснование научно-технической политики важно детально рассматривать в экономическом, экологическом, социальном и других контекстах, в которых формируются её характеристики, показатели и критерии её оценивания [1]. В условиях нарастания глобальных экологических уг-роз при прогнозировании и стратегическом планировании науки и технологического развития обязательным становится учёт морально-этического контекста [2–4]. При этом невозможно игнорировать критику как общественностью, так и самими учёными многих научных и технологических направлений [5]. Результатом концептуализации должен быть методологический проект, показывающий проблемное поле и задачи междисциплинарных исследований и системного проектирования политики в отношении научно-технической сферы.

2. В системной интерпретации научно-техническая политика представляет систему принятия решений, которая опосредует связи между суперсистемой “общество” и его когнитивной системой “наука”. Этот интеллектуальный интерфейс, с одной стороны, должен полно отражать актуальные и потенциальные проблемы общества, для решения которых нужны уже известные и новые научные знания и инновации. С другой, – современное состояние и когнитивные возможности научной системы, которой передаются задания на получение новых знаний. Прагматическая часть поставляемых наукой знаний необходима обществу для решения технологических, производственных, экономических, экологических, демографических, социальных, гуманитарных, политических, оборонных и др. проблем, характерных для стран, регионов, поселений и территориальных общин.

Функциональное предназначение научно-технической политики – разработка и реализация стратегии и тактик развития науки и техники, перспектив технологического обновления общества; определение структур и объёмов ресурсного обеспечения фундаментальных, поисковых и прикладных исследований и разработок, которые отражают потребности в научных знаниях разных секторов экономики и социальной жизни. В когнитивном плане важно раскрыть и открыто представить картину обоснования такой политики, показать – на базе каких знаний и фактологических данных разные правомочные субъекты формируют политические решения относительно науки и инновационной деятельности.

3. Субъектами научно-технической политики выступают: государства (правительства, представленные органами законодательной и исполнительной власти), бизнес (ТНК, ФПГ, национальные корпорации, предпринимательские структуры), политические силы (партии и движения), территориальные общины (представленные органами местного самоуправления), косвенно – творческие союзы и объединения. На формирование научно-технической политики Мир-Системы и отдельных стран сильные влияния оказывают международные организации (ООН и учреждённые ею организации), союзы государств (ЕС и его надгосударственные органы), военно-политические союзы (НАТО), различные международные и частные фонды, которые на конкурсных основаниях оказывают финансовую поддержку ученым и научным

коллективам (по заранее вызначенным направлениям исследований).

Эти акторы предъявляют к науке и инноватике разительно отличные требования и цели, задают направленность процессов научного познания, привносят в сферу НИОКР свои менеджмент и стили управления, которые отражают их интересы и мотивации. Каждый класс акторов в амбивалентной системе "наука" выделяет и усиливает лишь одну из её сторон, что искажает её гармоничное развитие. Бизнес занят милитаризацией и коммерциализацией науки, культивирует секретность НИОКР; государство бюрократизирует науку и заставляет оправдывать решения властей; политики идеологизируют и политизируют науку, используют её как средство борьбы за власть; общественность (гражданское общество) требует демократизировать науку и пытается установить гражданский контроль над планированием исследований и разработок, их содержанием, использованием научных результатов. При этом неизбежно возникают противоречия и конфликты интересов как разных акторов, заинтересованных в получении и использовании знаний, так и акторов (заказчиков НИОКР и ресурсодателей) и исполнителей научных работ.

Существует неустранимое противоречие между стремлением финансово и политически мощных акторов подчинить науку своим экономическим или политическим целям и теми интенциями развития науки, которые ей присущи как сфере творческой деятельности. В этом качестве наука старается сохранять академическую свободу, организационную автономию и самоуправление, придерживаться профессиональной культуры и этики, не присущих другим сферам деятельности. Но в сложно организованном потребительском обществе, основанном на рыночной экономике, остается всё меньше свободы для научных коллективов и отдельных ученых, которые самочинно определяют направленность и приоритеты своих исследований и разработок, и в случае востребованности полученных ими результатов, могут работать за счет самофинансирования.

4. В качестве средства системного описания и анализа научно-технической политики уже на концептуальном этапе можно применить метод структурных матриц (МСМ) [6]. В понятиях МСМ политика может быть представлена двояко:

- как активная часть внешней среды, воздействующей на систему "наука" – правая часть структурной матрицы (квадранты III и IV) (рисунок в работе [6, с.89]). В этом случае политику представляет кортеж переменных, воздействующих на подсистемы (диагональные элементы) ядра системы (науки);
- как относительно автономная когнитивная система, которая вырабатывает воздействия на систему "наука". В этом случае политика представляется как комплекс взаимосвязанных механизмов переработки информации и принятия решений, направляемых на внешнюю систему "наука".

Основные функциональные подсистемы субъекта, вырабатывающего и реализующего научно-техническую политику, можно обозначить как: рецептор (чувствительный орган); рефлектор (осознающий орган); акцептор (творческий орган); эффектор (исполнительный орган); продуктор (основной рабочий орган). Раскрыть содержание деятельности этих органов и синхронизировать их работу – задачи системного анализа и проектирования содержательно и структурно полной научно-технической политики.

Литература. **1.** Социальные показатели в системе научно-технической политики. Сборник переводов с английского. – М.: Прогресс, 1986. – 486 с. **2.** Агацци Э. Моральное измерение науки и техники. Пер. с англ. И.Борисовой. науч. ред. В.А. Лекторский. – М.: МФФ, 1998. – 344 с. **3.** Лейси Х. Свободна ли наука от ценностей? Ценности и научное понимание / Пер с англ. – М.: Логос, 2001. – 360 с. **4.** Барбур И. Этика в век технологии. – М.: ББИ им св. апостола Андрея, 2001. – 380 с. **5.** Каттон У.Р. мл. Конец техноутопии. Исследование экологических причин коллапса западной цивилизации. Пер. с англ. Под общей ред. В.И. Постникова. – К.: Экоправо-Киев, 2006. – 255 с. **6.** Заец Р.В. Метод структурных матриц как инструмент описания и анализа сложных систем // Системный анализ и информационные технологии: материалы 16-й Международной научно-технической конференции SAIT 2014, Киев, 26-30 мая 2014 г. / УНК "ИПСА" НТУУ "КПИ". – К.: УНК "ИПСА" НТУУ "КПИ", 2014. – С.88–89.

Зражевська Н.Г.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Оптимізаційна модель для оцінювання параметра Херста

Незважаючи на велику кількість методів оцінювання параметра Херста, що мають досконале теоретичне підгрунтя [1,2], при роботі з реальними даними дослідники часто зустрічаються з суттєвими проблемами. Оцінки параметра можуть сильно відрізнятися в залежності від метода, який використовувався [3]. Тому проблема вибору метода для дослідження ряду на наявність сильної залежності залишається актуальною.

У роботі було розглянуто вісім класичних методів [2] оцінювання параметру Херста та запропоновано новий підхід для знаходження параметра на основі оптимізаційної процедури. З метою налагодження техніки аналізу був розглянутий штучно змодельований по моделі FIGARCH часовий ряд з відомими характеристиками. Вказані методи були застосовані для побудови оцінок характеристик реального часового ряду РТС ("Российская Торговая Система") за 2014 год.

Для побудови АКФ використовується регресійна модель, що випливає з означення сильної залежності

$\rho(k) = \alpha_1 H^*(2H^* - 1)k^{2H^*-2} + \alpha_2 + \epsilon_k, \epsilon_k$ - н.о.р.с.в., $k_0 \leqslant k \leqslant N$

Оцінки для коефіцієнтів α_1, α_2, можуть бути знайдені по МНК. $\hat{H}_{mean}$ виступає в якості оцінки H^*. Для уточнення його значення була запропонована оптимізаційна процедура.

Для побудови задачі оптимізації розбиваємо фактичні дані на два інтервали: інтервал I ("in-sample") та інтервал II ("out-of-sample"). На інтервалі I знаходимо коефіцієнти моделі з міркувань мінімізації залишків моделі по методу найменших квадратів. На інтервалі II корегуємо параметри моделі так, щоб модель у сенсі середньо-квадратичної похибки була оптимальною.

Формально, оптимізаційна задача може бути записана у вигляді

$$\lambda q + (1 - \lambda)\frac{1}{k_2 - k_1}\sum_{k=k_1+1}^{k_2}(y_k - \hat{y}_k)^2 \to \min \quad (1)$$

$$(y_k - \hat{y}_k)^2 \leqslant \epsilon^2 + q, k = 1, 2, ...k_1 \quad (2)$$

$$q \geqslant 0 \quad (3)$$

де в інтервал I входять всі фактичні значення автокореляційної функції y_k, $k = 1, 2, ..k_1$, а в інтервал II - всі фактичні значення автокореляційної функції y_k для $k = k_1 + 1, 2, ..k_2$. Похибки $\epsilon_i = y_k - \hat{y}_k$ на інтервалі II мінімізуються в середньоквадратичному сенсі одночасно з релаксаційним параметром q, зваженими з коефіцієнтом штрафу λ. На інтервалі I похибки обмежуються похибкою агрегованої моделі з релаксаційним параметром q. Параметр λ є коефіцієнтом штрафу $0 \leqslant \lambda \leqslant 1$, що забезпечує можливість досягнення оптимального співвідношення похибок на інтервалах I і II. Отже, запропонована параметрична оптимізаційна процедура допускає налаштування параметрів для різних цілей моделювання автокореляційної функції. В результаті розв'язання отимізаційної задачі отримана оцінка $\hat{H}_{optim}$. Графіки АКФ, побудовані для $\hat{H}_{optim}$, більш точно наближають фактичні дані для старших значень аргументу.

Результати роботи можуть бути використані для побудови адаптивних робастних оцінок при прогнозуванні.

Література. **1.** W. Palma, "Long-Memory Time Series: Theory and Methods", John Wiley & Sons, Inc., New Jersey, 2007, pp. 304. **2.** M.S. Taqqu and V. Teverovsky and W. Willinger, "Estimators for long-range dependence: An empirical study", Fractals, 1995, pp.785-798. **3.** R.G. Clegg, "A practical guide to measuring the Hurst parameter", International Journal of Simulation: Systems, Science & Technology, 2006, pp.3-14.

Кірік О.Є.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Декомпозиційний підхід до аналізу великих просторово розподілених об'єктів мережевої структури

Системний підхід передбачає інтегроване дослідження об'єктів та явищ в цілому, на відміну від вивчення моделей елементів або окремих сторін тих чи інших явищ. Оскільки великі задачі виникають зазвичай в результаті врахування зв'язків, що існують між підсистемами в часі або просторі, то в них майже завжди виявляється деяка спеціальна структура. Апарат математичного програмування нараховує достатньо багато підходів для розв'язання спеціальних задач, проте інтерес до розробки нових ефективних методів не зменшується і наразі. Це особливо стосується методів розв'язання нелінійних задач великої розмірності. По-перше, нелінійні моделі часто зустрічаються у плануванні та управлінні. По-друге, область реального застосування системних підходів все ще недостатньо широка, що пояснюється, в основному, надвеликою складністю цих задач та великими затратами на їх розв'язання.

Визначилися два основні підходи до розв'язання складних задач великої розмірності. Це прямі методи та методи, що засновані на ідеях декомпозиції. У першому випадку обчислювальні труднощі долаються за допомогою різних способів прискорення алгоритмів, спеціальних засобів збереження інформації та інших прийомів. Методи, що належать до другого класу, базуються на розкладанні системи на підсистеми меншої розмірності. Виділяються підсистеми різного рівня і вибудовується ієрархія задач таким чином, що підсистеми верхнього рівня визначають відповідні зміни в підсистемах нижчого рівня.

Враховуючи, що для сучасних складних систем характерні різноманітні види ієрархій – часова, просторова, функціональна, інформаційна, а процес поділу системи на підсистеми не може бути однозначним, зростає роль кваліфікованих експертів у процесі системних досліджень складних об'єктів. Таким чином, системний підхід включає в себе математичні методи дослідження, які в достатній мірі формалізовані, а також неформальні (евристичні, інтуїтивні) процедури, прийоми і правила, що повинні застосовуватися дослідниками, експертами або особами, що приймають рішення.

Декомпозиція загальної задачі на деяку ієрархію або послідовність взаємопов'язаних підзадач, може мати двоякий характер. В одному випадку вона відповідає реальній послідовності задач, які можна виділити вже на етапі змістовної постановки вихідної загальної задачі у відповідності з реальною просторовою будовою системи або виходячи з окремих розглянутих інтервалів часу. В іншому випадку декомпозиція представляє більш формальний поділ задачі по будь-яких групах змінних або рівнянь, що значно менше враховує їх фізичний або технологічний зміст.

Як приклад, в енергетиці оптимізаційні задачі, як правило, нелінійні, динамічні, дискретні, мають великі розміри і недостатню (з формальної точки зору) інформаційну базу. З другого боку, серед багатьох змінних задач управління мережевими системами паливно-енергетичного комплексу майже завжди можна виділити ті, які характеризують структурні особливості, описують певні режими роботи систем, визначають зовнішні зв'язки або граничні умови [1].

Природним і широко вживаним прийомом є почергове фіксування різних груп змінних або зв'язків з формуванням і розв'язанням часткових підзадач. Далі проводиться перехід до координуючої задачі і т.д. У загальному випадку будуються і досліджуються складні алгоритми оптимізації у вигляді ітераційних процесів з взаємопов'язаними чисельними та логічними процедурами.

Оскільки мережеві оптимізаційні моделі для паливно-енергетичних систем є нелінійними і часто мають спеціальну (блочну) структуру обмежень, то пропонуємо розділити процес аналізу цих систем на два етапи.

На першому етапі заміняємо вихідну задачу послідовністю задач, що мають адитивно-сепарабельні цільові функції, за допомогою методів лінеаризації, послідовного квадратичного

програмування або іншої спеціальної апроксимації нелінійних залежностей [2,3]. На другому етапі піддаємо декомпозиції ці допоміжні задачі. Внутрішні кроки ітераційного процесу полягають у розв'язанні набору невеликих підзадач та застосуванні певних координаційних процедур для формування даних, що мають бути передані на верхній рівень для здійснення зовнішніх кроків процесу.

При такому підході відбувається зведення неадитивних систем до адитивних, тобто таких, що автоматично розкладаються на підсистеми меншої розмірності, що для широкого кола задач дозволяє перейти від евристичного до алгоритмічного виконання декомпозиції.

Ідея комбінації методів декомпозиції та квадратичної апроксимації нелінійних оптимізаційних моделей має особливий сенс саме для мережевих задач. По-перше, специфіка їхньої структури обмежень дозволяє отримати прийнятні за швидкістю збіжності процедури обчислень, що є критично важливим при розв'язанні великої кількості допоміжних задач. По-друге, при розв'язанні збалансованих нелінійних задач розподілу потоків допоміжні підзадачі квадратичного програмування для достатньо широких класів задач завжди мають обмежений розв'язок [4].

Відмітимо, що базове значення формалізованих методів моделювання і обґрунтування рішень як єдиного апарату вивчення систем енергетики та управління ними з часом зростає у зв'язку із зростанням складності самих систем і пов'язаних з їх розвитком і функціонуванням проблем. Зважаючи на складність досліджуваних систем необхідно гарантувати еквівалентне кваліфіковане спрощення моделей без шкоди для точності результатів. Тому процес створення більш точних і одночасно прийнятних за складністю методів логічно призводить до ієрархічних принципів аналізу, у тому числі для послідовного уточнення прийнятих рішень. При цьому важливо те, що в процесі перетворення моделей шляхом використання методів декомпозиції дослідник ґрунтується на деякій вихідній моделі, і якщо вона сформована надмірно наближено, то ніякі подальші математичні перетворення не зможуть дати змістовних результатів. Ще одним важливим моментом є співвіднесення точності результатів розрахунків з точністю використовуваної інформації. Особливо це стосується планових і проектних розрахунків, в яких доцільну точність результатів не слід мати більшою, ніж точність використовуваних вихідних показників.

Література. **1.** Системные исследования в энергетике /под ред. чл.-кор. РАН Н.И.Воропая. Новосибирск.: "НАУКА", 2010. – 686 с. **2.** Кірік О.Є. Алгоритми ітераційного квадратичного програмування для задач оптимального розподілу потоків // Системні дослідження та інформаційні технології. – 2008. - № 1. - С.101-113. **3.** Александрова В.М., Кірік О.Є. Оптимізаційні моделі та алгоритми для мережевих задач розподілу ресурсів // Наукові вісті НТУУ "КПІ". - Київ, 2014. - N 4. – 39-45. **4.** Кірік О.Є. Алгоритми лінеаризації та спряжених градієнтів для нелінійних задач розподілу потоків // Наукові вісті НТУУ "КПІ". - Київ, 2007. - N 3. – С. 67-73.

Козуля Т.В., Білова М.О.
Національний технічний університет "Харківський політехнічний інститут", Харків, Україна

Підвищення ефективності комплексної методики оцінки екологічності системних об'єктів різного рівня дослідження

Актуальність дослідження. Існуючі методики, що використовуються для розв'язання різнопланових еколого-економічних і еколого-соціальних задач, достатньо складно звести в єдину систему методичного забезпечення при потребі формування єдиного програмного комплексу для формування оцінки екологічності системного об'єкта дослідження. За вимогами сучасних потреб суспільства актуальними є питання сталого розвитку, в основі яких лежить розуміння тісного взаємозв'язку екологічних, економічних і соціальних проблем людства, необхідність багаторівневого рішення з урахуванням стану систем на глобальному, регіональному та локальному рівнях дослідження [1]. Таким чином, постає задача побудови універсальної методики комплексної оцінки екологічності складних систем на основі поєднання різних методів аналізу і кількісного оцінювання з метою надання узагальненої характеристики стану об'єкта дослідження і його відповідності вимогам стабільності та безпеки.

Постановка задачі дослідження. Запропонована система комплексної оцінки відповідності рівню екологічної безпеки природно-техногенних систем на основі компараторної ідентифікації якості в системних дослідженнях «стан системи – процес – стан змін в системі» [2] відповідає названим вище вимогам розв'язання комплексних задач, але потребує удосконалення з метою отримання практичних результатів з визначення дестабілізуючих факторів безпеки і надання інформації з регулювання якості техногенних систем. Для підвищення ефективності методичного забезпечення оцінки екологічності складних систем необхідним є створення алгоритмічно-програмного комплексу розв'язання різнопланових задач сталого розвитку.

Методика дослідження і аналіз результатів. Для отримання адекватних результатів при обробці моніторингової інформації про стан системних об'єктів важливим є усунення інформаційного шуму, що супроводжує вихідні дані для розрахункової системи комплексної оцінки екологічності [3]. У даній роботі пропонується впровадження методу головних компонент з метою вилучення необхідної для вирішення поставленої задачі інформації з усього доступного набору вихідних даних при збереженні його інформативності, подання її у більш компактній і вагомій формі для обробки та аналізу отриманих результатів [4]. Значення головних компонент за розрахунками в Statistica 6.0 [5] дозволяють використати вагомі параметри управління для стабілізації стану в аналізованій системі і оптимального регулювання структурних змін з збереженням функціональності елементів і об'єкта в цілому.

Алгоритмічне забезпечення з оцінки екологічності природно-техногенних об'єктів у зв'язку з комплексуванням методик головних компонент і компараторної ідентифікації за умови багатопланової вихідної інформації визначає послідовність усунення інформаційного шуму, встановлення узагальненої оцінки якості системного об'єкта (рис. 1).

Для прийняття рішень стосовно підвищення екологічної якості та загальної сталості функціонування природно-техногенних систем пропонується надавати аналіз вихідних характеристик об'єкта за визначеними трьома напрямками: екологічним, економічним та соціальним, тобто розділити параметри дослідження на три групи. Так, для визначення рейтингу стану областей України використано 20 параметрів вихідних даних, розподілених за змістом: екологічні – 7 параметрів, економічні – 9 параметрів, соціальні – 4 параметрів. За методом головних компонент встановлені більш вагомі оціночні параметри (рис. 2).

Впровадження методу головних компонент дозволило отримати підвищення адекватності отриманих практичних результатів за наданим методичним забезпеченням при визначенні екологічного рейтингу стану областей України відповідно до інформації про стан регіону за даними А. Б. Качинського [6].

Висновки. Аналіз можливостей розробленої методики екологічного компаратора в системі моніторингових досліджень оцінки якості природно-техногенних об'єктів дозволив обґрун-

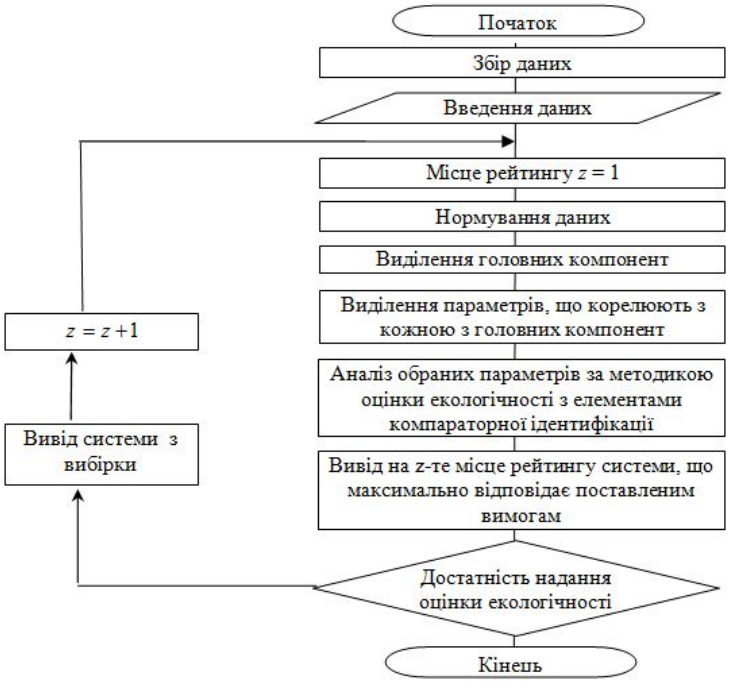

Рис. 1. Алгоритм оцінки екологічності об'єктів за методом компараторної ідентифікації

Eigenvalues of correlation matrix, and related statisti
Active variables only

Value number	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	12,01943	60,09717	12,01943	60,0972
2	2,27842	11,39209	14,29785	71,4893
3	1,61762	8,08808	15,91547	79,5773
4	1,26145	6,30727	17,17692	85,8846
5	1,01966	5,09831	18,19659	90,9829
6	0,67956	3,39778	18,87614	94,3807
7	0,45822	2,29108	19,33436	96,6718
8	0,22171	1,10853	19,55606	97,7803
9	0,17001	0,85005	19,72607	98,6304
10	0,09638	0,48190	19,82245	99,1123
11	0,05916	0,29578	19,88161	99,4081
12	0,05084	0,25418	19,93245	99,6622
13	0,02088	0,10439	19,95332	99,7666
14	0,01868	0,09339	19,97200	99,8600
15	0,01443	0,07217	19,98644	99,9322
16	0,00574	0,02870	19,99217	99,9609
17	0,00396	0,01982	19,99614	99,9807
18	0,00328	0,01640	19,99942	99,9971
19	0,00039	0,00197	19,99981	99,9991
20	0,00019	0,00094	20,00000	100,0000

Рис. 2. Результати встановлення головних параметрів оцінки екологічного рейтингу

тувати доцільність посилення методичного забезпечення у розв'язанні рейтингових завдань методом головних компонент. У роботі запропоновано алгоритмічне забезпечення удосконаленої методики, за отриманими результатами визначено підвищення адекватності кінцевої оцінки у разі відокремлення характеристик, що становлять інформаційний шум.

Література. **1.** Згуровський М. З. Основи системного аналізу / М. З. Згуровський, Н. Д. Панкратова – К.: Видавнича група BHV, 2007. – 544 с. **2.** Козуля Т. В. Методичне забезпечення комплексної оцінки екологічності системних об'єктів методом компараторної ідентифікації / Т. В. Козуля., М. О. Білова // Системный анализ и информационные технологии: материалы 16-й Международной научно-технической конференции SAIT 2014, Киев, 2014 /УНК «ИПСА» НТУУ «КПИ». – К.: УНК «ИПСА» НТУУ «КПИ», 2014. – С.107-108. **3.** Козуля Т. В Обоснование методики компараторной идентификации для системы экологического мониторинга на региональном уровне исследования / Т. В. Козуля., М. О. Білова // Проблеми інформаційних технологій. – №02(014). – 2013. – С. 45–49. **4.** Российское хемометрическое общество. Метод главных компонент [Електронний ресурс]. – Режим доступу: http://chemometrics.ru/materials/textbooks/pca.htm **5.** STATISTICA. Версия 6.1. Системный подход к анализу данных [Електронний ресурс]. – StatSoft Russia. – Режим доступу: http://www.statsoft.ru. **6.** Качинський А. Б. Екологічна безпека України: системний аналіз перспектив покращення / А. Б. Качинський. – К.: Екологічна безпека, 2001. – 251 с.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Козуля Т.В., Ємельянова Д.І.
Національний технічний університет "Харківський політехнічний інститут", Харків, Україна

Методичне забезпечення комплексної оцінки стану природно-техногенних систем

Актуальність дослідження. Природно техногенні об'єкти являють собою динамічні, самоорганізуючі, нестабільні, складно прогнозовані системи, які потребують розробки методів комплексної екологічної оцінки компонентів навколишнього природного середовища (НПС). Комплексність підходу до вивчення складних систем визначає за необхідне розгляд стану складових об'єкта дослідження і процесів взаємодії між ними.

Мета дослідження. Метою дослідження є розробка теоретичних та методичних підходів до комплексної оцінки екологічності стану природно техногенних комплексів (ПТК) в контексті сталого розвитку за допомогою системи екологічних і соціально-економічних показників, які характеризуються точністю оцінки і можливістю урахування окремих дестабілізуючих факторів впливу для різних компонентів НПС.

Реалізація мети обумовила вирішення таких завдань:

1. систематизація напрямків комплексної екологічної оцінки стану ПТК як інструменту забезпечення сталого розвитку регіону;
2. формування системи комплексної оцінки екологічності системних об'єктів.

Методика дослідження і аналіз результатів. Екологічна стійкість ПТК визначається результатами синергетичної дії на системи негативного впливу, їх ролі в самоорганізації і стабілізації стану природно техногенних систем і вивчається на базі причинної моделі «вплив – навантаження – стан – реакція». Ключовим у встановленні кількісної оцінки рівня екологічності і небезпечності впливу господарської діяльності на об'єкти НПС (атмосферне повітря, ґрунт, водні об'єкти) і здоров'я населення є характеристика процесів і явищ, пов'язаних з життєвими умовами виробництва, ступеню порушення прийнятих допустимих границь варіабельності у стані систем і прояву трансформаційних процесів. Звідси і необхідність комплексності у формуванні методичного забезпечення для оцінювання стану складних природно-техногенних утворень і визначення рівня екологічної безпеки. Саме з цією метою запропоновано послідовне застосування узагальнюючих методів екологічності – $MIPS$-аналізу і ризик-аналізу [1].

Для еколого економічної оцінки стану ПТК і обґрунтування превентивних заходів із захисту НПС доцільним є надання результатів $MIPS$-аналізу, що враховують усі матеріальні потоки природно-виробничого комплексу в територіальному чи економічному розрізі аналізу проблем охорони НПС. Перевагою $MIPS$-аналізу є здатність методу визначення небезпечності продукції на кожній стадії її життєвого циклу (ЖЦ), конкретного чинника збільшення екологічних збитків, кількісних характеристик використання ресурсів, починаючи з моменту їх вилучення з природного середовища і до знешкодження вироблених з них продуктів, тобто ЖЦ вилучених природних ресурсів [2].

Екологічність ПТК за MIPS-аналізом визначається на основі MI-чисел і показників кількості продукції:

$$MIPS_{ij} = \frac{\sum_{i=1}^{n} M_i C_i x_j}{S_j}, \tag{1}$$

де $MI_{\text{і}}$ - матеріальна інтенсивність i-ї забруднюючої речовини для компонента НПС, $C_{\text{і}}$ - концентрація i-ї забруднюючої речовини; x_{j} - загальна кількість викидів для j-го об'єкта дослідження; S_{j} - результати діяльності у вигляді продуктивності систем.

Основною метою соціально-екологічної складової комплексного аналізу ПТК є визначення екологічних ризиків у вигляді ступеню порушень за інтенсивністю дії джерела негативного впливу, що дає змогу оцінити стан екосистеми, дати оцінку її захищеності, а також визначити показники для безпечного функціонування об'єктів захисту на різних рівнях взаємодії. Застосування принципів управління екологічним ризиком передбачає раціональний розподіл

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

ресурсних витрат на зниження різних видів ризику з забезпеченням досягнення такого рівня безпеки населення й природного середовища, який досяжний тільки при існуючих у даному суспільстві економічних, соціальних умовах і технологічних можливостях. Оцінка екологічного ризику при сучасному стані i-го компоненту НПС визначається за формулою [3]:

$$P_i^c = f_i(K_i^c, H_i^c), \tag{2}$$

де ${K_i}^c$ – сучасний стан i-го компоненту НПС; ${H_i}^c$ - сучасний рівень антропогенного навантаження від впливу негативних чинників на i-й компонент навколишнього середовища.

Таким чином, комплексна оцінка екологічності стану регіонів є основою екологічно безпечного природокористування та екологічного нормування, оскільки дозволяє визначити компоненти навколишнього природного середовища, які перебувають у найгіршому стані і на основі аналізу причин і джерел забруднення розробити стратегічні плани розвитку регіонів з урахуванням необхідного обсягу зменшення антропогенного навантаження (рис.1).

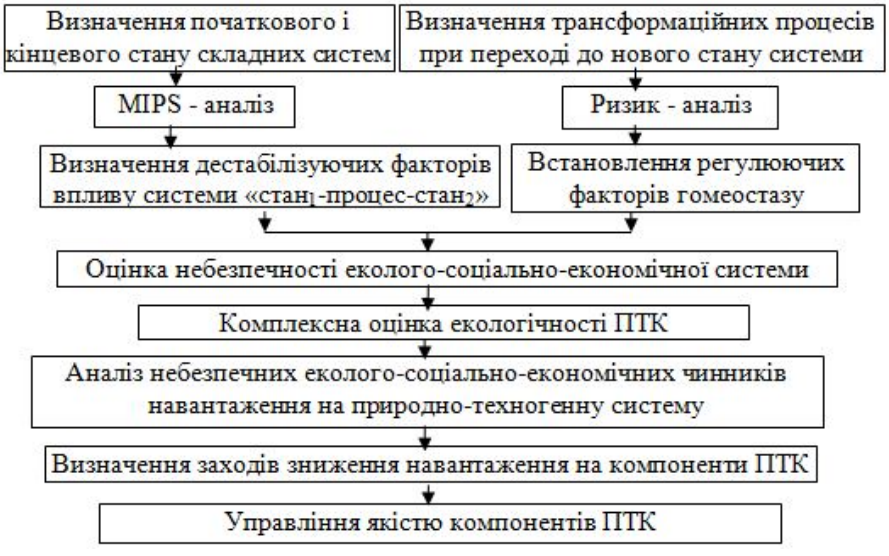

Рис. 1. Схема комплексної оцінки екологічності системних об'єктів

Висновки. Таким чином, у роботі запропоновано застосування елементу комплексності при створенні методології комплексної оцінки екологічності, що дозволяє вирішувати задачу екологічного управління якістю НПС, беручи за основу інформацію стосовно стану складних об'єктів природно-техногенного змісту, обґрунтування результатів удосконалення інформаційно-екологічної складової моніторингу НПС при використанні: 1) даних комплексної оцінки екологічності системного об'єкта дослідження, отриманих від реалізації методичного забезпечення для системи дослідження «фактори впливу – стан системи – навантаження – процес – реакція – стан системи»; 2) комплексного підходу у формуванні управлінського рішення з регулювання екологічної якості системних об'єктів на рівні територіально-об'єктового угрупування (рис. 1).

Література. **1.** Згуровский М.З. Глобальное моделирование процессов устойчивого развития в контексте качества и безопасности жизни людей / М. З. Згуровский, А. Д. Гвишиани. – К.: Політехніка, 2008. – 331 с. **2.** Козуля Т. В. Інформаційно-методичне забезпечення комплексної оцінки екологічності системних об'єктів [Текст] / Т. В. Козуля, Н. В. Шаронова, Д. І. Ємельянова, М.М. Козуля // Системні дослідження та інформаційні технології. – 2014. – №3. – С. 25 34. **3.** Лисиченко Г. В. Природний, техногенний та екологічний ризики: аналіз, оцінка, управління / Г. В. Лисиченко, Ю. Л. Забулонов, Г. А. Хміль – К.: Наук. думка. 2008. – 543 с.

Курмак И.С.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Модель популяции "хищник-жертва" с нелокальным потенциалом

Процесс развития диффузионной неустойчивости, соответствующий нескольким неустойчивым гармоникам, в отличие от диссипативных структур как следствия потери устойчивости однородного по пространству нетривиального равновесия системы, не имеет аналитических методов определения условий возникновения таких структур и их формы. Потому единственным методом является численный эксперимент. [1]

Исследования показывают что одно уравнение с усреднением:

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} + (1-u)\bar{u}, \qquad \bar{u} = \int\limits_{-\infty}^{\infty} u(t,\xi)e^{\frac{-(x-\xi)^2}{2\alpha^2}}$$

лучше описывает отдельную популяцию, чем без усреднения. [2] Поведение системы «хищник-жертва» принято описывать системой уравнений вида:

$$\begin{cases} \dfrac{\partial u}{\partial t} = D_1\dfrac{\partial^2 u}{\partial x^2} + (a-cv)u \\ \dfrac{\partial u}{\partial t} = D_2\dfrac{\partial^2 u}{\partial x^2} + (du-b)v \end{cases}$$

на одномерном ареале $x \in [0; L]$ с непроницаемыми границами.
Для популяционного уравнения (без усреднения) получены двусторонние оценки: [3]

$$u_1(t,x) < u(t,x) < u_2(t,x)$$

, где

$$u_1(t,x) = e^t \int\limits_{-\infty}^{\infty} u(t,\xi)\frac{f(y)}{1-f(y)+f(y)e^t}\frac{1}{2\sqrt{\pi t}}e^{\frac{-(x-y)^2}{4t}}, \qquad (u(0,x)=f(x))$$

$$u_2(t,x) = \frac{v(t,x)e^t}{1-v(t,x)+v(t,x)e^t}$$

В связи с этим предпринимаются попытки определить для систем 2х уравнений какая из моделей лучше описывает поведение популяции, с усреднением или без него. Модель с усреднением:

$$\begin{cases} \dfrac{\partial u}{\partial t} = D_1\dfrac{\partial^2 u}{\partial x^2} + (a-cv)\bar{u} \\ \dfrac{\partial u}{\partial t} = D_2\dfrac{\partial^2 u}{\partial x^2} + (du-b)\bar{v} \end{cases}$$

где

$$\bar{u} = \int\limits_{-\infty}^{\infty} u(t,\xi)e^{\frac{-(x-\xi)^2}{2\alpha^2}}$$

Численный эксперимент показывает что система с усреднением описывает поведение популяции более гладко.

Литература. **1.** . P. E. Souganidis front propagation for a jump process model arising in spatial ecology. Ecole Normale Superieure, DMA, UMR8553 45 rue d'Ulm, 75230 Paris, France. **2.** . Свирежев Ю. В. Нелинейные волны, диссипативные структуры и катастрофы в экологии. Украинский Мат Журнал 2008 №11. http://www.imath.kiev.ua/~umzh/Archiv/2008/11/UMZh_2008_11_1449.pdf **3.** Бондаренко В. Г., Прокопенко Ю. Ю. Барьерные функции для одного класса полулинейных параболических уравнений. Москва "Наука" 1987

Лазарович О.М.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Дослідження деяких властивостей гаусівської багатопараметричної півгрупи

Введення. Теорія півгруп виникла в середині 20-го століття в роботах таких відомих математиків, як Е.Хіллє, Р.Філіпса, К.Іосіди, В.Феллера. Основні застосування цієї теорії можна побачити в абстрактних задачах Коші, рівняннях матфізики, зокрема в параболічних рівняннях, випадкових процесах а також в теорії керування. Оскільки півгрупи мають широке застосування, тому виникає проблема ретельного дослідження їхніх властивостей.
В данній роботі розглядається багатопараметрична гаусівська півгрупа, породжена багатопараметричними групами. Вводиться поняття генератора багатопараметричної групи, а також області визначення знайденого генератора багатопараметричної гаусівської півгрупи.

Означення. Сім'я операторів $T(\overrightarrow{t})$, де $\overrightarrow{t}\epsilon R_n^+ = [\overrightarrow{t}|t_1,..,t_n \geqslant 0]$ називається багатопараметричною півгрупою, якщо:
1) $T(\overrightarrow{s}+\overrightarrow{t}) = T(\overrightarrow{s}) * T(\overrightarrow{t})$
2) $T(\overrightarrow{0}) = I$
3) C_0-півгрупа: $\forall x \epsilon R_n^+ \overrightarrow{t} \to T(\overrightarrow{t})x$ - неперервна функція

Постановка задачі. Задана $T(\overrightarrow{t})$ -багатопараметрична група в банаховому просторі X. $\overleftarrow{A} = (A_1,...,A_n)$ -генератор цієї групи,де $A_k = \frac{\partial}{\partial t_k}|_{\overrightarrow{0}} T(\overrightarrow{t})$ [1]

На основі $T(\overrightarrow{t})$ будується нова сім'я операторів: Нехай $t_{k_1}, t_{k_2}, ...t_{k_m} \neq 0$, якщо $j \neq k_i (i = 1,...,m)$.

$$S(\overrightarrow{t})x = \int_{R_m} \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi t_{k_i}}} e^{-\frac{1}{2}(\frac{\tau_{k_1}^2}{t_{k_1}}+...+\frac{\tau_{k_m}^2}{t_{k_m}})} T(\overrightarrow{\tau})x d\tau_{k_1}..d\tau_{k_m}$$

$$S(\overrightarrow{0}) = I$$

Мета роботи. Дослідити властивості нової сім'ї операторів і знайти її генератор.

Метод розв'язання. Дослідження проводено на базі методів сучасного функціонального аналізу, а саме теорії лінійних операторів та теорії операторних півгруп.

Результати. Доведено, що сім'я операторів $S(\overrightarrow{t})$, є багатопарметричною C_0 - півгрупою, а також знайдений її генератор.
Генератор: $S(\overrightarrow{t}) : \overleftarrow{B} = (B_1,..,B_n)$, де $B_k = \frac{1}{2}A_k^2$ [2]
Також, поставало питання про область визначення знайденого генератора. Під час дослідження було доведено, що:
$\bigcap_{k=1}^{n} D(B_k)$ щільна множина і $D(\overleftarrow{B}) = \bigcap_{k=1}^{n} D(B_k)$

Література. **1.** Голдстейн Дж.А. Півгрупи лінійних операторів і їх застосування/ Голдстейн Дж. А. ; .- К.:Вища школа, 1989. – 345 с. **2.** Данфорд Н. Линейные операторы : в3 т. Т.1. Общая теория / Н.Данфорд, Дж. Шварц .-М. : Издательство иностранной литературы, 1962. — 895 с.

***Лисецкий Ю.М.*[1], *Снитюк В.Е.*[2]**
[1]*Институт проблем математических машин и систем НАН Украины, Киев, Украина;* [2]*Киевский национальный университет им. Тараса Шевченко, Киев, Украина*

Формальное представление корпоративной интегрированной информационной системы как совокупности математических моделей

В настоящее время информационные системы и технлогии являются неотъемлемой частью жизнедеятельности человека. Крупные промышленные предприятия, операторы связи, финансовые учреждения в условиях конкуренции, интернационализации и глобализации экономики не могут функционировать без современных информормационных систем.

Современный этап разработки информационных систем характеризуется созданием различных систем корпоративного уровня, которые не могут работать эффективно, не будучи интегрированными.

Корпоративная интегрированная информационная система (КИИС) основывается на единой программно-аппаратной платформе и общей базе данных, в которой отдельные функциональные подсистемы объединены совокупностью разнородных связей на основе единого технологического процесса обработки информации.

Процесс построения КИИС как любой сложной системы представляет собой логическое единство трех процедур [1]:

- синтеза;
- анализа;
- выбора и принятия решений.

В основе эффективного выполнения этих процедур лежат такие предпосылки:

1) формирование комплекса моделей, которые позволят осуществить идентификацию критериальных функций;

2) разработка интегрального критерия, получение значений которого позволит установить предпочтения на множестве альтернативных вариантов.

Рассмотрим задачу формирования комплекса моделей, которые составляют информационно-аналитический базис исследования. Известно, что в процессе исследований и в процессе построения КИИС выступает пассивной категорией, процесс функционирования которой описывается моделями. Как сложную систему КИИС можно представить тремя моделями [2]: строения, функционирования и развития.

Модель строения (M_c) является теоретико-множественной моделью, реализующей отношение “часть-целое” и представляемая теоретико-множественной операцией объединения

$$A = \bigcup_{i \in I} A_i,$$

где I - индексное множество;

A_i - составляющие элементы A.

Модель функционирования (M_f) определяет процессы достижения целей объектом, которые осуществляются его составляющими и представляются отображениями

$$F_1 :< P, C_c^t, C_s^t, X^{t-1} >\rightarrow X^t,$$

$$F_2 :< P, C_c^t, C_s^t, X^{t-1} >\rightarrow Y^t,$$

где t - момент времени функционирования КИИС;

P - вектор прикладных задач, которые она должна решать;

C_c^t - ее структура в момент времени t;

C_s^t - стратегия управления;

X^t - состояние КИИС;

Y^t - ее выход в момент времени t.

Модель развития (M_p) описывает адаптивные процессы КИИС во внешней среде посредством отображения

$$F_3 :< E_z^n >\rightarrow< E_e^n, E_f^n, E_a^n, R >,$$

где E_z^n и P_r^n - новые цели и процессы целедостижения. соответственно, предполагающие существование в структуре КИИС таких элементов E_e^n, которые способны взять на себя новые функции E_f^n и посредством новых операций E_a^n привести к выработке новых нестандартных решений R. Множества E_z^n и P_r^n - нечеткие категории, что следует из неопределенности их на этапе проектирования, а также из нечеткости временного интервала между предположением о возможности реализации и реализацией E_z^n, определяемой функцией $\xi(Z)$. Оценка такой возможности и ее характеристики определяются на этапе исследования разрешимости задачи Ω. Рассматривая предполагаемое функционирование КИИС, аналогично существованию системы исследования в базисе системных свойств [3], обозначим Q^0 - пространство состояний КИИС. Тогда можно установить существование отображения

$$< E_z^n, P_r^n >\rightarrow Q^0.$$

Пространство Q^0 имеет три составляющие:

$$Q_c^0, Q_p^0, Q_n^0,$$

где Q_c^0 - компоненты состояния точно известные априори;
Q_p^0 - компоненты, распределение вероятностей которых известно;
Q_n^0 - компоненты, значения которых могут быть предложены экспертами.

Следовательно, Q_c^0 не зависит от $\xi(Z)$, для Q_p^0 функцию $\xi(Z)$ можно формально рассматривать как плотность вероятности состояния Q_p^0 КИИС [4]. Учитывая, что составляющие Q_n^0 - нечеткие множества, $\xi(Z)$ - многомерная функция принадлежности [5, 6], устанавливается соответствие

$$Q_n^0 \leftrightarrow Kern(\xi(Z)),$$

где

$$Kern(\xi(Z)) = q_n^m \in Q_n^o | \xi(q_n^m) = \max_{q_n \in Q_n^o} \xi(q_n).$$

Формально модель развития имеет вид:

$$F_3^0 :< E_z^n, P_r^n >\rightarrow Q^0 = Q^o(C, M\xi(Z), Kern(\xi(Z)),$$

где C - компоненты ссостояния системы, являющиеся константами.

Таким образом, разработка тройки моделей $< M_c, M_f, M_p >$ является дополнительным информационным фактором, позволяющим осуществлять структуризацию процесса создания и функционирования КИИС.

Также предложенные модели образуют базис для формирования критериев, которые будут использованы при принятии решений по выбору оптимальной альтернативы проектного варианта при построении КИИС.

Литература. **1.** Молчанов А.А. Моделирование и проектированик сложных систем / Молчанов А.А.- К.: Выща школа, 1988. - 339 с. **2.** Тимченко А.А. Основы информатики системного проектирования объектов новой техники / А.А. Тимченко, А.А. Родионов. - К.: Наукова думка, 1991. - 152 с. **3.** Снитюк В.Е. Эволюционные технологии принятия решений при пожаротушении / Снитюк В.Е., Быченко А.А., Джулай А.Н. - Черкассы: Маклаут, 2008. - 268 с. **4.** Матвеевский С.Ф. Основы системного проектирования комплексов летательных аппаратов / Матвеевский С.Ф. - М.: Машиностроение, 1987. - 240 с. **5.** Дюдуа Д. Теория возможностей / Д.Дюбуа, А.Прад. - М.: Радио и связь, 1990. - 286 с. **6.** Barlett E.B. Dymamic node architecture learning: An information theoretic approach / E.B. Barlett // Neural Networks. - 1994. - N 3. - P. 129 - 140.

Лопатин А.К.[1,2] ***Черненко О.Б.***[1]
[1] *Национальная академия управления, Киев, Украина;* [2] *Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина*

Вероятностные модели оценки фазы финансовых рынков на основе методов нелинейной динамики

В работе [1] выделены статистические циклы в эталонном диффузионном индексе PMI (Еврозона) методом нормированного размаха Херста. В настоящей работе строятся вероятностные модели оценки фазы финансовых рынков на основе указанного индекса с использованием следующих трех факторов:

- степени согласованности действий участников финансовых рынков,определяемой показателем толпы;
- воздействия внешних экономических условий (фундаментальных факторов) системы на поведение индекса (растущий или убывающий экономический тренд);
- расчета количества групп участников с примерно равными уровнями доходности («число участников рынка» в методе когерентного рынка).

Два первых фактора были впервые исследованы в работах по социальной имитации [2]. Дальнейшее развитие эта теория (теория когерентных рынков) получила в работах [3], [4]. В этих работах строится модель вероятностного распределения доходности на финансовом рынке с тремя параметрами: k- степень согласованности поведения толпы (показатель поведения толпы); h- фундаментальное смещение (результат влияния внешних экономических условий); N- число участников рынков, точнее количество групп участников, которые в пределах каждой группы ведут себя примерно одинаково.

Д.А. Филатов сделал предположение о зависимости показателя поведения толпы k от показателя Херста:

$$k = H + 1.3$$

В настоящей работе предлагается использовать алгоритм оценки фазы состояния рынка, используя следующие три показателя исследуемого временного ряда без предварительного построения какого либо вероятностного распределения доходности актива.

- Строятся статистические циклы, на которых значение показателя Херста примерно остается постоянным (показатель поведения толпы k на рассматриваемом цикле). Тем самым, выделяется определенный временной интервал, который является длиной цикла.
- Фундаментальные факторы экономики в пределах рассматриваемых статистических циклов характеризуются средним значением выборки на таком цикле, выраженным в стандартных отклонениях (h^*)
- Предполагается, что «число участников рынка» в методе когерентного рынка соответствует числу карманов, которое необходимо для построения оптимальной гистограммы вероятностного распределения данных рассматриваемого ряда.

В качестве объекта исследования выбран мировой финансовый кризис 2007-2009 годов. В таблице приведены результаты расчетов.

	Цикл1	Цикл2	Цикл3	Цикл4
Показатель Херста (H)	0.4527	0.4649	1.4202	0.2483
k (показатель поведения толпы)	1.8527	1.8649	2.8202	1.6483
Среднее (h^*), влияние внешних экономических условий)	-0.2934	-0.6032	0.1168	0.2054
Число карманов на циклах 1-4	31	31	31	31
Число карманов на каждом цикле	5	5	7	5
Асимметрия	-0.0641	-0.1674	-0.9398	-0.588
Эксцесс	-0.8464	0.0612	1.0793	-0.122

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

	Цикл1	Цикл2	Цикл3	Цикл4
Дата	04. 2006-07.2007	08. 2007-07.2008	08. 2008-01.2010	02. 2010-03.2011

На диаграме показана динамика изменения вероятностного распределения диффузионного индекса от первого до четвертого циклов.

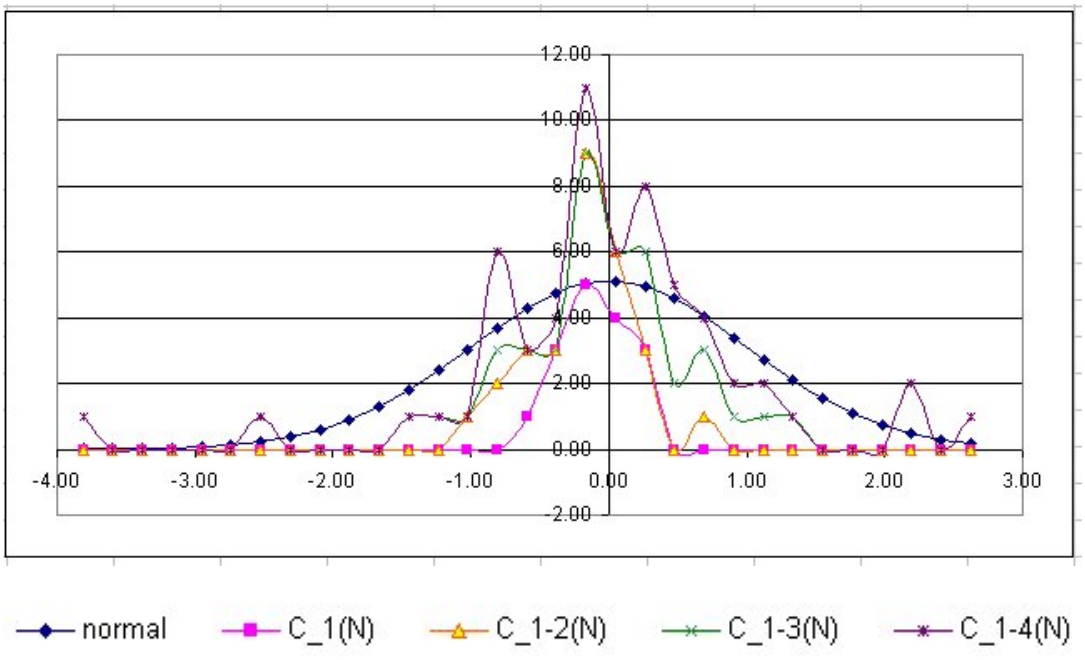

Рис. 1. Сравнение изменения вероятностного распределения диффузионного индекса от первого до четвертого циклов с нормальным стандартным распределением

Как видно из приведенных диаграмм и таблицы циклы определяют следующие состояния (фазы) финансовых рынков (экономики).

1 цикл: Случайное блуждание (эффективный рынок) (k=1.8527), медвежий рынок (h^*=-0.2934),

2 цикл: Случайное блуждание (эффективный рынок) ($k = 1.88649$), усиливающийся медвежий рынок ($h^* = -0.6032$).

3 цикл: Хаотический рынок ($k = 1.6483$), режим толпы на фоне слабого бычьего рынка ($h^* = 0.1168$).

4 цикл: Слабый когерентный бычий рынок ($h^* = 0.2054$). Сильный режим толпы с возвратом к среднему значению ($k = 1.6483$).

Выводы. Выделены фазы финансового состояния экономики по диффузионному индексу, что позволяет выбрать адекватные методы анализа и прогнозирования финансовых рынков.

Литература. **1.** Лопатин А.К., Черненко О.Б., Построение статистических циклов на основе метода нормированного размаха Херста, Матеріали 13 Міжнародної науково-технічної конференції «Системний аналіз та інформаційні технології», 27-31 травня 2013 року, Київ, 127-128. **2.** Callan E. A Theory of Social Imitation / E. Callan, D. Shapiro // Physics Today. 27, 1974. **3.** Vaga T. The Coherent Market Hypothesis / T. Vaga // Financial Analysts Journal. – December/January, 1991. **4.** Яновский Л.П., Анализ состояния финансовых рынков на основе методов нелинейной динамики / Л.П. Яновский, Д.А. Филатов // Научно- практический и аналитический журнал: «Финансы и кредит» № 32 (200), 2005.

Маслянко П.П., Маркова Н.М.
Національний технічний університет України "КПІ", факультет прикладної математики, Київ, Україна

Спосіб оптимізації маршрутів доставки товарно-матеріальних цінностей Рівненської атомної електростанції

Сьогодні Рівненська АЕС – одне з провідних підприємств атомно-енергетичного комплексу України, що динамічно розвивається і модернізується. Основою надійної експлуатації енергоблоків є якісна і своєчасна система товарно-матеріального забезпечення, адже на сьогоднішній день витрати на товарно-матеріальні цінності в собівартості виробництва енергії підприємства становлять не менше 20-30% [1]. Майже половина цих витрат припадає на організацію та обслуговування перевезення матеріалів, обладнання та комплектуючих АЕС [2]. Оптимізація маршрутів дозволить суттєво зменшити обсяг коштів, які витрачаються на доставку товарно-матеріальних цінностей.

Постановка задачі. Об'єкт дослідження – бізнес-процеси автотранспортного підприємства в складі РАЕС, існуючі рішення в області керування запасами, моделі логістики, алгоритми розв'язку задач маршрутизації транспорту, існуючі програмні засоби.

Предмет дослідження – спосіб оптимізації маршрутів доставки товарно-матеріальних цінностей Рівненської атомної електростанції.

Мета дослідження – автоматизувати процес доставки товарно-матеріальних цінностей РАЕС для скорочення термінів поставки, скорочення витрат на повторні поставки, створення довгострокових зв'язків з постачальниками і заводами-виробниками, підвищення ступеню забезпеченості виробництва матеріальними ресурсами, усунення залежаних і не використовуваних ресурсів, зменшення кількості збитків та пошкоджень.

Аналіз існуючих рішень. На теперішній час на вітчизняному ринку існує велика кількість автоматизованих систем керування як вітчизняного, так і зарубіжного виробництва. В області керування товарно-матеріальними цінностями підприємства виділяють наступні рішення. MRP – система, яка дозволяє визначити скільки і в які терміни необхідно виробити кінцеву продукцію. Також система визначає час і необхідну кількість матеріальних ресурсів для задоволення потреб виробничого плану [3]. Впровадження системи для керування матеріальним забезпеченням РАЕС буде невиправдане, тому що вона є застарілою технологією, адже на ринку з'явились більш розвинені програмні комплекси MRP II та ERP. MRP II – система, яка дозволяє детально планувати виробництво підприємства. Забезпечує більш широке охоплення ресурсів підприємства, ніж MRP [4]. Перевагами цієї системи для РАЕС є скорочення тривалості виробничих циклів, зменшення запасів, швидка реакція на зміни попиту і ін. Недоліком системи класу MRP II є організація постачань, яка не враховує ряд важливих обмежень: часових вікон пунктів доставок, обмеження транспортного парку, вантажопідйомності, доступності географічних точок. ERP – система, яка забезпечує збалансоване керування усіма ресурсами підприємства, не лише тими, які відносять до основної діяльності виробничого процесу, а й ресурсами, об'єднаними за допомогою загальної моделі даних про виробництво, фінанси, кадри, попит, споживання [5]. Основним недоліком цієї системи для керування матеріально-технічним забезпеченням РАЕС є висока складність, вартість впровадження, не врахування специфіки робіт станції (забезпечення виробництва електроенергії та безпечної роботи, проведення ремонтних та реконструктивних робіт).

Таким чином, враховуючи обмежену функціональність систем класу MRP та високу вартість систем класу ERP, встановлено, що для якісного та своєчасного керування ресурсами станції доцільно використовувати системи класу MRP II. Для подолання встановлених недоліків системи цього класу потрібно провести суттєве вдосконалення підсистеми організації поставок матеріально-технічних ресурсів. Тому актуальною є задача розробки моделі оптимізації маршрутів товарно-матеріальних цінностей станції.

Спосіб оптимізації маршрутів доставки товарно-матеріальних цінностей РАЕС. Процес доставки товарно-матеріальних цінностей атомної станції вимагає взаємодії виробників мате-

ріалів та обладнання, запасних частин, інструментів та приладдя, комплектуючих, які необхідні для виконання ремонтних та експлуатаційних робіт; паливних компаній; складів; підприємств з обробки відходів; пунктів захоронення відходів.

На основі сутностей системи матеріально-технічного забезпечення РАЕС та їх взаємодії, досліджено і побудовано компонентну модель оптимізації маршрутів доставки товарно-матеріальних цінностей РАЕС (рис. 1).

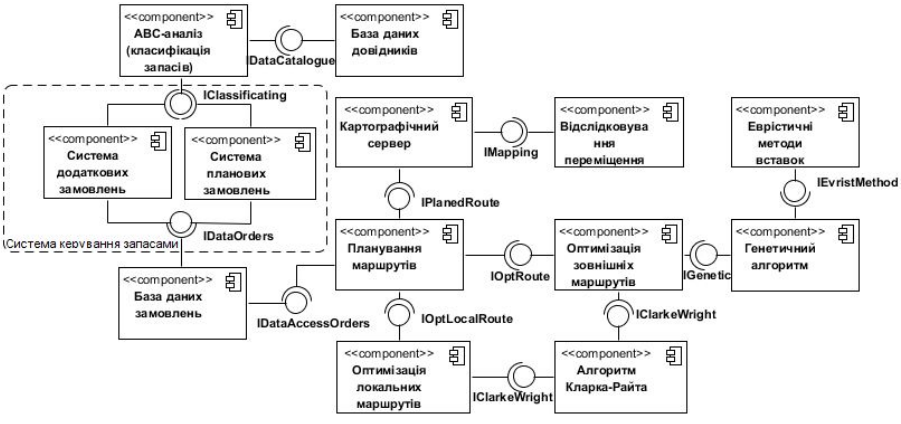

Рис. 1. Компонентна модель оптимізації маршрутів доставки товарно-матеріальних цінностей РАЕС

Одним з варіантів оптимізації локальних маршрутів розвезення партій вантажів від центрального складу до складів цехів, з врахуванням парку автомобілів різної вантажопідйомності, є використання алгоритму Кларка-Райта. Цей алгоритм також використовується як етап побудови початкової популяції для еволюційного пошуку генетичного алгоритму при оптимізації зовнішніх маршрутів. Для розвитку популяції та коректування рішень, що порушують обмеження, застосовуються евристичні методи вставок [6, 7].

За результатами експертних оцінок співробітників РАЕС встановлено, що використання розробленої моделі оптимізації маршрутів доставки товарно-матеріальних цінностей станції в складі системи матеріально-технічного забезпечення дозволить знизити витрати на товарно-матеріальні цінності в собівартості виробництва енергії підприємства приблизно на 10%.

Перевагами використання цієї системи є простота реалізації, можливість скорочення витрат на обслуговування перевезень, скорочення термінів поставки, підвищення ступеню забезпеченості виробництва матеріальними ресурсами, зменшення кількості збитків та пошкоджень.

Результати роботи можуть бути використані як практичний матеріал в організації керування запасами підприємств паливно-енергетичного комплексу.

Література. **1.** Асоціація "Український ядерний форум" [Електронний ресурс]. – Режим доступу: atomforum.org.ua **2.** Тюкаев Д.А. Методологические основы стратегического управления системами материально-технического обеспечения атомных электростанций в условиях неопределенности: дис. ... доктора экон. наук: 08.00.05 // Тюкаев Дмитрий Алексеевич; [Место защиты: Рос. хим.-технол. ун-т им. Д.И. Менделеева]. - Москва, 2014. - 362 с. **3.** MRP [Електронний ресурс]. – Режим доступу: wikipedia.org/wiki/MRP **4.** MRP II [Електронний ресурс]. – Режим доступу: wikipedia.org/wiki/MRP_II **5.** ERP [Електронний ресурс]. – Режим доступу: wikipedia.org/wiki/ERP **6.** Никоноров В.М. Метод Кларка-Райта и его применение для оптимизации логистических показателей развозки // Конкурентоспособность предприятий и организаций: Материалы VI Всероссийской научно-практической конференции. - Пенза: РИО ПГСХА, 2008.-С.25-27 **7.** Schmitt L.J. An evaluation of a genetic algorithmic approach to the vehicle routing problem // Working paper, Department of Information Technology Management., Christian Brothers University, Memphis, 1995.

Маслянко П.П., Оріщенко А.О.
Національний технічний університет України "КПІ", факультет прикладної математики, Київ, Україна

Компонентна модель кластерного аналізу організаційних систем

Організаційна система (Орг.С.) — це об'єднання автономних підприємств, організацій або структурних підрозділів з економічної, соціальної чи державної сфери діяльності, які функціонують під централізованим керівництвом і вирішують спільні завдання [1].

Вирішення задачі виділення необхідної множини Орг.С. за встановленими ознаками з метою організації їх ефективної діяльності вимагає створення концептуальних, математичних, інформаційних та інших моделей систематизації.

У літературі розглядають окремі методи і інструменти [2, 3], які лише частково вирішують це завдання. Крім того, велика кількість існуючих Орг.С. та їх ознак ускладнює процес виділення класів Орг.С. Через це виникає потреба автоматизації процесу кластеризації Орг.С.

Об'єкт дослідження – Орг.С., види і класи Орг.С., класифікація та критерії класифікації Орг.С., класифікація Орг.С. за видами економічної діяльності, концептуальні моделі Орг.С.

Предмет дослідження – компонентна модель кластеризації Орг.С. на основі методів кластерного аналізу.

Метою досліджень є розробка компонентної моделі для автоматизації виділення класів Орг.С. за встановленими ознаками та формалізації бізнес-процесів кластеризації.

Одним із способів систематизації множини Орг.С. є класифікація. Існують наступні класифікації Орг.С. за різними критеріями [4]:

- за «належністю до державних або недержавних»: урядові та неурядові;
- за «метою діяльності»: комерційні та некомерційні;
- за «видом фінансування»: бюджетні та небюджетні;
- за «призначенням»: громадські та господарські.

Такі класифікації поділяють усю множину Орг.С. усього на дві групи. Класифікація Орг.С. «за видом економічної діяльності» виділяє велику кількість груп, але в той же час враховує лише один критерій класифікації.

Нормативним документом, що описує класифікацію Орг.С. за критерієм «вид діяльності» в Україні, є «КВЕД (ДК 009:2010)» [5]. У ньому виділяють 21 економічну галузь, що поділяється на 615 класів.

У результаті дослідження встановлено, що існуючі способи систематизації не до кінця вирішують задачу виділення необхідної множини Орг.С. за виділеними ознаками. У зв'язку з цим актуальною є задача розробки компонентної моделі кластерного аналізу Орг.С.

На Рис.1 представлена компонентна модель кластерного аналізу Орг.С.

Специфікація компонентної моделі:

- компонент «Інтерфейс користувача» – призначений для взаємодії з користувачем. Забезпечує введення вхідних параметрів задачі і виведення результатів;
- компонент «База даних» – призначений для зберігання даних про Орг.С., а також результатів роботи програми;
- компонент «Система управління» – призначений для організації взаємодії між модулями, зокрема для передачі даних між ними;
- компонент «Розрахунок вагових коефіцієнтів ознак» – призначений для обчислення вагових коефіцієнтів ознак, наприклад, за алгоритмом Сааті [6];
- компонент «Розрахунок міри подібності об'єктів» – призначений для нормалізації значень ознак та обчислення «відстані» між об'єктами за обраною метрикою. У результаті формується матриця відстаней D_N;
- компонент «Виділення кластерів на основі правила зв'язку» – формує кластери на основі одного з правил зв'язку, таких як одинарний зв'язок, повний зв'язок, метод

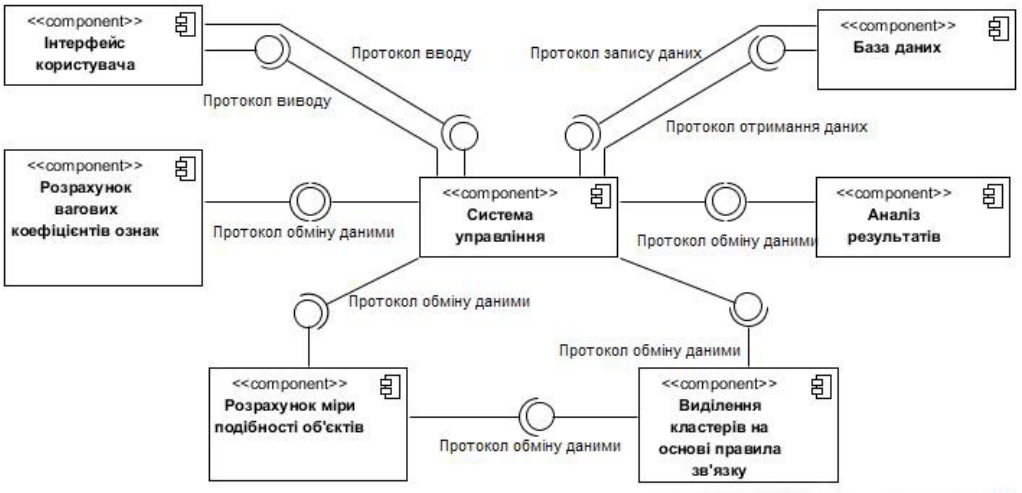

Рис. 1. Компонентна модель кластерного аналізу Орг.С.

Варда, центроїдний метод та ін. Вибір правила виконується в залежності від задачі, яка вирішується. Загальний алгоритм кластеризації виглядає наступним чином:

1. пошук у матриці відстаней D_N найближчих елементів (кластерів) i та j за обраним правилом;
2. заміна елементів (кластерів) i та j об'єднаним кластером h;
3. перехід від D_N до D_{N-1} у результаті об'єднання рядка та стовпця i та j у рядок та стовпець h;
4. повторення алгоритму доти, доки не виконається умова $N = 1$.

- компонент «Аналіз результатів»» – призначений для вибору необхідного рівня кластеризації, аналізу та корегування отриманих результатів.

Розроблена компонентна модель кластеризації Орг.С. дає можливість розробнику швидко і ефективно спроектувати систему, що вирішує задачу кластеризації Орг.С., використовуючи при цьому декілька ознак на відміну від існуючих класифікацій Орг.С. Практична цінність моделі полягає у тому, що вона дозволяє автоматизувати процес виділення класів та видів Орг.С., а це, в свою чергу, зменшує час на вирішення задач виділення класів та видів Орг.С. Перспективою подальших досліджень є порівняння роботи методів, що обчислюють ваговий коефіцієнт ознак систем, а також ієрархічних алгоритмів кластеризації.

Література. **1.** Маслянко П.П. Бізнес інжинірінг організаційних систем / П.П. Маслянко, О.С. Майстренко // Наукові вісті НТУУ "КПІ". – 2011. – №1. – С. 69–78. **2.** Дюран Б. Кластерный анализ. – М.: «Статистика», 1977. – 129 с. **3.** Мандель И.Д. Кластерный анализ. – М.: «Финансы и статистика», 1988. – 176 с. **4.** Дроздов Н.Д. Введенние в теорию организационных систем: Учеб. пособие; Твер. гос. ун-т. Тверь, 2001. – 172 с. **5.** Національний класифікатор України. Класифікатор видів економічної діяльності. – ДК 009:2010. – [Чинний від 2012–01–01]. — К. : Держспоживстандарт України, 2010. – 46 с. **6.** Саати Т.Л. Принятие решений. Метод анализа иерархий. — М.: Радио и связь, 1989. — 316 с.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Надточий А.В.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Оптимизация инвестиционного портфеля в условиях неопределенности

Одной из главных задач, которая стоит перед управляющей компанией или частным инвестором является задача оптимизации инвестиционного портфеля. Определить оптимальный инвестиционный портфель для каждого момента времени довольно сложно, так как ситуация на рынке ценных бумаг постоянно изменяется.

Эффективный портфель – это портфель ценных бумаг, который дает максимально возможную доходность при минимально допустимом уровне риска. [1]

В пятидесятые годы в работах американского математика-экономиста Марковица были заложены основы современной теории портфельного инвестирования. Марковиц привносит в теорию инвестирования стохастический подход. Согласно этому подходу, доходность капитала, что инвестируется, есть случайная величина, которая содержит характеристики, определяющие ожидаемое значение доходности и риск достижения ожидаемого значения. [2]

Пусть имеется фондовый портфель из N активов, в котором каждая компонента имеет свое прогнозное значение (относительное приращение цены актива за период).

Доходность по портфелю будет сумма произведений доходности актива и его веса в портфеле. Риск нечеткого портфеля определяется как возможность событий, при которых реальная доходность будет ниже критического значения.

Тогда, для оптимизации портфеля инвестиций будем использовать два критерия: доходность портфеля и его уровень риска. Рассчитав максимальную доходность при фиксированном уровне риска и минимальный риск при заданном уровне ожидаемой доходности для некоторого временного отрезка, можно определить эффективность инвестиционного портфеля.

Эффективным будет считаться тот портфель, который будет иметь риск, меньший риска начального портфеля, при одинаковом уровне доходности.

Сформулирована задача нечеткой портфельной оптимизации. Поставленная задача также требует построения прогнозных интервалов принадлежности для нечетких доходностей по предыстории котировок акций на рынке. Для этого предлагается использовать метод группового учета аргументов. [3]

Поскольку в общем случае риск носит негативный характер, то оптимальную структуру портфеля предлагается определять путем минимизации риска. Риск же предлагается вычислять, как стандартное отклонение ожидаемой доходности. Если же риск является фиксированным, то оптимальная структура будет достигаться путем максимизации функции доходности.

Литература. **1.** Оптимизация инвестиционного портфеля: [Электронный ресурс] URL: http://www.2stocks.ru/main/school/43/167 **2.** Оптимизация портфеля акций и опционов [Электронный ресурс] URL: Available:http://www.bsu.by/Cache/pdf/125763.pdf **3.** Ю.П. Зайченко. Оптимизация инвестиционного портфеля в условиях неопределенности. Состояние, проблемы, перспективы// I Международная научно-техническая конференция Вычислительный интелект. 2011. С. 33-34

Найдьонов І.М.
Національний технічний університет України "КПІ", факультет біомедичної інженерії, Київ, Україна

Побудова оптимальних маршрутів транспортних засобів з урахуванням топології дорожньої мережі

На сьогоднішній день, в умовах глобального дефіциту природних ресурсів і ситуативного підвищення цін на пальне, як ніколи актуальна проблема скорочення витрат на транспортну логістику. Для досягнення цієї мети пропонується цілий комплекс заходів організації робочого процесу, моніторингу, планування та мотивації персоналу, але ключовим з цієї низки є автоматична побудова оптимальних маршрутів транспортних засобів.

У реальних практичних задачах вхідні умови набагато складніші, ніж в класичній проблемі комівояжера — це в першу чергу розподілення задачі на наявні транспортні засоби (VRP), врахування максимального об'єму та ваги транспортного засобу (CVRP), дозволених часових вікон точок доставки (VRPTW), знання водіями місцевості і конкретних точок доставки, та багато інших. [1]

Всі наявні алгоритми вирішення проблем маршрутизації транспортних засобів мають один суттєвий недолік, який полягає в тому, що вони не враховують топології дорожньої мережі. Всі алгоритми працюють лише з матрицею відстаней між точками, за якою не очевидно, які точки знаходяться в одній місцевості, які точки вже лежать на шляху проїзду машини, і які точки зв'язані між собою таким чином, що витративши час на приїзд до однієї точки, необхідно обов'язково відвідати інші точки в цьому районі тим самим транспортним засобом. Для вирішення цього часто використовуються алгоритми кластеризації, але вони не дають добрих результатів, тому що в такому неевклідовому просторі як дорожня мережа складно визначити однозначні кластери, в тому числі з причин того, що в залежності від напряму проїзду відстані між точками відрізняються. У результаті цього рішення, отримані на основі наявних алгоритмів, не задовольняють вимог логістів, і їх практично неможливо впровадити в реальну практику. Тобто існує необхідність створення системного розв'язання проблеми.

Для вирішення цієї задачі була введена низка новацій (математичних рішень, алгоритмів, евристик, практичних прийомів), таких як:

- врахування у метриці відстані не тільки часу проїзду між точками, а і часу простою при несумісності часових вікон точок;
- побудова орієнтованого *графу сусідства* з урахуванням попутності точок;
- введення понять *ядр* і *хвостів*, де ядра - це скупчення точок, всі з яких транзитивно сусіди одна з одною, а хвости - це всі інші точки, які прив'язані до одного *чи більше* ядер;
- попередня оцінка необхідної кількості машин і одночасне наповнення машин з декількох *точок кристалізації* – початкових ядр для подальшого наповнення через конкуренцію;
- конкуренція об'єктів між машинами "за дельтами", тобто вибір найліпшого наступного ядра, хвоста або точки кристалізації не за принципом мінімізації вартості для цієї машини, а за принципом максимізації різниці між першою найкращою машиною і другою;
- введення поняття *протомашин*: коли точки, які не можуть бути розвезені наявним транспортним парком, не лишаються "за бортом" розкидані по всьому місту, а маршрутизуються так само як і інші, в одну чи декілька додаткових уявних машин, що дозволяє логісту легко прийняти рішення про необхідність залучення додаткових ресурсів.

Данна система реалізована на мові програмування C++, отримала позитивні відгуки експертів, і наразі готується до масового впровадження логістичними відділами кількох крупних компаній.

Література. **1.** L.C. Yeun, W.R. Ismail, K. Omar and M. Zirour, "Vehicle Routing Problem: Models and Solutions," *Journal of Quality Measurement and Analysis*, 4(1), pp 205-218, 2008

Недашковская Н.И.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Метод M_Outflow поиска наиболее несогласованных элементов матрицы парных сравнений

Рассмотрим матрицу парных сравнений (МПС) $D_{n\times n} = \{d_{ij}|i,j = 1,...,n\}$ альтернатив решений $a_1, a_2, ..., a_n$, удовлетворяющую условиям $d_{ij} > 0$ и $d_{ji} = 1/d_{ij}$. Такая МПС используется, например, в семействе методов анализа иерархий для поддержки принятия решений. Элементы МПС определяются по результатам выполненных экспертом сравнений альтернатив решений в шкале. Оценивание уровня противоречивости экспертных оценок парных сравнений и их пригодности для определения весов (приоритетов) альтернатив осуществляется с использованием понятий согласованности МПС [1–4]. Согласованной называется МПС $D_{n\times n}$, у которой транзитивности $d_{ij} = d_{ik}d_{kj}$ выполняются для $\forall i,j,k = 1,...,n$. Слабо (порядково) согласованной называется МПС $D_{n\times n}$, у которой имеют место порядковые транзитивности: $(d_{ij} > 1)\wedge(d_{jk} > 1) \Rightarrow (d_{ik} > 1)$, $(d_{ij} = 1)\wedge(d_{jk} > 1) \Rightarrow (d_{ik} > 1)$, $(d_{ki} > 1)\wedge(d_{ij} = 1) \Rightarrow (d_{kj} > 1)$, $(d_{ij} = 1) \wedge (d_{jk} = 1) \Rightarrow (d_{ik} = 1)$. У слабо несогласованной МПС $D_{n\times n}$ существует цикл, который определяется тройкой индексов (i,j,k), таких что нарушается, по крайней мере, одно из условий в определении слабо согласованной МПС. Существенный недостаток слабо несогласованной МПС состоит в том, что не существует ранжирования альтернатив решений, которое удовлетворяло бы всем элементам этой МПС в том понимании, что $a_i > a_j$ если $d_{ij} > 1$, $a_i < a_j$ если $d_{ij} < 1$ и $a_i = a_j$ если $d_{ij} = 1$.

Известные методы вычисления весов (приоритетов) альтернатив решений обоснованы для МПС с допустимым уровнем несогласованности. Допустимость, в свою очередь, определяется путем сравнения показателей согласованности МПС с пороговыми значениями [2].

Для нахождения наиболее несогласованного элемента МПС с целью повышения общего уровня ее согласованности, а также исключения циклов предлагается метод M_Outflow. Он состоит в вычислении входного и выходного потоков для каждой альтернативы решений с последующим выбором пары альтернатив, на которых достигается максимальное значение разницы этих потоков. Проверка эффективности предлагаемого метода осуществлялась с использованием моделирования на сгенерированных тестовых множествах МПС. В ходе моделирования вначале генерировался некоторый вектор реальных весов и соответствующая ему возмущенная МПС с циклом. Эта матрица подавалась на вход метода M_Outflow и определялся ее наиболее несогласованный элемент, который затем корректировался. Метод считался эффективным, если наиболее несогласованный элемент определялся правильно – тогда вектор весов, вычисленный на основании скорректированной МПС, будет ближе к вектору реальных весов, чем вектор весов до корректировки.

Метод M_Outflow правильно определял наиболее несогласованный элемент МПС более чем в 95 % экспериментов для МПС разных размерностей, и всегда приводил к исключению циклов на исследуемых тестовых множествах матриц. Следовательно, более чем с 95% вероятностью можно утверждать, что вектор весов, вычисленный на основании скорректированной по результатам метода M_Outflow матрицы, будет ближе к вектору реальных весов, чем вектор весов до корректировки этой МПС. Используя результаты моделирования, показано, что предлагаемый метод M_Outflow более эффективен по сравнению с другими известными методами, которые решают ту же задачу.

Литература. **1.** Панкратова Н.Д., Недашківська Н.І. Моделі і методи аналізу ієрархій: Теорія. Застосування: Навчальний посібник. – К: ІВЦ «Видавництво «Політехніка», 2010. –371 с. **2.** N.Pankratova, N.Nedashkovskaya. The Method of Estimating the Consistency of Paired Comparisons // International Journal «Information Technologies and Knowledge», vol.7, №4.-2013.–P.347-361. **3.** Недашківська Н.І. Метод узгоджених парних порівнянь при оцінюванні альтернатив рішень за якісним критерієм //Системні дослідження та інформаційні технології.–2013.- №4.–С.67–79. **4.** Недашківська Н.І. Коригування матриці парних порівнянь без участі експерта // Системний аналіз та інформаційні технології: Матеріали Міжнародної науково-технічної конференції SAIT 2013 (27 – 31 травня 2013 р., м. Київ), http://sait.kpi.ua/books/.

***Пазинич І.С.*[1], *Ланде Д.В.*[2]**
[1] *Фізико-технічний інститут НТУУ "КПІ", Київ, Україна;* [2] *Інститут проблем реєстрації інформації НАН України, Київ, Україна*

Дослідження алгоритмів інформаційного пошуку у піринговых мережах

В цій роботі розглянута проблематика піринговых мереж як елементу глобальної мережі Інтернет та алгоритми пошуку інформації в них (BFS, RBFS, ISM та ін.). Було встановлено, що існуючі алгоритми або не дуже оптимальні з точки зору часу виконання пошуку або з точки зору використання додаткових ресурсів. Було запропоновано новий алгоритм та наведені результати його тестування на основі піринговой мережі, змодельованої за алгоритмом Барабаша-Альберта.

Вступ. Пірінгові мережі (від англ. Peer-to-peer, P2P - один на один) - комп'ютерні мережі, засновані на рівноправності учасників. У таких мережах відсутні виділені сервери, а кожен вузол (peer) є як клієнтом, так і сервером. На практиці пірингові мережі складаються з вузлів, кожен з яких взаємодіє лише з деяким підмножиною інших вузлів [1]. З визначення легко бачити що основною характеристикою таких мереж є децентралізація, всі вузли мережі зв'язуються одне з одним на пряму. Ця децентралізація є й основною перевагою піринговых мереж над традиційними й основним фактором, що створює складності при пошуку певної інформації. В нас немає центрального сервера, до якого можна звернутися з запитом де можна знайти те що нам потрібно. Потрібно використовувати більш гнучкі, адаптивні, можливо, інтелектуальні алгоритмі пошуку. На даний момент вже відомо й активно використовуються п'ять алгоритмів пошуку інформації у піринговых мережах: BFS, RBFS, RWA, ISM, >RES. Кожен з них має свої переваги й недоліки, причому переваги в основному досягаються шляхом введення деякої допоміжної інформації, що зберігається на вузлах.

11-кроковий алгоритм. Для моделювання нашої мережі використовується дещо адаптована векторно-просторова модель пошуку. Кожен вузол мережі володіє певною кількістю документів, кожен документ в свою чергу описується вектором, в якому кожному терму, що присутні в документі, ставиться у відповідність одиниця, а тим, що не використовуються - нуль. Кількість термів (довжина вектора) n задана. Опис запиту, що відповідає необхідній користувачеві тематиці, також являє собою n-мірний вектор. Таким чином для усіх вузлів мережі можна розрахувати дві характеристики: богатство та релевантність. Перша з них відповідає кількості термів, що ними володіє вузол у своїх документах. Звертаючись до суто математичного змісту - це сумарна по всім документам кількість одиниць на різних місцях. Цей показник може приймати значення від 0 до n. Друга характеристика показує на скільки документ за своїм змістом відповідає запиту користувача. Для оцінки близькості запиту і документа використовується операція логічного "І". Тобто чим більша кількість одиниці на однакових місцях у векторі запиту й у векторі документа тим більша ймовірність, що цей документ слід повернути користувачу у відповідь на його запит. Суть нового пропонованого у цій статті алгоритму пошуку інформації у піринговых мережах полягає у плавному переході від прямування до абсолютного богатства до прямування до абсолютної релеванстності. Приведемо пояснення на прикладі:

1. На першому кроці алгоритму серед усіх сусідів стартового вузла перехід відбувається до найбагатшого з них. Тобто крок робиться абсолютно незалежно від змісту запиту.
2. На наступному кроці показник багатства сусідів поточного вузла множиться на 0,9 а релевантність на 0,1. Перехід відбувається до вузла з найбільшою сумою цих двох результатів множення.
3. На третьому кроці багатство сусідів множиться вже на 0,8 а релевантність на 0,2. Перехід відбувається аналогічним чином.
4. ...
5. На останньому одинадцятому кроці перехід відбувається до вузла, що є найбільш релевантним до запиту користувача серед усіх варіантів. Тобто фактично багатство множиться

на 0 а релевантність на 1.

Наведений алгоритм вигідно відрізняється від таких алгоритмів як ISM та >RES тим, що він не потребує зберігання відносно великих об'ємів інформації на вузлах для його роботи але при цьому містить у собі залежність від змісту запиту та характеристик вузла а не просто робить кроки поки не знайде найліпшу відповідь як BFS, RBFS та RWA. З іншого боку цей метод зовсім не гарантує 100% знаходження найкращого результату серед усіх можливих варіантів. Більш того, якщо не реалізовувати ніякого механізму запам'ятовування найкращого поточного результату в кінці можна отримати гірший результат ніж ті, що вже траплялися. Отримати вузол з документами з меншою релевантністю порівняно з вже переглянутими документами. Для тестування алгоритму використовувалася відносно невелика мережа на 50 вузлів, відповідно кількість кроків була обрана базова - 11. Генерування мережі виконувалося за алгоритмом Барабаші-Альберта. Приведемо приклади результатів роботи алгоритму (рис. 2 - 3). На наведених малюнках червоним кольором виділяються номери поточних вузлів, синім -

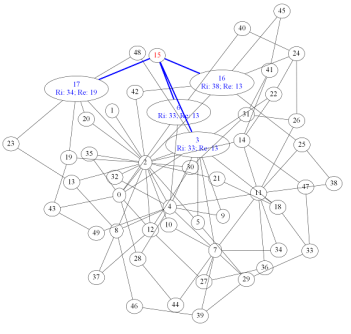

Рис. 1. Початок роботи алгоритму

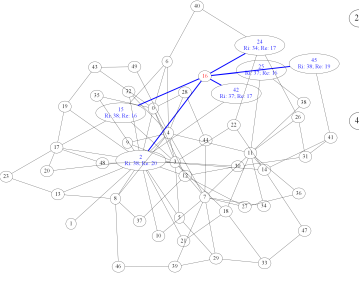

Рис. 2. Після виконання першого кроку

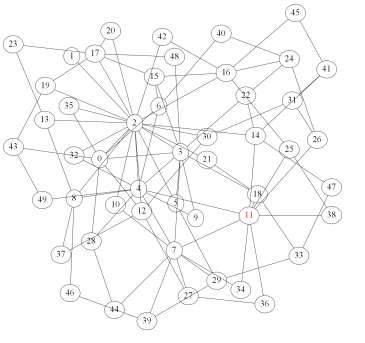

Рис. 3. Кінцевий результат

ребра до усіх сусідів та самі сусіди. Також в середині вузла-сусіда вказуються його показники багатства (Ri від Richness) та релевантності (Re від Relevance). Легко бачити, що алгоритм реалізований коректно й надає досить релевантні результати хоч й не найкращі. Загалом оцінка ефективності роботи алгоритму за релевантністю результату показала що практично завжди результат виходить або найкращім (найбільша релевантність серед можливих) або другим за релевантністю. Середнє відхилення від найкращого результату - 6%.

Висновки. У межах даної роботи було розглянуто поняття піринговых мереж та їх значної ролі у сучасному Інтернет. Основною суттю цього матеріалу є новий запропонований алгоритм пошуку інформації у відповідь на запит користувача, що враховує зміст запиту та певні характеристики вузлів мережі. Звісно аналіз характеристик не такий суттєвий як у алгоритмі ISM але він і не вимагає зберігання значної кількості додаткової інформації на вузлі (профайли у ISM). За наведеними ілюстраціями та деякими показниками ефективності можна побачити, що даний алгоритм в цілому є досить цікавим й, можливо з подальшими доробками, може бути досить актуальним й успішно використовуватись на реальних піринговых мережах.

Література. **1.** Ландэ Д.В. - P2P - по секрету всему свету... О пиринговых сетях - 2008. - 104 с.

Панкратов В.А.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Выявление основных отраслей экономики Украины до 2020 г.

Главным условием достижения позитивных темпов экономического развития как экономики страны в целом, так и отдельных ее предприятий, является активная инновационная и инвестиционная сферы деятельности. Поиск путей решения инновационного развития предприятия, мегаполиса, региона или государства в целом является довольно сложной системной проблемой. Для ее решения целесообразно привлекать синтез возможностей методологии предвидения с другими информационными технологиями, в частности, с методологией когнитивного моделирования, что открывает уникальную возможность в рамках единого программно-аналитического комплекса решать задачи стратегического планирования и оперативного реагирования.

Привлечение на первом этапе процесса моделирования методологии предвидения, которая в достаточной мере формализована (методы SWOT-анализа, Делфи, морфологического анализа и др.), позволяет с помощью экспертного оценивания выявить критические технологии и построить альтернативы сценариев с количественными значениями характеристик. Полученные характеристики являются исходными данными для начальной итерации когнитивного моделирования. Эффективный учет влияния всех релевантных факторов на итоговый результат возможно рассмотреть в рамках математической модели. Наилучшим типом такой модели является когнитивный математический граф, позволяющий структурировать эти взаимозависимости и представить сложную формализацию в простом и удобном для восприятия пользователем виде [1].

Использование модели когнитивных графов целесообразно для получения обоснованных решений поведения сложной системы на стратегическую перспективу при большом количестве взаимосвязей и взаимозависимостей. Ключевым преимуществом когнитивных графовых моделей является возможность с их помощью построить обоснованный сценарий, что делает их практически незаменимым инструментом в аналитической поддержке стратегического планирования развития экономики государства. Эти факторы влияния, не всегда, очевидно, проявляясь в краткосрочной перспективе, являются определяющими при рассмотрении прогноза на долгосрочных временных горизонтах.

Предлагаемый синтез методологий предвидения и когнитивного моделирования используется для решения актуальной задачи выявления основных отраслей экономики Украины до 2020 г. Применяя на этапе предвидения метод морфологического анализа, получено, что наиболее эффективными отраслями развития экономики Украины при существующих вероятностях влияния внешних факторов и взаимосвязей являются: энергетика (альтернативная энергетика), машиностроение (переход на высокотехнологичное производство), добывающая промышленность (разведка новых месторождений), сельское хозяйство (переориентация на энергетические культуры). После вычисления с применением метода морфологического анализа вероятностей альтернатив и результативности была выбрана альтернатива сценария, показатели которой являлись весами в матрице смежности для когнитивной карты.

При построении когнитивной карты выявилось, что для развития основных отраслей промышленности необходимо коренное обновление материальной базы, повышение технологического уровня и квалификации человеческого капитала, усиления политики инновационной деятельности. Важным фактором развития промышленности оказалось международное торговое сотрудничество, на которое влияет большинство положительных факторов, развивая их можно выйти на новый уровень в международном экономическом сотрудничестве. Также одним из основных факторов, который приводит к ухудшению значений всех других параметров, является коррупция. Таким образом, качественно анализируя количественные показатели, возможно построить конкретный курс развития украинской экономики, с четко акцентированными определенными приоритетными отраслями.

Литература. 1. Згуровский М.З., Панкратов В.А Стратегия инновационного развития региона на основе синтеза методологий предвидения и когнитивного моделирования // Системні дослідження та інформаційні технології. – №2. – 2014. – С.7–17.

Панкратова Н.Д., Пинчук В.А., Шишко Н.В.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

К решению задачи структурной оптимизации сложных объектов

Задачу структурной оптимизации сложных объектов решают, используя метод целенаправленного выбора рациональной иерархической структуры по заданным требованиям Q_0 к объекту в целом: $Q_0 = \left\{K_r^0 \;\middle|\; K_r^- \leqslant K_r^0 \leqslant K_r^+; r = \overline{1, R_0}\right\}$ (1), где K_r^0 - r-тый показатель качества объекта. При этом полагается, что структура $\hat{S}_0$ проектированного объекта состоит из $\hat{m}$ иерархических уровней S_i, $i = \overline{1, \hat{m}}$. Каждый иерархический уровень S_i состоит из n_i типов функциональных элементов Φ_{ij}, $j = \overline{1, n_i}$. Каждый функциональный элемент Φ_{ij} характеризуют параметры k_{ijq}, $q = \overline{1, \hat{q}_{ij}}$. Альтернативы функционального элемента Φ_{ij} находятся в множестве $M\Phi_{ij}$, которое можно представить как $M\Phi_{ij} = M\Phi_{ij}^- \cup M\Phi_{ij}^+$; $M\Phi_{ij}^- \cap M\Phi_{ij}^+ = \varnothing$, где $M\Phi_{ij}^- = \{\Phi_{ij\beta}^- \Leftrightarrow k_{ij\beta} \mid k_{ij\beta} \notin \hat{K}_{ij}; \beta = \overline{1, n_{ij}^-}\}$; $M\Phi_{ij}^+ = \left\{\Phi_{ij\beta}^+ \Leftrightarrow k_{ij\beta} \;\middle|\; k_{ij\beta} \in \hat{K}_{ij}; \beta = \overline{1, n_{ij}^+}\right\}$. Причем известны численные значения $n_{ij}^+ > 0$,$n_{ij}^- > 0$,но не известно, какому множеству принадлежит конкретный элемент Φ_{ij} множества $M\Phi_{ij}$.

Предлагается другой способ метод поиска рациональной структуры и множества Парето. Суть метода состоит в следующем. Как и в методе целенаправленного выбора, элементы выбираются последовательно из каждого множества $M\Phi_{ij}$, и если при очередной попытке оказывается, что текущий элемент принадлежит множеству $M\Phi_{ij}^-$,то этот элемент исключают из множества $M\Phi_{ij}$. Отличие данного метода состоит в следующем: если при очередной попытке выбранный элемент Φ_{ij} принадлежит множеству $M\Phi_{ij}^+$, то дальнейший поиск элементов данного типа не прекращают. Выполнение данной процедуры для всех i, j дает нам возможность за линейное время получить множества альтернатив ФЭ, которые удовлетворяют заданным ограничениям (1). Множество Парето рациональных структур представляют в виде всех возможных комбинаций альтернатив из множеств Парето-рациональных альтернатив для каждого ФЭ. Для нахождения этих множеств используется аналог алгоритма линейного поиска с последовательным попарным сравнением альтернатив из множеств $M\Phi_{ij}^+$. Вычислительная сложность такого алгоритма для каждого ФЭ оценивается $O(\hat{q}_{ij} \cdot n_{ij}^+)$, где $\hat{q}_{ij}$ - размерность вектора исходных параметров. Все сложные структуры, полученные таким образом, также удовлетворяют условию (1). При необходимости учесть наиболее весомый критерий, например, цену или объём, рациональную структуру объекта определяют, выбирая из каждого множества альтернатив, полученного на предыдущем шаге, наилучший по данному критерию ФЭ и затем объединяют их. Вычислительная сложность алгоритма оценивается как $O(n_{ij}^+)$.

Чаще всего, в реальных задачах, ЛПР выделяет параметры, которые являются более весомыми, чем остальные. В таком случае возникает задача многокритериальной оптимизации для поиска рациональной альтернативы каждого ФЭ. Для этого используют один из стандартных методов решения такой задачи. Но нельзя априори оценить преимущество того или иного метода. Для конкретной задачи метод должен выбираться непосредственно ЛПР. Множество Парето, в таком случае, включает иерархические структуры сложного объекта, сформированные как всевозможные комбинации полученных рациональных ФЭ. Большинство методов решения задач многокритериальной оптимизации так или иначе сводят задачу к одномерному случаю, что дает возможность применить описанный алгоритм.

Преимущества предложенного метода: при выделении наиболее весомого критерия, данный метод позволяет за линейное время найти рациональную структуру сложной системы, ограничиваясь только вычислительными ресурсами ПК. Если же такой критерий не выделен или их несколько, метод позволяет уйти от задачи транс-вычислительной сложности и не только выявить рациональную структуру, а и предложить решение, более близкое к оптимальному.

Литература. 1. Панкратова Н.Д. Системная оптимизация конструктивных элементов современной техники. //Кибернетика и системный анализ, 2001, 3. С.119-131.

***Печурин Н.К.*[1], *Кондратова Л.П.*[2], *Печурин С.Н.*[1]**
[1] *Институт компьютерных информационных технологий Национального авиационного университета, Киев, Украина;* [2] *Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина*

Модель распределения трафика в беспроводных компьютерных сетях

В стандартах беспроводной компьютерной сети представлена организация сотовой архитектуры с базовыми станциями, являющимися точками доступа сети, которые управляют сотами и взаимодействуют между собой по аналогии с маршрутизаторами, используемыми в проводной компьютерной сети, образуя распределительную систему [1]. Исследование методов определения экстремальных потоков в компьютерной сети связывают с теоремой Форда-Фалкерсона, в соответствии с которой величина максимального потока равна пропускной способности минимального разреза, и анализом уязвимости относительно характеристик, влияющих на показатели работоспособности элементов компьютерной сети [2, 3].

В докладе предлагается экстремальная модель распределения трафика физического уровня в компьютерных сетях с сотовой архитектурой, использующих радиоэфир в качестве среды передачи данных, представленная в классе линейного целочисленного программированиия, что позволяет использовать классические алгоритмы для её исследования. Рассматривается задача поиска потока минимальной стоимости как обобщение задачи поиска максимального потока, которая заключается в определении минимальной стоимости транспортировки потока заданной величины $f_{st} = v$ между узлами s и t, являющимися источником и стоком соответственно.

В связном ориентированном графе G(V,E) с множествами V узлов и E дуг каждой дуге множества E соответствуют веса f_{ij} потока по дуге (i,j), U_{ij} пропускной способности дуги (i,j), c_{ij} стоимости транспортировки единицы потока по дуге (i,j); множество потоков в графе принято называть распределением потоков. Математическая модель задачи включает целевую функцию минимизации стоимости транспортировки потока заданной величины между источником и стоком с ограничениями, характеризующими сохранение потока по дугам, их неотрицательность и ограниченность.

В задаче поиска многополюсных потоков рассматривается неориентированный граф G = (V, E) с пропускными способностями U_{ij} дуг множества E, причем $U_{ij} = U_{ji}$. Здесь используются обозначения f_{ij} для максимального потока между вершинами i и j; $(V_1, \overline{V_1})$ для минимального разреза, разделяющего i от j $i \in V_1, j \in \overline{V_1}$; $U(V_1, \overline{V_1})$ - пропускной способности минимального разреза, разделяющего i от j; $\overline{N_{ij}}$ для множества узлов, сформированного после конденсации всех узлов, кроме i и j; $\overline{E_{ij}}$ для множества дуг, объединяющих узлы множества $\overline{N_{ij}}$. Ребра, соединяющие узел $i \in V_1$ с узлами множества $\overline{V_1}$, сливаются в одно ребро между вершиной i и конденсированной вершиной $\overline{V_1}$; пропускная способность данного ребра определяется суммой пропускных способностей объединяемых ребер: $U_{i\overline{V_1}} = \sum_{m=1}^{r} U_{ij_m}$.

Поиск потоков минимальной стоимости выполняется с использованием алгоритмов Басакера-Гоуэна и Клейна [4], которые характеризуются трудоемкостями, связанными соответственно с формированием потоков и нахождением отрицательного цикла. Данные алгоритмы начинаются с назначения потокам по всем дугам нулевых значений или выбора произвольного допустимого потока величины v, используя алгоритм Форда-Фалкерсона, в котором следует увеличивать поток до достижения заданной величины. Определяется путь минимальной стоимости или отрицательный цикл с учетом модифицированных стоимостей дуг и увеличивается поток на некоторую величину. Формирование искомого потока завершается при достижении заданной величины или отсутствии отрицательного цикла.

Основу алгоритма Форда-Фалкерсона составляет поиск увеличивающейся цепи на множествах дуг, где поток может увеличиваться или уменьшаться; для всех дуг определяется величина пометки. Возможны три состояния узлов сети: не помечен, помечен, помечен и просмотрен. Процедура расстановки продолжается, пока сток не окажется помеченным либо нельзя больше пометить ни один узел и сток остался не помечен. Сформированный поток является максимальным в случае необнаружения увеличивающейся цепи. По окончании вы-

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

полнения алгоритма поиска максимального потока в каждой дуге с помеченным начальным и непомеченным конечным узлами величина потока равна величине пропускной способности, в противном случае непомеченный конечный узел каждой из этих дуг был бы помечен в результате выполнения процедуры поиска увеличивающей цепи. Выполнение алгоритма за конечное число шагов обеспечивается в процессе поиска увеличивающей цепи увеличением на положительную целую величину значения максимального потока между узлами, представляющими источник и сток, и его ограничением сверху пропускной способностью любого разреза, отделяющего данные узлы.

Поиск максимальных потоков в многополюсной сети, представляющей неориентированный граф, выполняется с использованием алгоритма Гомори-Ху, идея которого состоит в построении по итерациям максимального остовного дерева. Алгоритм Гомори-Ху включает повторяющиеся шаги по выбору минимального разреза между узлами, представляющими источник и сток; конденсации в один узел каждой связной подсети, соединенной с группой узлов, включающей источник и сток; определению величины максимального потока из источника в сток и дополнению множества ребер дерева разрезов. Факт построения максимального остовного дерева обеспечивают максимальные значения потоков его ребер.

Алгоритмы поиска потоков минимальной стоимости и максимальных потоков в своей основе используют математические модели с ограничениями, включающими систему линейных алгебраических уравнений (СЛАУ). СЛАУ содержит n^2 переменных (n - число узлов), характеризующих потоки, коэффициенты для которых представляются в виде блочной матрицы. Каждый элемент матрицы представляет вектор-строку коэффициентов потоков для заданного узла. Результат решения СЛАУ обеспечивается в соответствии с теоремой Форда-Фалкерсона. Решение СЛАУ начинается с назначения начальных значений допустимого распределения потоков $f_{ij}|i,j \in V$, соответствующих их верхней границе U_{ij}, и обеспечивает сопоставимость с результатом выполнения указанных выше потоковых алгоритмов.

Альтернативой описанным выше специальным потоковым алгоритмам распределения является использование универсальных или специализированных декомпозиционных алгоритмов исследования непосредственно модели целочисленного линейного программирования. Особенностью матрицы коэффициентов является наличие в системе ограничений блочно-диагональной структуры, что позволяет использовать декомпозиционные подходы Данцига – Вулфа или Корнаи – Липтака с двухуровневой декомпозицией: выделение в исходной модели независимых блоков и формирование модели координирующей задачи со связывающими переменными [5].

В этом случае использование потоковых алгоритмов по сути сводится к поиску решения на основе классических декомпозиционных методов целочисленного программирования для модели с системой линейных алгебраических уравнений, характеризующих сохранение потоков в узлах, их неотрицательность и ограниченность.

Литература. **1.** Рошан П., Лиэри Дж. Основы построения беспроводных локальных сетей стандарта 802.11. – М.: Издательский дом «Вильямс», 2004. – 304 с. **2.** Алгоритмы и программы решения задач на графах и сетях / М.И.Нечипуренко, В.К.Попков, С.М.Майнагашев и др.–Новосибирск: Наука. Сиб. отд-ние, 1990. – 515 с. **3.** Назарова И.А. Анализ переборных алгоритмов для задачи оценки уязвимости многопродуктовых сетей. – М.:ВЦ РАН, 2006. – 72 с. **4.** Бахтин В.И., Коваленок А.П., Лебедев А.В., Лысенко Ю.В. Исследование операций.– Минск: БГУ, 2003.– 199 с. **5.** Жариков А.В. Исследование скорости сходимости некоторых алгоритмов блочного линейного программирования // Управление, вычислительная техника и информатика. – 2011. – С.100-105.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Потебня А.В., Погорілий С.Д.
Київський національний університет ім. Тараса Шевченка, Київ, Україна

Дослідження процесів поширення інформації у наукометричних базах даних

У статті виконано аналіз процесів інформаційного обміну в наукометричних базах даних. Виявлено тенденції розподілу публікацій на окремі групи та встановлено основні способи підтримки цілісності систем. Виконано експериментальне дослідження бази даних *ArXiv*.

Вступ. Моделі складних систем (*complex system*), які описуються набором окремих компонентів та способом взаємодії між ними, є надзвичайно поширеними в сучасних наукових дослідженнях. До них належать соціальні мережі, наукометричні бази даних, всесвітня павутина WWW, комунікаційні структури і т. п. При цьому, формування складних систем пов'язано з появою низки нових властивостей, зокрема синергізму (*synergy*) та емерджентності (*emergence*). Явище синергії пов'язано зі стрімким зростанням ефективності всієї системи порівняно з діяльністю окремих компонентів. Емерджентність є проявом цілісності системи та визначає появу якісно нових властивостей, які не є притаманними жодному з окремих учасників [1].

Нині загальноприйнятою є глобальна інтеграція наукових досліджень, яка здійснюється засобами наукометричних баз даних та соціальних мереж. Однак, при цьому ефективність наукового розвитку виявляється залежною від стану відповідних складних систем, які у зв'язку з активізацією дисинергетичних процесів можуть увійти у фазу рецесії або зазнати повного руйнування. Як наслідок, важливою задачею системного аналізу є дослідження процесів інформаційного обміну в таких системах та діагностика їх загального стану.

Наприклад, найпопулярніша нині база *Scopus* індексує понад 53 млн наукових праць. Крім того, розмірність цієї системи зростає щорічно більш, ніж на 2 млн статей. Подібні тенденції розвитку стабільно демонструють інші наукометричні бази (*Web of Science*, *Google Scholar*, *IEEE Explore*) та академічні соціальні мережі, призначені для обміну документами та їх обговорення (*ReseachGate*, *Academia.edu*, *Mendeley*). Разом з тим, стрімке розширення баз даних супроводжується постійним невдоволенням користувачів щодо закритості провідних систем, платного доступу до публікацій, непрозорості механізмів розрахунку імпакт-факторів та обмеження доступу системи *Google Scholar* деякими видавцями. Наслідком стало формування каталогу відкритих журналів *DOAJ* та низки баз регіонального призначення (наприклад, *РИНЦ*), що призвело до суттєвого розділення основних наукових результатів. Крім того, після короткої фази зростання, структура кожної з таких систем стає дуже нестійкою і потребує додаткової зовнішньої підтримки для збереження життєздатності [1, 2].

Слід зазначити, що цілісність наукометричної бази визначається особливостями її структурної організації та процесами циркулювання інформації. Тому, *метою цієї роботи* є розробка методу діагностики топології та встановлення впливовості вузлів, необхідного сучасним системам для уникнення катастрофічного розпаду центрального ядра основного фрактального компонента на окремі «острови», нездатні до обміну інформацією. Крім того, такий аналіз є необхідним для встановлення здатності системи до проведення мережевої мобілізації, за якої більшість учасників, отримуючи оперативну інформацію та працюючи в найбільш перспективному напрямі, реалізують ефект синергії.

Формування математичної моделі наукометричної бази. Надзвичайно важливою для дослідження організації складних мереж є задача виявлення в них прихованих структур та встановлення режиму їх функціонування. Відомо, що в процесі еволюції складні системи схильні до формування спільнот (модулів) з високою густиною сильних внутрішніх зв'язків за незначної кількості слабких зовнішніх з'єднань. З цим пов'язані основні властивості таких систем – феномени *«малих світів»*, *«клубу багатіїв»*, *«малого діаметру»*, *«важких хвостів»* та ін. Встановлено, що для ефективної роботи мережі найбільш важливими є слабкі міжмодульні зв'язки, за якими інформація передається між окремими спільнотами [2]. Наявність

таких зв'язків скорочує середню відстань між будь-якими науковими статтями до 6 – 7 ребер навіть для найбільших баз даних, а їх руйнування неминуче призводить до розпаду системи на окремі фрагменти. Таким чином, архітектори наукометричних баз обов'язково повинні враховувати стан слабких з'єднань при підтримці своїх розробок.

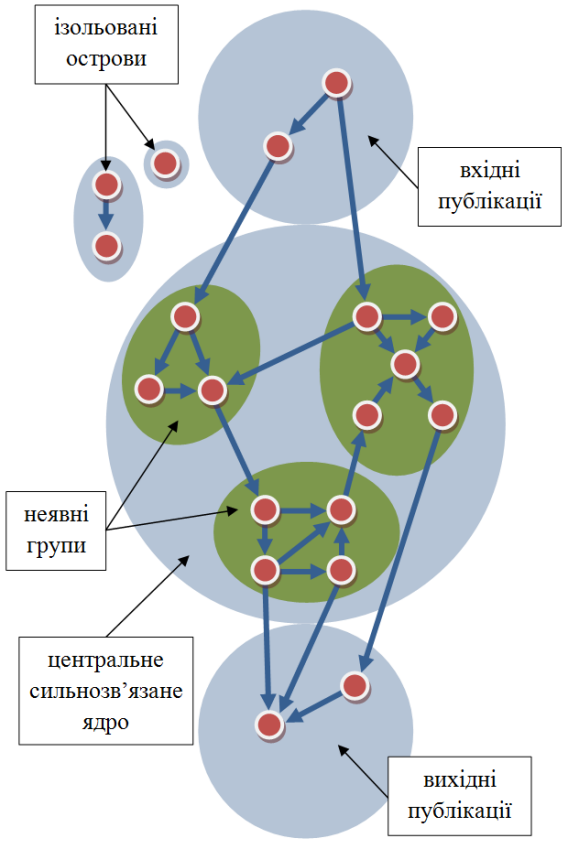

Рис. 1. Приклад структури основного фрактального компонента наукометричної бази

Проблема виділення спільнот у складних мережах є втіленням поширеної задачі розбиття графів (*graph partition problem*). Вона потребує пошуку оптимального розподілу всіх статей за підмножинами, при якому досягає екстремуму задана цільова функція (*objective function*). Важливо, що задача розбиття графів належить до NP-складних задач комбінаторної оптимізації (*combinatorial optimization problem*), тобто час її розв'язання експоненційно залежить від розмірності вхідних даних [3,4]. Як наслідок, при обробці графів реальних систем застосовуються евристичні алгоритми, які передбачають більш ефективний розподіл ресурсів при обчисленні розв'язків.

Застосувавши структурний підхід, представимо наукометричну базу у вигляді орієнтованого графа $G = (V, E)$, де V – набір публікацій, а E – множина цитувань між ними. Задача розбиття такого графа передбачає розподіл множини статей V на k неперерізних підмножин V_i, для яких $V = \bigcup_{i=1}^{k} V_i$ та $V_i \bigcap V_j = \varnothing$ при $i \neq j$. Нехай, E_i – множина внутрішніх ребер групи V_i, а E_{ij} – набір зовнішніх з'єднань між спільнотами V_i та V_j. Слід зазначити, що на відміну від соціальних мереж або світової павутини, графи наукометричних баз даних є ациклічними у зв'язку з неможливістю цитування ще не написаних статей. На рис. 1 наведено приклад спрощеної структури основного фрактального компонента такого графа.

Для встановлення якості розбиття графа найчастіше застосовується функція модулярності (*modularity*) Ньюмана-Гірвана Q, що описує співвідношення густини внутрішніх та міжгрупових зв'язків за обраного розподілу вузлів:

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \sigma\left(c_i, c_j\right)$$

Тут A – матриця суміжності графа, k_i – ступінь вершини i, m – загальна кількість ребер $|E|$,

$$\sigma\left(c_i, c_j\right) = \begin{cases} 1, & c_i = c_j \\ 0, & c_i \neq c_j \end{cases},$$

де c_i – номер класу, до якого належить вершина i. Експериментально показано, що значення модулярності, які перевищують 0,3 є свідченням реальної наявності спільнот у системі [2].

Методи оптимального розбиття наукометричних баз та визначення впливовості публікацій. Таким чином, задача оптимального розбиття графа може бути представлена у вигляді задачі максимізації функції модулярності Q. Однак, її розв'язання шляхом простого перебору всіх варіантів є неможливим у зв'язку з явищем *«комбінаторного вибуху»* [4]. Тому, у праці [2] запропоновано ефективний жадібний алгоритм, заснований на послідовному об'єднанні спільнот V_i та V_j, які забезпечують найбільший приріст функції модулярності $\Delta Q(V_i, V_j) = |E_{ij}| - \frac{d(V_i)d(V_j)}{2|E|}$, де $|E_{ij}|$ – кількість ребер між цими множинами, а $d(V_i)$ – ступінь множини V_i, що є сумою ступенів її складових вузлів $v \in V_i$. При цьому, найбільш вигідним є сполучення підмножин, щільність зв'язку між якими є високою за незначної кількості внутрішніх з'єднань. Важливо, що об'єднання ізольованих наборів виключається, оскільки при $|E_{ij}| = 0$ значення приросту $\Delta Q(V_i, V_j)$ не може набувати додатних значень.

У цій роботі для дослідження реальних наукометричних баз застосовано прискорений ітераційний метод, який складається з двох стадій [5]. Перший етап передбачає формування низькорозмірних спільнот шляхом оптимізації функції модулярності на локальному рівні. При цьому визначається можливість об'єднання кожного набору V_i з його сусідами V_j та розраховуються відповідні значення ΔQ. Обчислення на першій стадії закінчуються отриманням локального максимуму функції модулярності, а другий етап потребує здійснення агрегації підмножин, утворення кластерів більшої потужності та визначення ваги міжкластерних дуг. Ітерації алгоритму продовжуються до одержання стійких множин, склад яких надалі залишається незмінним.

Для визначення відносної впливовості наукових публікацій вводиться метрика центральності вузлів графа, яка може бути визначена різними способами:

1. Центральність за ступенем (*degree centrality*) обчислюється як кількість зв'язків, інцидентних до заданої вершини: $C_D(v) = d(v)$. Однак, деякі статті за високих рівнів цитування можуть бути зв'язані з іншими кластерами в графі малою кількістю ребер. Тому, застосування цієї метрики не є вичерпним при визначенні впливовості публікацій.
2. Центральність за близькістю (*closeness centrality*) показує швидкість поширення інформації між вузлами: $C_C(v) = \frac{|V|-1}{\sum_{t \in V \setminus v} d_G(v,t)}$, де $d_G(v,t)$ – найкоротший шлях від статті v до вершини t. Вона визначає близькість окремої публікації до всіх інших документів у базі. При цьому враховується не лише наявність сусідніх вузлів, але й стан їх зв'язків з іншими документами.
3. Центральність за посередництвом (*betweenness centrality*) обчислюється як кількість найкоротших шляхів між всіма парами документів, що проходять через заданий вузол, тобто $C_B(v) = \sum_{s \neq t \in V \setminus v} \frac{\sigma_{st}(v)}{\sigma_{st}}$, де σ_{st} – загальна кількість найкоротших маршрутів між вершинами s і t, а $\sigma_s t(v)$ – їх число за умови проходження через точку v. Однак, суттєвим недоліком цієї метрики є її обчислювальна складність.
4. Центральність за власним вектором (*eigenvector centrality*) демонструє залежність впливовості вузла $C_E(v)$ від значень впливовості його сусідів $C_E(t)$ при $t \in N(v)$. Обчислення цієї метрики пов'язане зі розв'язанням системи рівнянь виду $C_E(v) = \frac{1}{\lambda} \sum_{t \in N(v)} C_E(t)$, де λ – деякі константи.

Крім того, авторитетність публікацій може бути розрахована за допомогою відомих методів ранжування *PageRank* та HITS (*Hyperlink Induced Topic Search*), які широко застосовуються пошуковими системами при обробці веб-сторінок. Алгоритм *PageRank* встановлює значимість статті шляхом підрахунку впливовості всіх посилань на неї, а метрика HITS передбачає розрахунок для кожного вузла оцінок авторитетності та посередництва. Важливо, що при аналізі графів наукометричних баз застосування алгоритму HITS є надзвичайно ефективним, оскільки він дозволяє виявити як цитовані дослідницькі праці, так і якісні аналітичні огляди (посередники), які посилаються на найбільш авторитетні вузли. Таким чином, крім індексів цитування для визначення впливовості документів у наукометричних базах мають бути застосовані інші метрики, що враховують представлення статті у різних неявних спільнотах [1].

Дослідження процесів інформаційного обміну в наукометричній базі ArXiv. *ArXiv* є однією з найвідоміших безкоштовних наукометричних баз, що містить більше 700 000 статей та препринтів з фізики, математики, комп'ютерних наук і біології. У статті для дослідження обрано фрагмент цієї системи (секцію фізики високих енергій), що охоплює 34 546 публікацій та 421 578 посилань між ними [6]. Зв'язки зі статтями, які не входять до бази *ArXiv*, були вилучені зі сформованого графа.

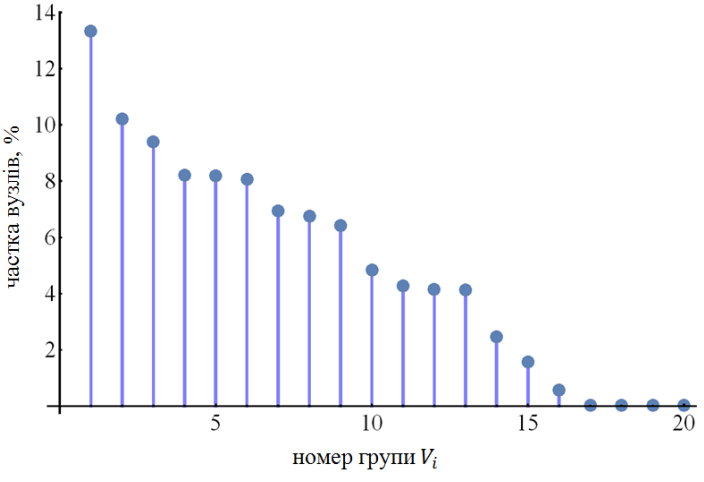

Рис. 2. Розмірність найбільших кластерів у системі *ArXiv*

Встановлено, що граф наукометричної бази містить 61 компоненту слабкої зв'язаності (*weakly connected component*), що має зв'язок між всіма вершинами принаймні в одному напрямі. Найбільший з виділених підграфів охоплює 99,58% всіх статей, а всі інші компоненти є островами, що об'єднують лише декілька публікацій. Разом з тим, у графі виділено 21 608 компонент сильної зв'язаності (*strongly connected component*), в межах яких існує взаємний зв'язок між будь-якими вузлами. При цьому одна з них є гігантською та містить 36,79% вузлів, а розмірність решти островів не перевищує десяти публікацій. Тобто, більшість статей намагаються цитувати найбільш авторитетні публікації, однак зворотний зв'язок мають лише третина з них.

Шляхом оптимального розбиття графа було отримано 80 неявних спільнот за значення модулярності $Q = 0,727$, що є свідченням суттєвої кластеризації. На рис. 2 показано розмірність 20 найбільших груп. При цьому лише дві зі сформованих множин охоплюють більше 10% документів. На рис. 3 наведено залежність модулярності від значення порогу цитування публікацій, за яким відбувалася фільтрація вузлів у графі.

Видно, що статті з низькими індексами цитування, як правило, пов'язані лише з вузлами своєї спільноти та не сприяють збільшенню цілісності системи. Однак, вилучення з графа публікацій з середнім рівнем цитування (50 – 120 разів) призводить до стрімкого зростання модулярності, руйнування «мостів» між кластерами та відділення нових островів від гігантської компоненти. При цьому найбільш цитовані праці рідко пов'язані між собою і за високих рівнів порогу відсічки розташовуються на окремих ізольованих островах. Таким чином, основну зв'язувальну роль у наукометричній базі відіграють статті з середнім рівнем цитування.

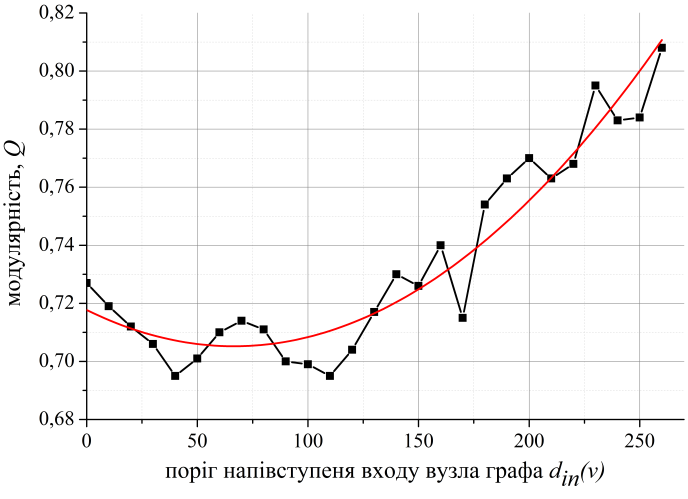

Рис. 3. Залежність цілісності бази *ArXiv* від значення порогу цитування публікацій

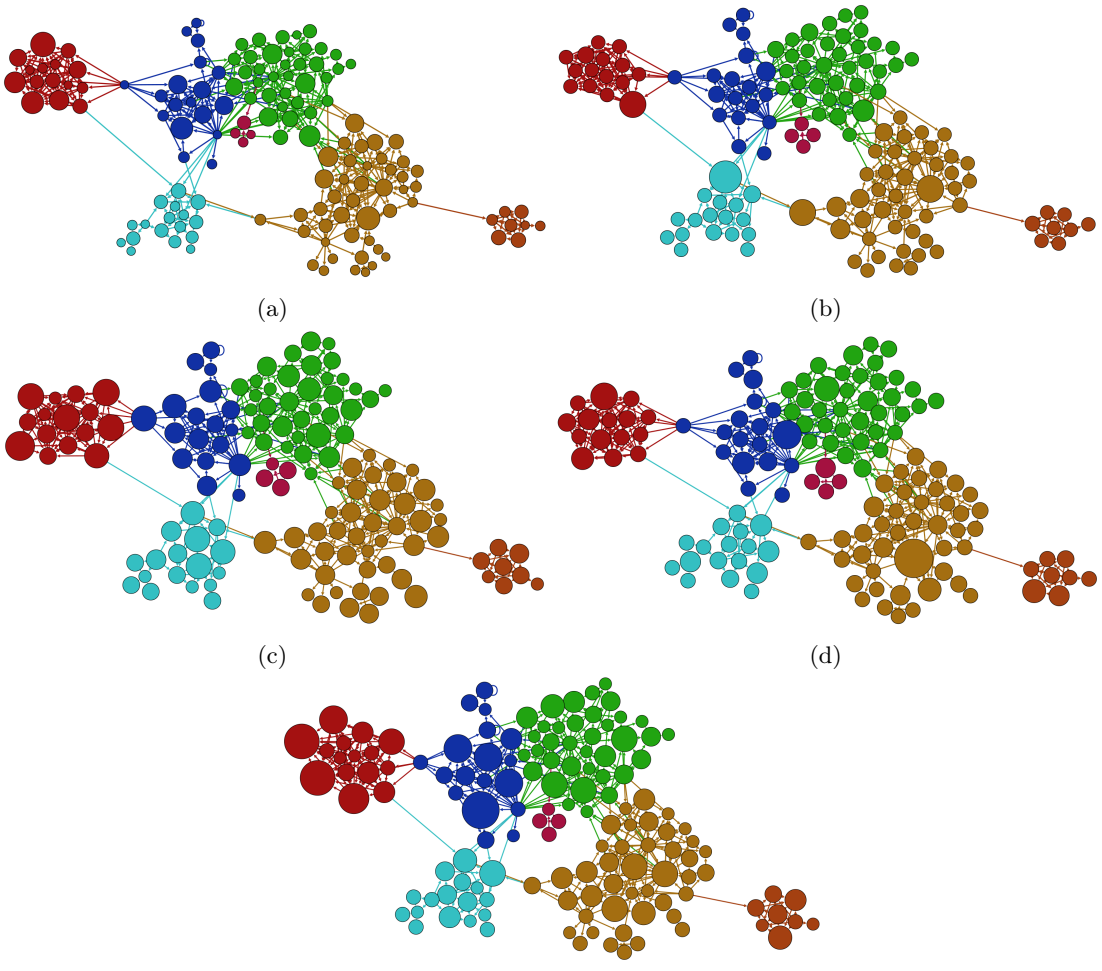

Рис. 4. Визначення за різними метриками впливовості публікацій, цитованих у системі *ArXiv* більше 150 разів. Кольорами показано різні спільноти в графі. Розмір вузлів є пропорційним до їх авторитетності

Розглянемо найбільший компонент зв'язаності графа системи *ArXiv* за порогу відсічки, що складає 150 цитувань (рис. 4). Видно, що високі значення модулярності та значна автономія спільнот обумовлюють недостатність застосування напівстепенів входу для опису впливовості публікацій (рис. 4a). При цьому мостові вузли, вилучення яких призводить до катастрофічного розпаду, мають найменші значення цитування, а найбільші значення метрики здобувають вузли, «занурені» в глибину своїх спільнот. Натомість, при використанні в якості метрики центральності за посередництвом (рис. 4b) або за близькістю (рис. 4c) впливовість таких документів є набагато вищою.

Застосування алгоритму *PageRank* (рис. 4d) призводить до здобуття найбільшого рангу статтями, зв'язки яких, як правило, зосереджені в межах однієї спільноти. Показовим є використання алгоритму HITS для знаходження найкращих посередників, які також містяться всередині кластерів поруч з найавторитетнішими дослідницькими працями (рис. 4e). Загальна структура системи *ArXiv* для статей, цитованих більше 100 разів, міститься на рис. 5.

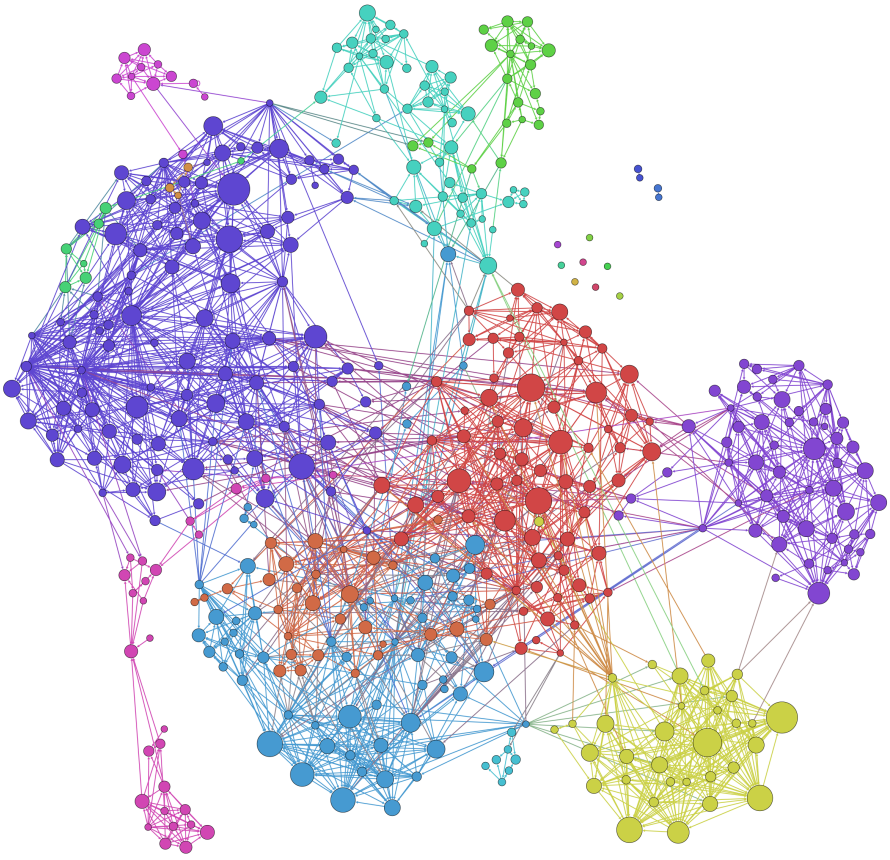

Рис. 5. Структура графа наукометричної бази *ArXiv* для публікацій, що мають більше 100 посилань. Розмір вузлів є пропорційним до індексу їх цитування

Висновки. Таким чином, на прикладі системи *ArXiv* встановлено, що структура провідних наукометричних баз є розподіленою на окремі неявні спільноти, зв'язки між якими підтримуються за рахунок публікацій з середнім рівнем цитування. Саме вони, як правило, одночасно містять посилання на різні авторитетні праці та формують між ними мостові з'єднання, що відіграють важливу роль в процесі обміну інформацією. Однак, використання індексів цитування, поширене в сучасних системах, не дозволяє виявити такі критичні вузли. Для цього більш доцільним є застосування метрик центральності за посередництвом та за близькістю.

Література. **1.** M. Newman, Networks: An Introduction. – Oxford University Press, 2010. – 720 p. **2.** M. Newman, Modularity and community structure in networks // Proceedings of the National Academy of Sciences. – 2006. – Vol. 103(23). P. 8577 – 8582. **3.** А.В. Потебня, С.Д. Погорілий, Архітектура високопродуктивного нейрокомп'ютера для розв'язання задач комбінаторної оптимізації // Праці міжнародної конференції «Високопродуктивні обчислення HPC-UA 2014». (Україна, Київ, 14 жовтня, 2014). – С. 88 – 99. DOI: 10.5281/zenodo.16941. **4.** A.V. Potebnia, S.D. Pogorilyy, Exploration of data coding methods in wireless computer networks // Fourth International Conference on Theoretical and Applied Aspects of Cybernetics (TAAC), Kyiv. – 24 – 28 November 2014). – P. 17 – 31. DOI: 10.13140/RG.2.1.3186.3844. **5.** A. Lancichinetti, F. Radicchi, J. Ramasco, S. Fortunato, Finding Statistically Significant Communities in Networks // PLoS ONE 6(4): e18961. – 2011. DOI: 10.1371/journal.pone.0018961. **6.** J. Leskovec, A. Krevl, SNAP Datasets: Stanford Large Network Dataset Collection. – June 2014. URL: http://snap.stanford.edu/data.

Путренко В.В.
Світовий центр даних з геоінформатики та сталого розвитку НТУУ "КПІ", Київ, Україна

Використання глобальних моделей інтерполяції у інтелектуальному аналізі геопросторових даних

Одним із напрямів аналізу даних є інтелектуальний аналіз геопросторових даних, який базується на використанні інформації про геопросторове розташування об'єктів в локальних та глобальних системах координат. Серед широкого кола методів обробки та аналізу даних, які використовуються у сфері інтелектуального аналізу значне місце посідають методи трендового аналізу на основі глобальних інтерполяційних моделей. Найбільший розвиток методи статистичного трендового аналізу здобули при аналізі часових рядів та побудові прогнозних моделей динаміки явища [1]. Просторові тренди було достатньо добре вивчені в рамках статистичної обробки інформації, відображення багатомірних просторів та зменшення розмірності даних. Сьогодні моделі трендового аналізу є корисними при вивченні геологічних структур, екологічних проблем, соціально-економічних явищ тощо.

Особливістю використання трендової моделі є глобальність та абстрактність отриманих результатів оцінювання, що може ускладнювати процес їх інтерпретації при підтримці прийняття рішень. З іншої сторони, результати аналізу трендової моделі можуть бути основою для виявлення глибинних зв'язків, які існують між даними.

Глобальні методи інтерполяції одночасно використовують всі наявні дані для виконання прогнозу для всієї даної території, тоді як локальні методи оперують у межах невеликих зон навкруги належних інтерполяції вузлів для того, щоб забезпечити виконання оцінки тільки за даними, розміщеними в безпосередній близькості від точок прогнозу або оцінювання.

Глобальні інтерполяції, як правило, використовуються не для безпосередньої інтерполяції, а для дослідження і можливого видалення ефекту глобальних варіацій (тренда), обумовлених зовнішніми чинниками. Після того як глобальні ефекти будуть видалені, відхилення від глобальних варіацій можуть інтерполюватися з використанням локальних методів. Глобальні методи у більшості випадків базуються на стандартних статистичних ідеях варіаційного аналізу і регресії [2]. До них відносять:

- класифікації з використанням зовнішньої інформації;
- поліноміальну регресію з геометричними координатами;
- регресійні моделі.

Класифікаційні методи використовують досяжну інформацію (таку, як ґрунтові типи або адміністративні території) для того, щоб поділити досліджувану територію на регіони, які можуть бути охарактеризовані статистичними моментами (середньою, дисперсією) атрибутів, виміряних у точках, розміщених у межах цих регіонів.

Поліноміальні інтерполятори відображають значення у точці у вигляді поліному від координат. В двовимірному випадку – для точки x з координатами (x, y) Z*(x, y) = Pn(x, y), де Pn — полином n-го ступеня. Зазвичай на практиці для двовимірного випадку використовують один з чотирьох типів поліномів:

- площина: $P_1(x, y) = a + bx + cy$;
- білінійно-сідловий: $P_{1.5}(x, y) = a + bx + cy + dxy$;
- квадратичний: $P_2(x, y) = a + bx + cy + dxy + ex^1 + fy^2$;
- кубічний: $P_3(x, y) = a + bx + cy + dxy + ex^2 + fy^2 + gx^2y + hxy^2 + ix^3 + jy^3$.

Теоретично можна використовувати і поліноми більш високого порядку. Вони визначаються максимальним ступенем при x, максимальним ступенем при y і спільним максимальним ступенем. Всі проміжні ступені в поліномі будуть присутні.

Таким чином, горизонтальна поверхня має порядок нуль, похила плоска поверхня - перший порядок, квадратична поверхня – другий порядок, кубічна поверхня з десятьма коефіцієнтами має третій порядок. Знаходження коефіцієнтів a_i є стандартною процедурою в задачах на множинну регресію, тому обчислення легко виконуються за допомогою стандартних статистичних

пакетів. Після знаходження коефіцієнтів ai трендова поверхня може бути відображена оцінкою значень $Z(x,)$ у всіх точках регулярної мережі. Порядок трендової поверхні P з формальної точки зору може бути скільки завгодно великим. Збільшення порядку полінома, як правило, призводить до збільшення ступеня його відповідності реальній поверхні. Проте при цьому, з одного боку, втрачається фізичне значення побудови, з іншого - різко збільшуються вимоги до обсягу даних, необхідних для знаходження коефіцієнтів a_i.

Завдання поліноміальної інтерполяції зводиться до того, щоб визначити невідомі коефіцієнти ai. так, щоб поліноми максимально добре відповідали даним в заданих точках. Для цього знаходять мінімум по всіх коефіцієнтах (a, b, c, d і т. д.) функції χ^2, яка задає інтегральну помилку інтерполяції та обумовлена наступним чином:

$$\chi^2 = \Sigma_{i=1}^{N}(Z(x_i, y_i) - P_n(x_i, y_i))^2.$$

Мінімізація полягає у вирішенні системи лінійних рівнянь з числом невідомих, що дорівнює кількості рівнянь. Число рівнянь (невідомих) залежить від обраного полінома. Глобальний поліноміальний метод не є інтерполятором в строгому сенсі, скоріше він відноситься до апроксиматорів. Його можна використовувати для виділення великомасштабного тренда.

Можна скористатися і локальним варіантом поліноміального методу, коли пошук коефіцієнтів проводиться тільки на основі даних, які потрапили в зону пошуку. Глобальні методи інтерполяції було використано з метою перевірки складових індексу сталого розвитку на наявність просторових трендів.

Для проведення дослідження було використано набір даних, який описує структуру індексу сталого розвитку України в 2013 році. З метою перевірки гіпотези про існування просторового тренду в розподілі даних про сталий розвиток було використано спочатку картографічний метод, який дозволяє оцінити загальний паттерн просторового розподілу значень. З цією метою було побудовано тематичну карту значень індексу сталого розвитку за областями на основі поділу на п'ять груп за природними межами класифікації значень.

Аналіз карти дозволяє стверджувати про наявність певної зміни у розподілі значення індексу у географічному просторі. Для підтвердження цього припущення було використано метод побудови поліноміальної трендової поверхні, яка була побудована на основі точкового файлу, який містить дані про геометричні центри областей України. В результаті проведення експериментальних побудов поверхонь було з'ясовано, що практичне значення при виявленні тенденцій мають поверхні утворені поліномами 1 – 3 ступенів. Якщо поліноми першого ступеня допомагають встановити загальний напрям та силу тренду в даних, то поліноми 2 – 3 ступенів дозволяють підтвердити існування просторової тенденції у даних та визначити загальні риси її розподілу.

Допоміжним інструментом аналізу тренду є графік аналізу тренду, який у площинах північ – південь, захід – схід демонструє наявність трендів у розподілі даних, які у вигляді результуючого тренду визначаються на трендовій поверхні.

Трендові поверхні можуть бути оцінені як за допомогою значень коефіцієнтів, які використовувались у поліномі, так і описом характеристик трендової поверхні. Такими характеристиками є ухил поверхні, кут між лініями скату поверхонь першого та другого поліномів. Додатковими інструментами аналізу можуть виступати локальні методи інтерполяції даних, такі як обернено-зважені відстані та розподіл ухилів отриманих статистичних поверхонь.

Таким чином, використання моделей глобальної інтерполяції для виявлення просторових трендів у розподілі геопросторових даних дозволяє виявляти закономірності у структурі даних, що є одним із завдань інтелектуального аналізу.

Література. **1.** Згуровський М.З., Болдак А.О., Єфремов К.В. Интеллектуальный анализ и системное согласование научных данных в междисциплинарных исследованях // Кибернетика и системный анализ. – 2013. – № 4. – С. 62-75. **2.** Burrough., McDonnel, Principles of Geographical Information systems for land resources assesment. – Oxford Univercity Press, 1998. – 333 p.

Романенко В.Д., Милявский Ю.Л.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Инструментарий адаптивного и координирующего управления импульсными процессами когнитивных моделей сложных систем

Когнитивное моделирование в настоящее время является одним из распространенных подходов для исследования проблем слабоструктурированных социально-экономических систем. Оно применяется в тех случаях, когда объектом исследования является сложная система большой размерности, структуру которой можно представить взвешенным ориентированным графом. В когнитивном моделировании этот граф принято называть когнитивной картой (КК) сложной системы. При этом вершины КК отражают координаты (факторы), а ребра - связи между этими координатами. Правило изменения значений координат вершин КК при импульсном процессе, когда на одну или несколько вершин вершин Y_i подействовали возмущения, сформулировано в виде разностного уравнения первого порядка в приращениях

$$\Delta Y_i(k+1) = \sum_{j=1}^{n} w_{ji}\Delta Y_j(k), \tag{1}$$

где $\Delta Y_i(k) = Y_i(k) - Y_i(k-1), i = 1, 2, ..., n$, w_{ij} - весовой коэффициент ребра КК, которое идет от j-ой вершины к i-ой, а n - количество вершин КК. В векторной форме выражение (1) имеет вид $\Delta\overline{Y}(k+1) = W^T\Delta\overline{Y}(k)$, где W - весовая матрица смежности, а $\Delta\overline{Y}$ - вектор приращений кооринат вершин Y_i.

В данном докладе рассмотрен инструментарий решения задач управления импульсными процессами КК. При этом разработаны условия управляемости КК в режиме импульсных процессов, при выполнении которых можно реализовать адаптивное и координирующее управление.

Часто оказывается, что для КК различных сложных систем импульсные процессы являются неустойчивыми. Поэтому первой задачей, которая решается авторами, является стабилизация неустойчивых импульсных процессов КК на основе синтеза модальных регуляторов состояния, которые формируют управления, непосредственно воздействующие на вершины КК.

Вторая задача посвящена разработке системы стабилизации координат вершин КК на заданных уровнях. Для этого разработана управляемая динамическая модель импульсного процесса типа "вход - выход", а в качестве критерия синтеза дискретного регулятора используется эталонная модель характеристического уравнения замкнутой системы управления.

Третья задача посвящена координации значений вершин КК во время импульсных процессов с целью регулирования заданных соотношений между координатами отдельно взятой КК.

Следующей задачей является управление двумя взаимосвязанными КК. Их совместная динамика представлена в виде системы уравнений импульсных процессов

$$\Delta\overline{X}(k+1) = W_1\Delta\overline{X}(k) + V_1\Delta\overline{Y}(k),$$
$$\Delta\overline{Y}(k+1) = W_2\Delta\overline{Y}(k) + V_2\Delta\overline{X}(k),$$

где $\overline{X}, \overline{Y}$ - векторы координат вершин первой и второй КК. Задача координирующего управления сформулирована и решена путем синтеза управлений на основе векторно-матричного уравнения соотношений $S\Delta\overline{X}(k) = \overline{f}(\Delta\overline{Y}(k))$, где S - заданная матрица размерности $s \times n$ при $s < n$, а $\overline{f}$ - некоторая вектор-функция.

Как правило, коэффициенты w_{ji} импульсных процессов КК (1) определены неточно либо могут изменяться во времени. В докладе проведено исследование по адаптивному оцениванию указанных коэффициентов на основе рекуррентного метода наименьших квадратов.

Последняя задача посвящена синтезу управления импульсными процессами КК при наличии запаздываний в передаче энергии ии информации между вершинами Y_i КК или запаздываний при формировании и передаче внешних управляющих воздействий.

Романкевич А.М., Романкевич В.А., Морозов К.В.
Национальный технический университет Украины "КПИ", кафедра системного программирования и специализированных компьютерных систем, Киев, Украина

Об одном условии модификации реберных функций GL-модели

Отказоустойчивые многопроцессорные системы (ОМС) с каждым годом находят все более широкое применение, особенно в системах управления (СУ) сложными объектами. Это касается в первую очередь систем и объектов критического применения. ОМС позволяют решать важнейшие задачи построения СУ: с одной стороны, обеспечить достаточную вычислительную мощность, а с другой – сделать ее в достаточной мере надежной. В процессе построения ОМС разработчику важно иметь возможность рассчитывать уровень ее надежности. Это позволит ему выбрать оптимальный вариант архитектуры системы. В то же время, рассчитать уровень надежности ОМС часто бывает весьма сложно в силу того, что помимо большого количества процессоров, СУ могут иметь сложную и неоднородную структуру. Существуют различные методы расчета показателей надежности ОМС, все они имеют свои достоинства и недостатки. Один из таких методов был предложен в [1] и основан на проведении статистических экспериментов с так называемыми графо-логическими или GL-моделями. Они сочетают в себе свойства графов и логических функций и позволяют моделировать поведение ОМС в потоке отказов.

Модель принимает на вход вектор состояния системы (содержащий булевы переменные, которые соответствуют каждому из процессоров системы и принимают значение 1, если процессор исправен, и 0 – если он вышел из строя) и определяет, останется ли в такой ситуации система работоспособной или выйдет из строя. Модель основана на графе с переменной структурой, каждому ребру которого соответствует булева реберная функция, принимающая значение 0 или 1 в зависимости от значений переменных вектора состояния. Если функция принимает нулевое значение, то соответствующее ей ребро исключается из графа. Если граф модели на некотором векторе остается связным, это соответствует работоспособному состоянию системы, потеря же связности соответствует выходу системы из строя.

ОМС, состоящая из n процессоров, которая остается работоспособной ровно до тех пор, пока из строя вышли не более, чем m любых ее процессоров, называется базовой. Ей соответствует базовая GL-модель, которая обозначается $K(m,n)$. К сожалению, не все реальные системы являются базовыми. Также бывает, что система отличается от базовой m-отказоустойчивой лишь тем, что остается работоспособной на некотором множестве векторов с $m+1$ нулем. Такие векторы называются блокируемыми. В [2] был предложен метод модификации базовой модели путем внесения изменений в одну из ее реберных функций, в результате применения которого блокировалось некоторое множество векторов с $m+1$ нулем. Однако, такое решение не всегда оказывается достаточным. Приходится модифицировать несколько функций одновременно.

В результате анализа выяснилось, что в случае, если в модели $K(m,n)$ модифицируется M реберных функций, множество блокируемых векторов может содержать векторы с количеством нулей от $m+1$ до $m+M$. При этом множество векторов с $m+1$ нулем действительно будет представлять собой объединение множеств векторов, блокируемых в результате модификации каждой из функций отдельно. Присутствие среди блокируемых векторов с количеством нулей большим, чем $m+1$, является нежелательным побочным эффектом.

В докладе сформулировано условие, достаточное для того, чтобы вышеописанный побочный эффект не проявлялся: на любом векторе с $m+t$ нулями ($2 \leqslant t \leqslant M$) базовая и модифицированная модели отличались значениями менее, чем у t функций.

Литература. **1.** Романкевич В.А., Потапова Е.Р., Бахтари Хедаятоллах, Назаренко В.В. GL-модель поведения отказоустойчивых многопроцессорных систем с минимальным числом теряемых рёбер // Вісник НТУУ "КПІ". – Інформатика, управління та ОТ. 2006. №45. С.93–100. **2.** Романкевич В.А., Морозов К.В., Фесенюк А.П. Об одном методе модификации реберных функций GL-моделей // Радіоелектронні і комп'ютерні системи. 2014. №6 (70). С.95–99.

Романчук К.Г., Стефанишин Д.В.
Інститут телекомунікацій і глобального інформаційного простору НАН України, Київ, Україна

Про використання сценарного підходу та методу Байєса при оцінці ризиків системних аварій на гідровузлах

Гідровузли — інженерні об'єкти, які мають у своєму складі гідротехнічні споруди різного типу і призначення (греблі, водосховища, водоскиди, гідроелектростанції, шлюзи та ін.), що поєднані територіально й функціонально, являють собою складні природно-технічні системи, аварії на яких можуть мати катастрофічні наслідки. В багатьох випадках потенційна небезпека для населення, яке проживає в зонах аварійних впливів гідровузлів, може бути не меншою, ніж для людей, що живуть біля атомних чи хімічних виробництв, з якими спеціалісти й громадськість зазвичай пов'язують проблеми техногенної безпеки [1].

Практика показує, що аварії на гідровузлах носять системний характер і можуть виникати і розвиватися за різними сценаріями навіть у випадках, коли причини аварій, тип і конструкція споруди, обладнання, вид основи тощо були схожими [2]. Причинно-наслідкові відношення між різними подіями і станами, що призводили до аварій на гідровузлах, часто виявлялися занадто складними, щоб їх розглядати як статистичні факти в сукупності подібних випадкових подій або явищ, які можна аналізувати в рамках однієї моделі. В тій чи іншій мірі і ретроспективний аналіз аварій, що вже відбувалися на гідровузлах, і прогнозування можливих аварій, потребує аналізу індивідуалізованих сценаріїв їх виникнення і розвитку у формі певних логічних побудов та суб'єктивних припущень.

Сценарний підхід [3, 4] дозволяє здійснити декомпозицію складної задачі моделювання й прогнозування аварій на гідровузлах. В результаті сценарного моделювання неструктурована («слабо структурована») задача прогнозування аварії на гідровузлі зводиться до кількох (за кількістю модельних сценаріїв) «більш структурованих» задач, що можуть описуватися простішими математичними моделями з меншим числом визначальних факторів та параметрів тощо. Сценарне моделювання в поєднанні з методом Байєса дозволяє поєднувати можливості різних методів та моделей, здійснювати формалізоване узагальнення ризиків системної аварії, отриманих за різними модельними сценаріями. В роботі сформульовано основні узагальнення, означення і твердження, що відносяться до оцінки ризиків системних аварій на гідровузлах в рамках сценарного підходу. Показано, що сценарний підхід з використанням байєсівського перетворення ймовірностей дозволяє здійснити узагальнення повного (сумарного) ризику системних аварій на гідровузлі, які можуть відбуватися внаслідок довільних, в тому числі і сумісних, аварійних подій, призводити до різних наслідків, за деякими модельними, несумісними сценаріями виникнення й перебігу аварії, що формують повну групу подій.

Література. **1.** Векслер А.Б. Надежность, социальная и экологическая безопасность гидротехнических объектов: оценка риска и принятие решений / А.Б. Векслер, Д.А. Ивашинцов, Д.В. Стефанишин. — СПб.: ВНИИГ им. Б.Е. Веденеева, 2002. — 591 с. **2.** Стефанишин Д.В. Прогнозування аварій на греблях в задачах оцінки й забезпечення їх надійності та безпеки / Д.В. Стефанишин // Гідроенергетика України. — № 3-4. — 2011. — С. 52-60. **3.** Стефанишин Д.В. Сценарный подход к оценке вероятностей аварий на плотинах / Д.В. Стефанишин // Мониторинг. Наука и безопасность. №1 (9), 2013. – С. 26-33. **4.** Стефанишин Д.В. Логіко-імовірнісна оцінка ризику збитків від аварійного виливу води з басейну добового регулювання Зарамагської ГЕС-1 / Д.В. Стефанишин, Е.Г. Романчук // Системні дослідження та інформаційні технології. — 2013. №3. — С. 130-141.

Савін Є.В., Олефір О.С.
Національний технічний університет України "КПІ", факультет прикладної математики, Київ, Україна

Моделювання залежних збитків від нещасних випадків на сонові копула-моделей

Моделювання випадкових процесів лежить в основі страхового бізнесу. Основною ціллю передових актуарних розрахунків являється розробка та калібрування моделей, що найкраще повторюють реальні події. Саме тому при створенні повної та максимально точної моделі необхідно враховувати залежності, що існують між випадковими велечинами.

Одним з найпоширеніших способів, яким описували та продовжують описувати залежнесті випадкового типу, являється коефіцієнт лінійної кореліції. Проте гнучкість та точність таких моделей дуже слабка, адже залежність між випадковими величинами в більшості випадків не є лінійною.

Для моделювання складних типів залежностей використовують копули. Вони широко застосовані для оцінки фінансових ризиків і в страховому аналізі – наприклад для ціноутворення забезпечених боргових зобов'язань (CDOs). Також копули застосовуються до інших страхових задач як гнучкий інструмент. Нещодавно, копули булі успішно використані для формування бази даних для аналізу надійності автострадних мостів і для різноманітних моделювань з багатьма змінними в цивільному, механічному та шельфо-видобувному машинобудуванні [1].

Постановка задачі. Опираючись на реальні статистичні дані, побудувати копула-модель залежності збитків нанесених пожежою застрахованій будівлі, майну і матеріалам, що були в цій будувлі та збитків через припинення фінансової діяльності внаслідок пожежі.

Опис вхідних даних. Для побудови моделі ми використали реальні історичні дані страхової компанії X, по застрахованим від пожежі будівлям. Вони представляють собою таблицю з 2167 спостережень та з наступними колонками:

1. Дата;
2. Збиткі нанесені будівлі;
3. Збиткі нанесені майну та матеріалам;
4. Втрачений прибуток від використання будівлі;
5. Сумарні збитки спричинені пожежою.

Збитки в таблиці наведені в млн. данських крон (1 даньска крона дорінює приблизно 0,13 евро).

Результати розрахунків. Враховуючи курс данської крони та розміри збитків заданих незначними пожежами, було вирішено відібрати для розрахунків лише страхові випадки ціної білше 1 млн. (327 спостережень). Також для побудови двовимірної копули ми виділили дві випадкових величини: build - представляє збитки нанесені будівлі та other - сума двох інших видів збитків. Всі розрахунки та моделювання проводилися за допомогою безкоштовної статистичної мови програмування - R. Перший етап моделювання передбачає встановлення типів маргінальних розподілів та оцінку їх параметрів. Потенціальними кандидатами було обрано наступні розподіли: лог нормальний, Парето, Вейбул та Гамма. Ці розподіли мають гарні характеристики та найчастіше використовуються для моделювання збитків. Оцінка параметрів проводилась за допомогою методу максимальної вірогідності (Maximum Likehood Estimation, MLE) [2]. Отримані розподіли були порівняні з емпіричними данними за допомогою тесту Колмогорова-Смирнова [3]. Найкращими маргінальними розподілами для двох випадкових величин виявилися розподіли Парето. Оцінені параметри та результати тесту наведені на рисунках та в таблиці (рис. 1, 2, табл. 1).

Другий етап моделювання передбачає підбір та оцінку параметрів для копул. Для цього були обрані копули основних сімейств: Гаусівська та Стьюдента (еліптичні копули) для перевірки лінійних залежностей, Гамбела (екстремальна копула) та Франка (архімедова копула) для перевірки екстремальних залежностей. Їх параметри також були оцінені MLE методом.

Табл. 1. Обрані розподіли та результати КС тесту

Випадкова величина	Тип розподілу	Параметр 1 (Shape)	Параметр 2 (Scale)	D-stat	P-Value
build	Парето	2.693317	5.051742	0.0298	0.9333
other	Парето	1.342932	2.480009	0.0291	0.9451

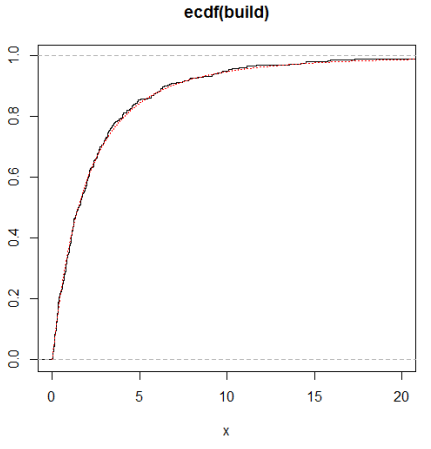

Рис. 1. Емпірична функція розподілу і розподіл Парето для build

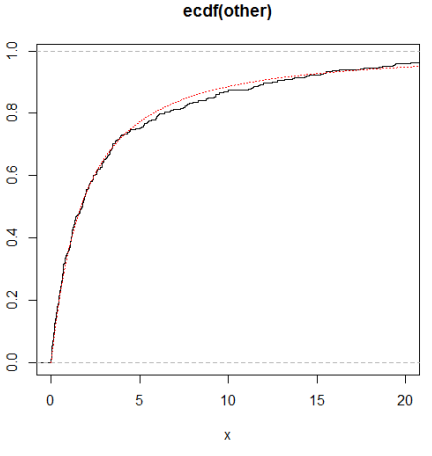

Рис. 2. Емпірична функція розподілу і розподіл Парето для other

Для оцінки якості моделей використано обрали коефіцієнт конкордації Кендала - τ [4]. Вибір оптимальної копули базувався на найближчому наближенні τ моделі до емпіричного (табл. 2).

Табл. 2. Коефіцієнти конкордації Кендала τ

Копула	Емпірична	Гаусівська	Стьюдента	Гамбела	Франка
τ	0.2259513	0.236616	0.236616	0.2157642	0.2160951

Найближче до емпіричного значення τ має копула Франка (оцінений параметр копули - α = 2.022223), тому оберемо її як модель, що можна використовувати для поставленої задачі.

Висновки.

1. Найкращими маргінальними розподілами для двох випадкових величин виявилися розподіли Парето;
2. Для моделювання залежних збитків найкраща копула-модель базується на копулі Франка;
3. Для обраних копул розраховані параметри для моделювання;
4. Ці моделі не є остаточними, необхідно провести детальний аналіз екстремальних значень і визначити поведінку моделей на хвостах розподілів. Це є ціллю подальших досліджень.

Література. **1.** Meneguzzo, David, «Copula sensitivity in collateralized debt obligations and basket default swaps», Journal of Futures Markets T.24, 2003.-pp. 37–70. **2.** Карташов М.В. «Ймовірність, процеси, статистика», К.: ВПЦ Київський університет, 2007.-сс. 148-163. **3.** Постовалов Н.С. «Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть II. Непараметрические критерии.» – М.: Изд-во стандартов, 2002.-с 64. **4.** Кобзарь А.И. «Прикладная математическая статистика.» — М.: Физматлит, 2006.-с 202.

***Стефанишин Д.В.*[1], *Ходневич Я.В.*[1], *Щодро О.Є.*[2]**
[1]*Інститут телекомунікацій і глобального інформаційного простору НАН України, Київ, Україна;*
[2]*Національний університет водного господарства та природокористування, Рівне, Україна*

Про ідентифікацію шорсткості природних русел для математичного моделювання руслових потоків в нижніх б'єфах гідроспоруд

Пропонується підхід для ідентифікації шорсткості природних русел в нижніх б'єфах гідроспоруд для заданого набору умов (витрата води, рівень вільної поверхні, геометрія русла). Ця характеристика визначає гідравлічний режим русел, його динаміку. Вона залежить від морфології русла, крупності наносів, наявності рослинності та різного роду перешкод в руслі тощо, а також від перебігу акумулятивно-ерозійної діяльності ріки, яка може суттєво змінюватися в часі на різних ділянках нижніх б'єфів гідроспоруд. Наприклад, донні гряди наносів можуть значно більше впливати на гідравлічний опір русла, ніж крупність частинок дна.

При розрахунках руслових потоків значення коефіцієнтів шорсткості, зазвичай, задають наближено, орієнтуючись на емпіричні дані гідравлічних довідників. Для умов природних русел такі дані є досить ненадійними. Тому пропонується при розв'язанні задач математичного моделювання руслових потоків, при можливості, визначати шорсткість русла зворотним шляхом, в рамках відповідної математичної моделі. При цьому ідентифікація шорсткості русла для різних наборів умов, що мали місце в ретроспективні моменти часу, дозволить встановити кількісні показники еволюції цього важливого гідравлічного параметра.

Відомо, що існує зв'язок між витратами і рівнями води руслових потоків. Для нижніх б'єфів гідроспоруд, завдяки моніторингу і матеріалам вимірювань, дані щодо основних характеристик потоку (витрат, рівнів води тощо), зазвичай, є відомими або ж легко встановлюються. Рівні води в створах залежать від витрат і геометричних параметрів русла та його морфологічних характеристик. І якщо глибину і ширину руслового потоку можна зрештою контролювати, то шорсткість русла в потрібних нам створах вимірювати безпосередньо ми не можемо. Інтенсивність руслових переформувань в нижніх б'єфах залежить, як відомо, від абсолютних значень швидкостей потоку на відповідних ділянках, які в свою чергу пов'язані з опором русла на цих ділянках. Відповідно і шорсткість потрібно знати пофрагментно для різних ділянок русла (біля берегів, на перекатах, поворотах тощо), яка в результаті місцевого і загального розмиву, відкладення продуктів розмиву, змінюється в часі, змінюючи опір русла.

Підхід, що нами пропонується, полягає у диференційованому врахуванні окремих факторів опору русла шляхом проведення детальних розрахунків елементів гідравлічної структури потоку. При відомому глобальному нахилі водної поверхні можна визначити глобальний коефіцієнт шорсткості всього русла шляхом обчислення вільної поверхні в рамках одновимірної моделі так, як це пропонується в [1]. Більш детальні розрахунки в рамках двовимірної задачі гідродинаміки дозволяють отримати розподіл швидкостей в плані [2]. В рамках цієї моделі при варіюванні коефіцієнта шорсткості враховується вплив руслових форм і зернистості русла на гідравлічний опір на заданій ділянці русла або окремих її фрагментах (біля берегів, на поворотах тощо). Тривимірне моделювання гідродинамічної структури потоку застосовується для обчислення розподілів швидкостей, характеристик турбулентності, перемішуючих властивостей та ін. у всьому об'ємі характерного фрагменту потоку [3]. Це дозволяє уточнити опір в зазначеній області руслового потоку.

Література. **1.** Ходневич Я.В. Задача обчислення рівнів вільної поверхні у руслі. Проблема визначення коефіцієнта шорсткості / Я. В. Ходневич // Зб. наук. праць «Вісник НУВГП». Серія «Технічні науки». – Рівне: НУВГП, 2010. – Вип. 3(51). – С. 87-94. **2.** Шеренков И.А. Прикладные плановые задачи гидравлики свободных потоков / И.А. Шеренков. – М.: Энергия, 1978. – 240 с. **3.** Ходневич Я. В. Математична модель динаміки водного потоку в місці обтікання гряди наносів / Я. В. Ходневич // XIV Міжнар. наук. конф. ім. акад. М. Кравчука. Т. 1. Диферен.та інтегр. рівняння, їх застосування. – К.: НТУУ «КПІ», 2012. – С. 428.

Стефанишина-Гаврилюк Ю.Д., Стефанишин Д.В.
Інститут телекомунікацій і глобального інформаційного простору НАН України, Київ, Україна

Про один метод прийняття рішень в природокористуванні з врахуванням ризику невикористаних можливостей на основі попарного порівняння альтернатив

Формальний підхід до прийняття рішень у природокористуванні передбачає порівняння варіантів (альтернатив) та вибір серед них оптимальних (квазіоптимальних, кращих, компромісних тощо) [1, 2].

Оскільки ефекти від природокористування реалізуються в майбутньому, то оцінка і порівняння відповідних альтернатив здійснюються в умовах невизначеності. З метою формалізації задачі прийняття рішення в умовах невизначеності використовується ризик, як інструмент її розкриття і подолання.

Сучасні національні стратегії екологічної й техногенної безпеки багатьох країн орієнтуються на концепцію прийнятного («ненульового», «практично досяжного мінімального» тощо) ризику в природокористуванні. При цьому приймається, що і відмова від природокористування теж може бути пов'язана з ризиком, зокрема з ризиком невикористаних (втрачених) можливостей [3], коли суспільство не використовуючи раціонально природних ресурсів, ризикує відстати в своєму поступальному розвитку.

Обґрунтовано необхідність врахування ризику невикористаних можливостей як складової повного ризику [4] при прийнятті рішень в природокористуванні за умов невизначеності та ризику на основі порівняння альтернатив. Сформульовано загальні положення й правила та наведено принципові узагальнення – означення і твердження, що стосуються прийняття рішень в природокористуванні з врахуванням ризику невикористаних можливостей на основі попарного порівняння альтернатив. Запропоновано функції повного ризику альтернатив у вигляді лінійних комбінацій, які поєднують системні (власні) ризики альтернатив та ризики невикористаних можливостей, які трактуються як виграшні ефекти альтернатив, що відкидаються. Сформульовано твердження про те, що множина оптимальних альтернатив за критерієм мінімального ризику як результат розв'язання задачі прийняття рішень з врахуванням ризику невикористаних можливостей співпадає з множиною оптимальних альтернатив за критерієм максимальної корисності. Наведено способи визначення компонент повного ризику альтернатив при заповненні таблиць рішень та запропоновано алгоритми вибору оптимальної альтернативи з врахуванням ризику невикористаних можливостей. Відмічено, що під ризиком невикористаних можливостей в природокористуванні слід розуміти не тільки втрачені ймовірні виграші (економічні та інші вигоди тощо), які можна отримати від використання природних ресурсів, а й відмову від зменшення негативних наслідків (збитків, шкоди, втрат), зниження величини штрафних та інших санкцій від впровадження більш безпечних технологій тощо.

Література. **1.** Бейко І.В. Задачі, методи і алгоритми оптимізації: навчальний посібник / І.В. Бейко, П.М. Зінько, О.Г. Наконечний. — Рівне: НУВГП, 2011. — 624 с. **2.** Мушик Э. Методы принятия технических решений / Э. Мушик, П. Мюллер / Пер. с нем. — М.: Мир, 1990. — 206 с. **3.** Ястремський О.І. Основи теорії економічного ризику / О.І. Ястремський. — К.: АртЕк, 1997. — 248 с. **4.** Stefanyshyn D.V. A method of decision making at risk in natural resources use by pairwise comparison of alternatives with taking account of risks of lost opportunities / D.V. Stefanyshyn, Yu.D. Stefanyshyna // Proc. of Int. Scientific School «Modelling and Analysis of Safety and Risk in Complex Systems». July 7-11, 2009. — S.-Petersburg, Russia. — P.P. 435-439.

Стефанович Т.О., Щербовських С.В.
Національний університет "Львівська політехніка", Львів, Україна

Аналіз причин непрацездатності системи із містковою структурою, в якій внутрішній елемент функціонує у ненавантаженому резерві

Запропоновано підхід для аналізу причин непрацездатності системи із містковою структурою, в якій внутрішній елемент функціонує у ненавантаженому резерві. Математичний опис системи виконано на основі динамічного дерева відмов, кількісний аналіз на основі марковської моделі.

Системи із містковою структурою є проміжним різновидом між системами із роздільним резервуванням та системами із загальним резервуванням. Іншими словами, систему із містковою структурою слід розглядати як систему із роздільним резервуванням, в якій внутрішній елемент, який з'єднує виходи елементів першої групи та входи елементів другої групи, не є абсолютно надійним. Відповідно, постає проблема як для такого елемента урахувати зміну його навантаження залежно від стану хордових елементів системи. Особливість проблеми полягає у тому, що можна виділити кілька різних алгоритмів, за якими відбувається зміна навантаження, що залежить від функціональних особливостей досліджуваної системи.

Метою дослідження є розроблення математичної моделі надійності, призначеної для аналізу причин непрацездатності системи із містковою структурою, в якій внутрішній елемент функціонує у ненавантаженому резерві. Новизна моделі полягає у тому, що в ній враховано як впливає перерозподіл навантаження у внутрішньому елементі такої системи на ймовірнісні характеристики причин її непрацездатності.

Розглянемо систему, яка складена із п'яти елементів (рис.1). Генератор G1 та трансформатор TV1 утворюють першу підсистему, а генератор G2 і трансформатор TV2 — другу. Підсистеми взаємно дублюють одна одну. Вважаємо, що явищами перерозподілу навантаження між елементами цих підсистем можна знехтувати. Між генератором G1 та трансформатором TV1 першої підгрупи і генератором G2 та трансформатором TV2 другої підгрупи включено силовий перемикач SA, який є внутрішнім елементом. Цей елемент функціонує за таким алгоритмом. Якщо система здатна виконувати свою функцію за допомогою елементів обох підсистем, то перемикач перебуває у ненавантаженому режимі. Якщо елементів обох підсистем недостатньо для виконання функції системи, то здійснюється вмикання цього елемента. Якщо одержана конфігурація забезпечує працездатність, то елемент переходить у навантажений режим, а у зворотному випадку настає непрацездатність системи. Вважаємо, що навантаження елементів основної та резервної системи не залежить від стану інших елементів. Тобто, якщо хордовий елемент працездатний, то для будь-якої працездатної конфігурації системи він працює у номінальному режимі. Засоби технічної діагностики та перемикання ідеальні, а зміна навантаження елементами миттєва. Напрацювання до відмови усіх елементів системи розподілено за законом Вейбулла, а тривалість ремонтування — експоненціально.

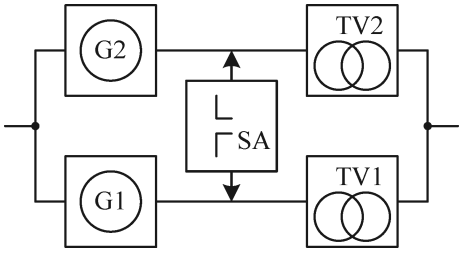

Рис. 1. Структурна схема системи

Методи, використані у дослідженні, докладно описані у роботах [1–5]. Для розв'язання поставленої проблеми надійність системи формалізовано динамічним деревом відмов (рис.2). Дерево відмов складене із двох частин. Перша частина задає логічну умову настання непрацездатності системи, а друга — логічні умови зміни навантаження. Щоб друга частина не впливала на умову непрацездатності системи, її включено через блок 0, який здійснює відсікання логічного сигналу. Зміна навантаження відбувається, якщо система переходить в один із двох контрольних станів. Перший контрольний стан полягає у тому, що генератор G1 та трансформатор TV2 непрацездатні, а трансформатор TV1 та генератор G2 працездатні.

Другий контрольний стан, навпаки, — генератор G1 та трансформатор TV2 працездатні, а трансформатор TV1 та генератор G2 непрацездатні.

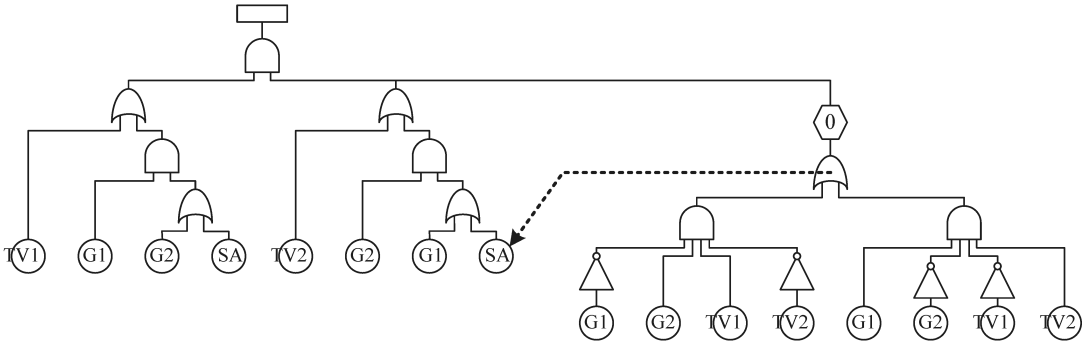

Рис. 2. Динамічне дерево відмов системи

Напрацювання до відмови усіх елементів системи розподілено за законом Вейбулла, а тривалість ремонтування — експоненціально.

На підставі динамічного дерева відмов побудовано граф станів та переходів системи (рис.3). Система перебуває у сімнадцяти станах, із яких дев'ять відповідають працездатності. Непрацездатні стани згруповано у чотири множини C1–C4, кожна із яких відповідає окремій причині непрацездатності системи. Перша множина C1 відповідає непрацездатності системи внаслідок відмови генераторів G1 та G2, друга множина C2 — внаслідок відмови трансформаторів TV1 та TV2, третя множина C3 — внаслідок відмови генератора G2, силового перемикача SA та трансформатора TV1 та четверта множина C4 — внаслідок відмови генератора G1, силового перемикача SA та трансформатора TV2.

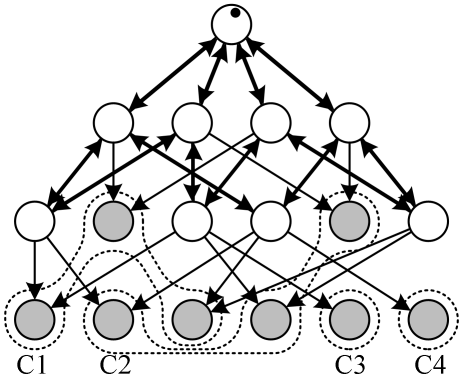

Рис. 3. Граф станів та переходів системи

На підставі графу станів та переходів системи побудовано розщеплену однорідну марковську модель, яка містить 595 диференціальних рівнянь. На основі марковської моделі можна обчислити ймовірнісні характеристики причин непрацездатності та для заданих параметрів визначити найімовірнішу із них.

Література. **1.** Щербовських С. В. Математичні моделі та методи для визначення характеристик надійності багатотермінальних систем із урахуванням перерозподілу навантаження: монографія / С. В. Щербовських. — Львів: Вид-во Львівської політехніки, 2012. — 296 с. **2.** Бобало Ю. Я. Математичні моделі та методи для аналізу надійності радіоелектронних, електротехнічних та програмних систем : монографія / Ю. Я. Бобало, Б. Ю. Волочій, О. Ю. Лозинський, Б. А. Мандзій, Л. Д. Озірковський, Д. В. Федасюк, С. В. Щербовських, В. С. Яковина. — Львів: Вид-во Львівської політехніки, 2013. — 300 с. **3.** Mandziy B. Mathematical model for failure cause analysis of electrical systems with load-sharing redundancy of component / B. Mandziy, O. Lozynsky, S. Shcherbovskykh // Przeglad Elektrotechniczny. — 2013. — Vol. 89, № 11. — P. 244–247. **4.** Стефанович Т. О. Математична модель для аналізу причин непрацездатності системи із роздільним полегшеним резервуванням / Т. О. Стефанович, С. В. Щербовських // Відбір та обробка інформації. — 2014. — № 41(117). — С. 28–35. **5.** Щербовських С. В. Розроблення моделі надійності захисної арматури, яка враховує вплив перерозподілу навантаження / Т. О. Стефанович, С. В. Щербовських // Східноєвропейський журнал передових технологій. — 2015. — Т. 1(73), № 3. — С. 37–44.

***Стецюк Р.В.*[1], *Валькман Ю.Р.*[2]**
[1]*Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина;* [2]*Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины, Киев, Украина*

Проектирование онтологии для правовой предметной области

Понятие онтологии является достаточно широким по степени формализации. Так, к онтологиям можно отнести даже неструктурированный словарь с определениями. Поскольку в текстах правовых актов достаточно широко практикуется введение определений (например, в преамбулах некоторых законов), построить словарь терминов конкретной отрасли права не так сложно, но такая простая онтология не будет эффективна в использовании. [1]

Проектирование онтологий в правовой сфере имеет следующие характерные черты [2]:

- большое количество общепринятых понятий с особой спецификой использования;
- существенное различие в структурах разных отраслей права;
- наличие промежуточного общетеоретического правового уровня между онтологией верхнего уровня и онтологией предметной области;
- большое количество теоретических допущений и абстрактных конструкций, зависящих от специфики правовых взглядов.

Первоначально, эксперты в предметной области определяют онтологию, используя базу знаний онтологии и инструменты редактирования *RDF*, язык *Protege* [3]. Затем, слова и фразы (например, речь, фрагменты текстов и аннотации) сопоставляются с соответствующими *концептами* онтологии области. Эксперты, далее, создают учебный набор *шаблонов*, используя простую и удобную в работе обработку ЕЯ и используя инструмент под названием *MontyLingua* [4]. После этого, вычисляются *вероятности концептов* в заданных фрагментах документа. Вероятности концепта рассчитываются в каждом конкретном документе, а затем используются для группировки шаблонов. Таким образом, иерархический алгоритм кластеризации уточняется, адаптируясь к процессу вывода онтологического понятия.

В случае с правовым материалом, особенно при построении отраслевой онтологии, основной текстовый массив совпадает с законодательной базой данной отрасли, т.е. представляет собой четко определенный перечень правовых актов, доступ к которым обеспечен всем гражданам. Несмотря на разнообразие правовых актов и возможные недостатки юридической техники, тексты правовых актов в идеальном случае всегда должны отличаться свойствами, которые облегчают формализацию предметной области. Среди таких свойств можно назвать следующие:

- нейтральность (отсутствие экспрессивности),
- безличный характер,
- связность,
- последовательность,
- точность мысли,
- наличие большого количества стандартных оборотов,
- сильно выраженная структурированность текста.

Во многих случаях для решения прикладных задач онтологического инжиниринга в правовой сфере удобно использовать некоторую общую базовую юридическую онтологию, которая разрабатывается без учета конкретной решаемой задачи и которая может быть легко расширена. В настоящий момент *не существует* такая единая общеприменимая правовая онтология, в то время как опыт зарубежных исследователей показывает, что создание такой онтологии не только возможно, но и необходимо. Такая онтология могла бы быть определена в качестве основы для разработки практически применимых онтологий для решения задач в конкретных отраслях права.

Представляется нецелесообразным проводить всю работу по формализации правовой сферы без использования данных других онтологических проектов. Правовая онтология, которая является частью общей архитектуры для систем правовых знаний, которые обеспечивают обмен

знаниями между существующими системами правовых знаний (это формат обмена правовыми знаниями – *LKIF* [5]). *LKIF* выполняет две основные роли: *во-первых*, включает перевод между правовыми базами знаний, представленными в различных форматах и формах и, *во-вторых*, представляет формализм знаний, который является частью большой архитектуры для разработки правовых систем знаний.

Эти случаи использования *LKIF* приводят нас к классическому компромиссу между уступчивостью и выразительностью. Дополнительные требования заключается в том, что формат *LKIF* должен соответствовать текущему *Semantic Web* стандарту для включения службы правовой информации через web из комбинации *OWL DL* и *SWRL*, предлагая гибридное решение задачи.

За основу для построения онтологии взята *«родительская»* онтология *LKIF* [5]. Онтология *LKIF* не только обеспечивает теоретическое понимание правовой области, но ее прагматическое использование - как инструмент, чтобы облегчить приобретение знания, обмена и его представления (в частности, формализовать части существующего законодательства).

Онтология *LKIF* должна содержать *«фундаментальные понятия закона»*. От потенциальных пользователей зависит, какой словарь будет использоваться. Выделено и идентифицировано три главных группы пользователей: граждане (непрофессионалы), юристы и правоведы. И, хотя юристы используют правовой словарь более грамотно и тщательно, чем непрофессионалы, для большинства из используемых терминов по-прежнему существует некоторое общее понимание (и можно рассматривать соответствующие термины как более или менее схожие).

Вместе с тем множество основных терминов имеет определенное юридически-техническое значение, например, *«ответственность»* или *«правовой факт»*. В роли стратегии моделирования была выбрана система *METHONTOLOGY* [6]. *METHONTOLOGY*, как подход к построению и сопровождению онтологий, разработан Асунсьон Гомез-Перез с коллегами. В рамках этой методологии реализуются принципы Тома Грубера, а также разработан программный комплекс спецификации онтологии *ODE* (*ODE - Ontology Design Environment*).

Данный подход выделяет следующие процедуры в *«жизненном цикле»* создания онтологии: управление проектом, собственно разработка и поддержка разработки. Процедуры управления проектом включают *планирование*, *контроль* и *гарантии качества*.

Планирование определяет, *какие задачи должны быть выполнены, как они организуются, как много времени и какие ресурсы нужны для их выполнения.*

Контроль гарантирует, что *запланированные задачи выполнены и именно так, как это предполагалось.*

Гарантии качества нужны для того, чтобы быть *уверенным в том, что компоненты и продукт в целом находятся на заданном уровне.*

Таким образом, актуальной является следующая задача: на основе базовой юридической онтологии *LKIF* создать юридическую онтологию верхнего уровня применительно к системе права с использованием русскоязычных концепций.

Литература. **1.** Соловьев В. Д. Онтологии и тезаурусы / В. Д. Соловьев, Б. В. Добров, В. В. Иванов, Н. В. Лукашевич – Казань: Казанский государственный университет, 2006. – 157 с. **2.** Деринг Ю. Онтологии в правовой сфере. / Ю. Деринг, М. Хербергер // Научно-правовой журнал «Проблемы правовой информатизации» – 2006. – № 2(12) . **3.** Wu Z. and Palmer M. Verb semantics and lexical selection.- in Proc. 12nd Anna. Meeting Assoc. Comput. Linguist. Las Cruccs. NM. Jun.27-30. 1994. pp. 133-138. **4.** Witten I. H. Adaptive text mining: Interring structure from sequences/ Discret Algorithms, vol. 2. no. 2. Jun. 2004. pp. 137-159. **5.** LKIF Ontology. A core ontology of basic legal concepts. – (http://www.estrellaproject.org/lkifcore). **6.** Fernбndez-Lypez, M., Gymez-Pйrez, A. Overview and Analysis of methodologies for building ontologies. Knowledge Engineering Review (KER). Vol. 17[2]. – 2002.

Ус С.А., Станіна О.Д.
Національний гірничий університет, Дніпропетровськ, Україна

Алгоритм розв'язування задачи оптимального розбиття множин з додатковими зв'язками

Розглядається задача оптимального розбиття множин з додатковими зв'язками за наявністю обмежень. Запропоновано алгоритм вирішення такої задачі.

Останнім часом, все більша кількість дослідників звертає свою увагу на вирішення різноманітних проблем в області економії матеріальних і природних ресурсів. Одним з можливих варіантів таких задач можуть буть задачі розміщення виробництва з урахуванням транспортування сировини. В даній роботі розглядаються задача оптимального розбиття множин з додатковими зв'язками. Змістовну постановку задача оптимального розбиття множин з додатковими зв'язками можна сформулювати у такий спосіб: необхідно поділити область на зони обслуговування для підприємств-виробників, таким чином, щоб сумарні витрати на доставку сировини і продукції були мінімальні. Передбачається, що місця розташування підприємства-виробників і підприємств-постачальників заздалегідь відомі, а об'єм доставляємої продукції визначається в ході вирішення задачі. Математична постановка такого виду задач можна представити у вигляді:

$$\sum_{i=1}^{N} \int_{\Omega_i} c_i(x,\tau_i)\rho(x)\,dx + \sum_{i=1}^{N}\sum_{=1}^{M} c_{ij}(\tau_i,\tau_j)\upsilon_{ij} \tag{1}$$

при обмеженнях

$$(\Omega_i \cap \Omega_i) = 0, i \neq j, \tag{2}$$

$$\bigcup \Omega_i = \Omega, \tag{3}$$

$$\sum_{i=1}^{N} \upsilon_{ij} = b_j, j = 1,2,...M, \tag{4}$$

$$\sum_{j=1}^{M} \upsilon_{ij} = \int_{\Omega_i} \rho(x)\,dx \tag{5}$$

де $c_i(x,\tau_i)$ - вартість доставки до виробництва, $\rho(x)$- попит, $c_{ij}(\tau_i,\tau_j)$ - вартість доставки від виробництва до постачальників, υ_{ij} - об'єм доставляємої продукції, b_j - потужність підприємств-постачальників.

Задача оптимального розбиття множин з додатковими зв'язками в такій постановці є комбінуванням двоетапної задачі розміщення виробництва, а також задач розміщення-розподеління. Крім того, вона представляє собою частковий випадок задач ОРМ [1]. Необхідно зазначити, що подібні задачи розглядалися в [2]. Особливістю безпосередньо даної задачі є зв'язність задачи ОРМ, та транспортної задачі. В роботі запропоновано алгоритм, які базуються на використанні теорії оптимального розбиття множин та r-алгоритму Шора. Характерною особливістю даного алгоритму є врахування додаткових зв'язків між підприємтсвами-виробниками та підприємствами-постачальниками. Представлений алгоритм може бути використаний для вирішення транспортних задач, задач розподілення, задач формування портфелю замовлень та задачі об'ємно-календарного планування.

Література. 1. Киселева Е.М., Шор Н.З., Непрерывные задачи оптимального разбиения множеств: теория, алгоритмы, приложения: Монография – К.: Наукова Думка 2005 – 564 с. **2.** Ус С.А., Станина О.Д., О математических моделях многоэтапных задач размещения предприятий, Питания прикладної математики і математичного моделювання: зб.наук.пр./редкол. О.М. Кісельова (відп. ред.) та ін. – Д.: Вид-во «Ліра», 2014, с.258-268

Шаронова Н.В., Козуля М.М.
Національний технічний університет "Харківський політехнічний інститут", Харків, Україна

Інформаційні та об'єктивні загальносистемні особливості функціонування природно-техногенних систем

Актуальність дослідження. Сучасна теорія систем передбачає існування системних утворень у вигляді соціально-економічних, соціально-екологічних, еколого-економічних систем і об'єктів природно-техногенних об'єктів, що відносять до складних систем, з урахуванням положень сталого розвитку їх об'єднують у соціально-еколого-економічні об'єкти дослідження [1,2]. Особливим для таких систем є інформаційна компонента, яка враховує якісні характеристики як окремих елементів, так і їх відношень. Нестійкий стан системи призводить до криз, які умовно прирівнюють до якісних перетворень у системі. Отже, актуальним стає у дослідженні сучасних складних систем обробка інформаційних даних відповідно до стандартного системного підходу і новітніх методів теорії інформації.

Постановка задачі дослідження. Мета даної роботи спрямована на визначення комплексу системних методів пізнання складних системних об'єктів для підвищення ефективності застосовуваних інформаційних технологій інтелектуальної обробки даних і якості представлення результатів для прийняття рішень урегулювання соціально-еколого-економічних відносин на основі ідентифікації знань.

Методика дослідження і аналіз результатів. Якісні зміни, що відбуваються у кризових станах, визначають як відмінні (розвал системи), так і позитивні (створення нової системи). Загалом при наявності порушень природного порядку у навколишньому середовищі відбуваються зміни з погіршенням якості, а для створення нової високої якості необхідним стає управління, ефективність якого безпосередньо ґрунтується на знанні про систему. Саме на створення такої ефективної системи управління спрямовані дослідження стосовно ідентифікації ситуації і оцінки відповідності вимогам якості для «стан системи – процес – трансформаційні зміни – стан системи». Об'єктом визначено за концепцією сталого розвитку соціально-еколого-економічні утворення будь-якого рівня дослідження і призначення. Для вирішення задачі дослідження запропоновано комплексне використання послідовного аналізу даних за методом компараторної ідентифікації і ризик-оцінки стану природно-техногенних систем за умови пріоритетної їх відповідності екологічній якості [3].

Екологічність системи складається з таких основних етапів:

- виділення економічного, екологічного і соціального аспектів аналізу;
- визначення перебігу процесів відповідно до їх стабілізуючої дії на основі використання положень теорії ентропії, синергетики.

Для комплексної оцінки екологічності територіально-об'єктових систем запроваджена компараторна ідентифікація у вигляді функції відповідності як кількісне визначення за двома параметрами – 0 і 1, що дозволяє поєднати зміни у стані систем і об'єкта й імовірності порушення зв'язків у середовищі [3,4].

За результатами ідентифікації стану систем і відповідності умовам стійкості процесів, що відбуваються в об'єктах дослідження, за методом компаратора екологічності надають узагальнюючу характеристику об'єкта на основі ризик-аналізу. На останньому етапі реалізації комплексного оцінювання визначається стан усіх трьох складових сталого розвитку, виділяються процеси стабілізації і дестабілізації в аналізованих об'єктах навколишнього середовища, які впливають на стійкість систем.

Реалізація методики комплексної оцінки якості навколишнього природного середовища (НПС) розглянуто на прикладі дослідження екологічного стану техногенно-навантажених ландшафтно-геохімічних комплексів з використанням запропонованої програми реалізації аналітичного комплексу оцінки екологічності системних об'єктів дослідження на основі моніторингової інформації (рис. 1).

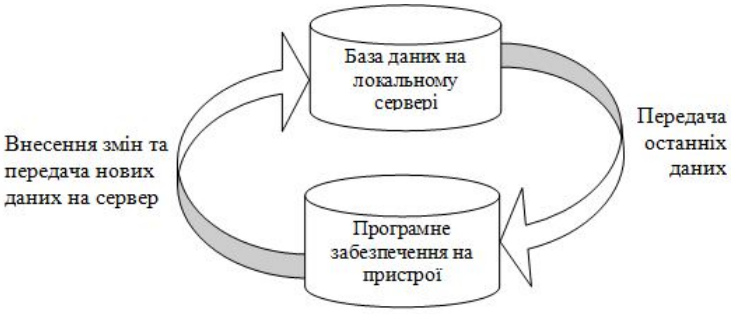

Рис. 1. Принцип роботи програмного продукту

Прикладом реалізації наданої методики є результати оцінки екологічності природно-техногенних систем, що входять до складу ландшафтно-геохімічних комплексів території Зміївського району. За наданою методикою розрахунки показників стану ґрунтів і оцінки їх екологічності представлені у програмі, розробленої для операційної системи Android (рис. 2).

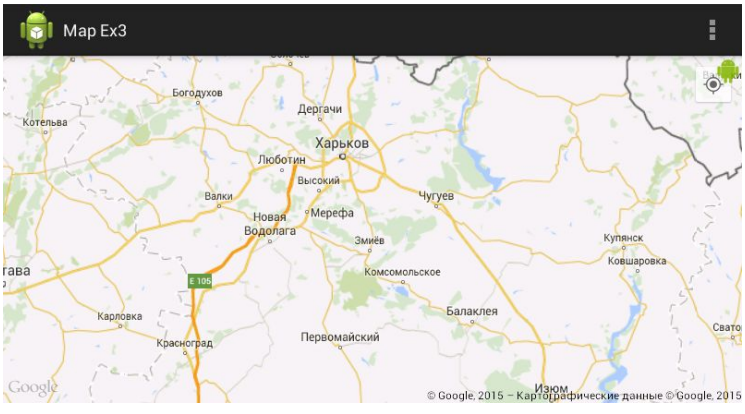

Рис. 2. Територіальне визначення точок дослідження

Висновки. Таким чином, у роботі обґрунтована доцільність комплексного використання методу компараторной ідентифікації для оцінки рівня екологічності природно-техногенних об'єктів за характеристикою їх стану й процесів у них з метою виявлення факторів дестабілізації систем і визначення ризик-ситуації з метою подальшого прийняття рішень щодо її стабілізації. Запропоновано програмну реалізацію даного підходу з дослідження складних об'єктів з метою автоматизації комплексної обробки вхідної інформації і отримання знань про стан системи і її складових, відповідності певним вимогам, основні фактори стабільного розвитку чи дестабілізації системи.

Література. **1.** Згуровский М. З. Глобальное моделирование процессов устойчивого развития в контексте качества и безопасности жизни людей / М. З. Згуровский, А. Д. Гвишиани. – К.: Політехніка, 2008. – 331 с. **2.** Згуровський М. З. Основи системного аналізу / М. З. Згуровський, Н. Д. Панкратова. – К.: Видавнича група BHV, 2007. – 544 с. **3.** Козуля Т. В. Процеси екологічного регулювання. Концепція корпоративної екологічної системи: монографія / Т. В. Козуля. – Харків : НТУ «ХПИ», 2010. – 588 с. **4.** Sharonova N. V. Entropy as Substratum of identifying the Corporative Ecological system (CES) condition / N. V. Sharonova, T. V. Kozulia // Вестник Херсонского национального технического университета. – 2008. – № 2 (31). – С. 518–527.

Шелест А.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

До побудови стратегії сталого розвитку економіки України за допомогою кластерного аналізу

На сьогодняшній день Україна переживає найважчі часи з моменту отримання незалежності. Окрім вже «традиційних» проблем, таких як, технологічна відсталість, корупція, важкі умови ведення бізнесу, низька якість освіти і медицини, бюрократія в органах державного апарату, з'явились нові, не менш серьозні проблеми – військовий конфлікт на сході України та втрата контролю над значною частиною території.

Розглянуті питання вимагають негайного вирішення, що ускладнюється важкими геополітичними та екномічними умовами. Саме тому сьогодні як ніколи важливо вибрати правильну стратегію розвитку.

На початку 2015 року президентом України було схвалено стратегію сталого розвитку «Україна – 2020», яка визначає цілі, основні напрями руху, реформи та програми, що планується впровадити до 2020 року [1].

В даній статті основна увага приділяється економічному аспекту сталого розвитку. Основною метою є висвітлення розробленої автором методології, яка полегшує задачу побудови стратегії сталого розвитку економіки України. Під сталим розвитком економіки будемо розуміти такий розвиток, що забезпечує макроекономічну стабільність та стійке зростання економіки.

В роботі [2] була представлена формальна модель стійкості економіки з використанням поняття економічної рівноваги, як вектора з компонент, які характеризують рівновагу в різних секторах. Але в результаті застосування такого підходу виявилось, що найбільш стійкі світові економіки не є стійкими у класичному математичному сенсі. Тим не менш, запропонований підхід дає можливість оцінити міру нестійкості економіки шляхом визначення інтегрального показника, що має пряму кореляцію з індексом сталого розвитку.

Найбільша проблема класичного математичного підходу полягає в тому, що поведінка побудованої формальної моделі стійкості може суттєво відрізнятись від поведінки реальної економіки. Це відбувається, перш за все, за рахунок того, що на процес розвитку економіки впливає дуже велика сукупність факторів, характер впливу яких може бути досить скланим та постійно змінюватись протягом часу.

Фінансово-економічна криза 2008 року показала неспроможність існуючих економічних моделей описати динаміку реальної економіки. На думку автора причина полягає у тому, що ані дослідники, ані політичні діячі не взмозі виявити і оцінити вплив на екеноміку одразу усіх ключових факторів, зважаючи на їхню кількість та складний характер впливу.

В роботі застосовується кластерний підхід до аналізу стійкості економіки, що призводить до більш чіткого розуміння основних економічних процесів всередині кластеру та ключових факторів впливу на нього. З іншого боку такий підхід враховує важливі міжгалузеві зв'язки, які відіграють важливу роль в економіці.

Очевидно, що виведення економіки України на рівень світової конкурентоспроможності є одною з головних цілей стратегії. Зважаючи на це, в даному дослідженні проводиться детальний аналіз кластеру ІКТ (інформаційно-комунікаційних технологій), як одного з найбільш перспективних та конкурентоспроможних. Результатом аналізу є список рекомендованих дій, які потрібно впровадити для того, щоб у найкращий спосіб забезпечити стійкість економічного розвитку і високу конкурентоспроможність кластеру ІКТ.

Література. **1.** http://www.president.gov.ua/documents/18688.html **2.** К построению модели равновесия экономической макросистемы / Панкратова Н.Д., Шелест А.В. // Системні дослідження та інформаційні технології. — 2013. — № 3. — С. 77–88.

Шумада К.О.
Національний технічний університет України "КПІ", факультет інформатики та очислювальної техніки, Київ, Україна

Цифрова система управління штучної вентиляції легенів

Система управління - це з'єднання окремих елементів в певну конфігурацію, що забезпечує задані характеристики. В основі її аналізу лежить теорія лінійних систем, що передбачає наявність причинно-наслідкових зв'язків між елементами. Тому процес або об'єкт, який піддають керуванню, може бути представлений у вигляді певного блоку. Зв'язок між входом і виходом - це, по суті, перетворення одного сигналу (причини) в інший (наслідок), при чому доволі часто з підсиленням потужності.[1]

У зв'язку з постійним ускладненням об'єктів керування і бажанням добитись оптимальних показників якості, за останні десятиріччя різко підвищилась роль автоматичного управління, в тому числі і в медицині.

ШВЛ - метод, за допомогою якого відновлюються і підтримуються порушена функція легенів - вентиляція і газообмін. Її суть полягає у примусовому введенні повітря в легені. Вона застосовується у випадках зупинки дихання, а також за наявності неправильного чи майже непомітного дихання. ШВЛ покликана вирішувати завдання, яке у нормі виконують дихальні м'язи. Це завдання включає в себе забезпечення оксигенації та вентиляції (видаленні вуглекислого газу) легенів пацієнта.[2]

Найефективніше ШВЛ проводити за допомогою спеціальних апаратів для штучного дихання. Для підтримки оптимальної роботи таких апаратів розроблено систему управління ШВЛ. Мета полягає в синтезі системи з підтриманням тиску введення повітря певного об'єму з певною концентрацією кисню. Для середньостатистичної людини ці значення будуть становити відповідно 0.3 кПа, 500мл, а концентрація кисню в повітрі має становити близько 40 відсотків, проте ці значення підбираються лікарем індивідуально для кожного пацієнта. Власне контролю піддається два значення: тиск та об'єм. Структура системи містить наступні елементи:

1. Об'єкт керування
2. Блок задання еталонних значень
3. Блок управління (мікроконтролер)
4. Виконуючі механізми (на лінії вдоху і лінії видиху)
5. Клапани (на лінії вдоху і лінії видиху)
6. Блок підготовки суміші (змішувач, фільтр, зволожувач)
7. Датчик потоку (на лінії видиху)
8. Датчик тиску (на лінії видиху)
9. Блок узгодження
10. Блок відхилення

Для досягнення бажаних показників якості дану систему синтезуємо та налаштовуємо потрібним чином. На виході отримаємо цифрову систему управління з дотриманням конкретних оптимальних значень тиску та об'єму, можливістю корегування та підлаштовування під конкретного пацієнта, тобто з хорошою компенсацією збурень, бажаним видом реакції на вхідний вплив, адекватними вихідними сигналами виконавчого пристрою, малою чутливістю до зміни параметрів та робастністю.

Розвиток систем управління іде шляхом вдосконалення їх гнучкості і забезпечення високого ступені автономності. Вони володіють достатньою можливістю до пристосування, проте з участю оператора. Щоб зменшити навантаження на людину і підвищити ефективність її роботи, ведуться інтенсивні дослідження в області супервізорного управління, людино-машинного інтерфейсу і керування комп'ютерними базами даних. Багато досліджень пов'язані також з покращенням методів передачі інформації і подальшим розвитком мов програмування.

Література. **1.** Р. Дорф, Р. Бишоп Современные системы управления 2002, Москва **2.** http://intranet.tdmu.edu.ua/data/kafedra/internal/biofiz

Яковлева А.П., Курдуп І.О.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Задача про розподіл ресурсів між підприємствами

Поставлено та вирішено методом динамічного програмування Беллмана [1]. задачу про розподіл ресурсів між підприємствами. Запропоновані різні варіанти постановки задачі, враховуючи постановку власних функцій віддачі та врахування часу.

Задача розподілу ресурсів, як і будь-яка інша задача динамічного програмування, є багатоетапною. Тобто вона розпадається на ряд послідовно розв'язуваних задач з меншою кількістю змінних. Оптимальна стратегія, за принципом оптимальності Беллмана, володіє властивістю, що якими б не були початкові умови і початковий стан, подальші розв'язки повинні становити оптимальну поведінку відносно стану, отриманого в результаті першочергового рішення [1]. Математично цей проинцип описується наступним чином:

$$f_{n-l}(S_l) = optimum[R_{l+1}(S_l, U_{l+1}) + (f_{n-l-1}(S_{l+1})], \quad (1)$$

де U_l - керування, обране на першому кроці, S_l - стан системи на першому кроці, R_l - безпосередній ефект, досягнутий на першому кроці, f_{n-1} - оптимальне значення ефекту, що досягається за $n-1$ кроків.

Постановка задачі. Нехай є N підприємств, між якими розподіляються ресурси v_1, v_2 і т. д. в сумарній кількості V. По закінченню робочого терміну вони повертають прибуток у розмірі відповідно $P_1(v_1)$, $P_2(v_2)$, ... Треба знайти оптимальну стратегію розподілу ресурсів, тобто визначити, скільки ресурсів необхідно виділити кожному підприємству для отримання максимального сумарного прибутку.

На кожному n-му кроці розв'язку задачі для n-го підприємства, для кожного можливого виділеного йому об'єму ресурсів v визначається його результуюча функція $P_n(v)$ за основним функціональним рівнянням Беллмана [2]. Важливим є той факт, що на кожному етапі ця функція Беллмана стає все більш складною, що викликає необхідність обчислень за допомогою ЕОМ. Функції $P_n(v)$ можуть бути різної природи. Так, вони можуть бути виражені дискретною вибіркою в залежності від керуючої функції v, можуть набувати лінійних, експоненційних, тригонометричних чи інших залежностей. Зокрема функції вигляду $P_n(v) = K * \sqrt{v}$, де – певний невід'ємний коефіцієнт продуктивності, набувають значної розповсюдженості в економічних задачах, адже вона є невід'ємною, зростаючою, забезпечує наявність екстремальних точок досліджуваної функції всередині відрізка зміни аргумента і задовольняє економічний закон убуваючої ефективності. Керуюче діяння v може набувати більш складних форм. Зокрема, його можна представити у вигляді :

$$v = a * s + b * t, \quad (2)$$

Тут s – грошові ресурси, t – час, a,b – нормуючі коефіцієнти. Подібна функція необхідна, якщо час виконання роботи підприємством є важливим. В залежності від розподілених на попередніх кроках ресурсів будуються наступні кроки.

Можливе також створення часових етапів роботи підприємств і оптимізація розподілення ресурсів на кожному етапі. В цьому випадку на кожному часовому етапі з загальної суми ресурсів V виділяється частина, що розподіляється між підприємствами. Проте в цьому випадку необхідно проводити для кожного етапу умовну оптимізацію, в процесі якої проводиться обчислення функцій Беллмана та умовно-оптимальних керувань.

Висновок. Як результат було створено програмний продукт на C Sharp, що реалізує розв'язок задачі розподілу ресурсів для різних випадків функцій віддачі підприємств та критерію часу.

Література. **1.** Беллман, Р. Динамічне програмування - М.: Вид-во Иностранная литература, 1960 г. **2.** Беллман, Р., Калаба Р. Динамическое програмирование и современная теория управления - М.: Изд-во Наука, 1969 г.

Ясинський В.К., Юрченко І.В.
Чернівецький національний університет ім. Юрія Федьковича, Чернівці, Україна

Поведінка сильного розв'язку стохастичних рівнянь в частинних похідних з марковськими параметрами

Після введення поняття стохастичного диференціала та інтеграла, визначення сильного розв'язку стохастичного диференціального рівняння у відомих монографіях [1–3] та їх подальше розповсюдження на класи стохастичних диференціально-функціональних рівнянь [4] (див. наведену велику бібліографію в цих роботах) стало можливим дослідження асимптотично сильного розв'язку для стохастичного диференціального рівняння з частинними похідними (СДРЧП) з урахуванням випадкових параметрів. (див. [5], [6], [7] та ін.).

Розглянемо стохастичний експеримент з базовим імовірнісним простором $(\Omega, \mathcal{F}, F, \mathbb{P})$, $F \equiv \{\mathcal{F}_t, t \geqslant 0\}$ - фільтрація, $\mathbb{E}\{\cdot\}$ — математичне сподівання, $T \subset [0, \infty)$. Простір функцій $u(t,x,\omega) : [0,T] \times \mathbb{R}^1 \times \Omega \to \mathbb{R}^1$, вимірних за t та x з імовірністю одиниця відносно σ-алгебри борелевих множин фазового простору $\mathcal{B}([0,T], \mathbb{R}^1)$, для яких існує невласний інтеграл $\int\limits_{-\infty}^{+\infty} \mathbb{E}\{|u(t,x,\omega)|^2\}dx < \infty$, позначимо $\mathfrak{M}_T$ і введемо для нього відповідні норми [4,5].

Позначимо через $Q(A(\cdot),q,p) \equiv \sum\limits_{k=1}^{n}\sum\limits_{j=1}^{m} a_{kj}(t,\xi(t))q^k p^j$, $Q(B(\cdot),q,p) \equiv \sum\limits_{k=1}^{n}\sum\limits_{j=1}^{m} b_{kj}(t,\xi(t))q^k p^j$, $Q(C(\cdot),q,p) \equiv \sum\limits_{k=1}^{n}\sum\limits_{j=1}^{m} c_{kj}(t,\xi(t))q^k p^j$, де $A(\cdot)$, $B(\cdot)$, $C(\cdot)$ – матриці розмірності $n \times m$, що містять відповідні берові функції, які залежать від t. Розглядаємо на $(\Omega, \mathcal{F}, F, \mathbb{P})$ задачу Коші для СДРЧП вигляду

$$\frac{\partial}{\partial t}\left[Q\left(A(t,\xi(t)), \frac{\partial}{\partial t}, \frac{\partial}{\partial x}\right)u(t,x,\omega)\right] =$$

$$= f(\beta(\omega))\left[Q\left(B(t,\xi(t)), \frac{\partial}{\partial t}, \frac{\partial}{\partial x}\right)u(t,x,\omega) + \frac{\partial}{\partial t}\left[Q\left(C(t,\xi(t)), \frac{\partial}{\partial t}, \frac{\partial}{\partial x}\right)u(t,x,\omega)\right]\frac{dw(t,\omega)}{dt}\right],$$

$$Q\left(A(t,\xi(t)), \frac{\partial}{\partial t}, \frac{\partial}{\partial x}\right)u(t,x,\omega)\Big|_{t=0} = [Qu]_0,$$

$f(\cdot)$ — берова функція з областю значень $\mathbb{R}^1$, $\beta(\omega)$ — випадкова величина, задана щільністю $p_\beta(x)$ (або функцією розподілу), $\xi(t) \equiv \xi(t,\omega) \in \mathbb{Y}$ для довільного $t \geqslant t_0$, $\omega \in \Omega$ — стохастично неперервний феллерів марковський процес з неперервними справа реалізаціями на компактному фазовому просторі $\mathbb{Y}$, $w(t,\omega)$ — одновимірний стандартний вінерів процес.

Отримані результати щодо поведінки в середньому квадратичному сильного розв'язку даного рівняння.

Література. **1.** Гихман И.И., Скороход А.В. Стохастические дифференциальные уравнения и их применение.– К.: Наук. думка, 1980.– 612 с. **2.** Гихман И.И., Скороход А.В. Управляемые случайные процессы.– К.: Наук. думка, 1977.– 251 с. **3.** Гихман И.И., Скороход А.В. Стохастические дифференциальные уравнения с частными производными.– К.: Ин-т математики АН УССР .– 1981.– С.25–59. **4.** Царьков Е.Ф., Ясинский В.К. Квазилинейные стохастические дифференциально-функциональные уравнения.– Рига: Ориентир, 1992.– 301 с. **5.** Перун Г.М., Ясинский В.К. Исследование задачи Коши для стохастических уравнений в частных производных // Укр. мат. журн.– 1993.– Т.45, № 9.– С.1773–1781. **6.** Donez N. P., Yurchenko I. V., Yasynskyy V. K. Mean Square Behavior of the Strong Solution of a Linear non-Autonomous Stochastic Partial Differential Equation with Markov Parameters // Cybernetics and System Analysis.- 2014.- Vol.50., №6.- P.930-939. **7.** Koroliuk V. S., Yurchenko I. V., Yasynskyy V. K. Behavior of the Second Moment of the Solution to the Autonomous Stochastic Linear Partial Differential Equation with Random Parameters in the Right-Hand Side // Cybernetics and Systems Analysis.- 2015.- Vol. 51, Issue 1.- PP.56-63.

This page was left blank intentionally

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Plenary talks

System analysis of complex systems of various nature

Intelligent systems for decision-making

High Performance Computing and Microsystems Engineering

Progressive information technologies

2 Intelligent systems for decision-making

Section 2

Intelligent systems for decision-making

1. Intelligent decision-making systems (IDMS) in finance-economical area (micro- and macro economical systems, banks, stock exchanges, insurance companies, etc.).
2. Decision-making systems in social processes' management.
3. Decision-making systems in technological processes' management in industry.
4. Intellectual analysis of data and knowledge; problems of data and knowledge mining, knowledge bases for IDMS.
5. Mathematical modelling and forecasting of complex objects and processes.
6. Decision-making under the data uncertainty conditions (systems of fuzzy logical deduction).
7. Modern methods and algorithms for IDMS (genetic and evolutional algorithms, neural networks, etc.).

Секция 2

Интеллектуальные системы принятия решений

1. Интеллектуальные системы принятия решений (ИСПР) в финансово-экономической сфере (микро- и макроэкономические системы, банки, биржи, страховые компании и т.п.).
2. Системы принятия решений в управлении социальными процессами.
3. Системы принятия решений в управлении технологическими процессами в промышленности.
4. Интеллектуальный анализ данных и знаний; проблемы добывания данных и знаний (Data&Knowledge Mining), базы знаний для ИСПР.
5. Математическое моделирование и прогнозирование сложных объектов и процессов.
6. Принятие решений в условиях неопределенности данных (системы нечеткого логического вывода).
7. Современные методы и алгоритмы ИСПР (генетические и эволюционные алгоритмы, нейронные сети и т.п.).

Секція 2

Інтелектуальні системи прийняття рішень

1. Інтелектуальні системи прийняття рішень (ІСПР) в фінансово-економічній сфері (мікро- та макроекономічні системи, банки, біржі, страхові компанії і т.д.).
2. Системи прийняття рішень в управлінні соціальними процесами.
3. Системи прийняття рішень в управлінні технологічними процесами в промисловості.
4. Інтелектуальний аналіз даних і знань; проблеми добування даних і знань (Data&Knowledge Mining), бази знань для ІСПР.
5. Математичне моделювання та прогнозування складних об'єктів і процесів.
6. Прийняття рішень в умовах невизначеності даних (системи нечіткого логічного виведення).
7. Сучасні методи й алгоритми ІСПР (генетичні та еволюційні алгоритми, нейронні мережі і т.д.).

***Fridman R.*[1] *Sagan V.Yu.*[1] *Slashcheva T.*[2] *Tymoshenko Yu.O.*[1]**
[1]*Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine;* [2]*Clinical hospital "Feofaniya", Kyiv, Ukraine*

Approach to automated ECG analysis

Introduction. Cardiovascular diseases(CVD) are the main cause of death globally. Approximately 17.5 million people died from CVD in 2005, representing at least 30% of all global deaths according to the world health organization (WHO) report. By 2016, almost 20 million people will die from CVD [1]. Some of the cardiovascular diseases causing death are caused by ventricular arrhythmia, valve disease, and coronary artery disease. Hence the early detection of these diseases in the initial stages is of great importance to the treatment and life prolonging of the patients.

Description of research. The electrocardiogram (ECG) signal is one of the most important tools in clinical practice to assess the cardiac status of patients. This signal represents the potential difference between two points on the body surface versus time. The main way of cardiac diseases diagnostic is ECG analysis. In ECG analysis held by doctors important role is given to detection of different features, that can indicate pathologies. Machine learning can be implemented in the process of ECG analysis to facilitate the work of doctors and enable them to use patterns that require huge amount of data to be analyzed. Raw data in this case is ECG itself, which is usually represented as time series with frequencies of 120-360 dots per second. Hence, a one minute of ECG recording may consist of 21600 dots and so ECG recordings can't be used as input data in machine learning. Situation becomes even more complicated when it comes to long-term ECG analysis.

Reseacrh 1. The first deep research, which was held by Italian scientists A. De Gaetano, S. Panunzi,F. Rinaldi, A. Risi, M. Sciandrone from the Institute for Systems Analysis and Computer Science, Rome. In the proposed method, the pattern under classification is defined to be the ECG signal from an RRR interval, i.e. the ECG signal from two successive heart beats, marked by three successive R peaks. It was remarked that RRR intervals can be easily detected and contain enough information for pattern recognition. The pattern under classification is the ECG digital recording of two consecutive RR-segments (hereafter defined RRR interval).Moreover R peaks can be easily detected in the digitized signal, since they are local maximum that can be exactly determined once known the period of the heart cycle and the sampling frequency. So The RRR interval contains two complete heart cycles and hence it constitutes suitably informative input of a pattern recognition system. Problem was faced by team of researches:

- Sampled ECG over an RRR interval is a vector lying on a high dimensional space whose dimension depends on cardiac frequency and sampling frequency.
- The presence of noise affects the measurements made in the recorded RRR intervals.

The proposed strategy is that of considering the ECG data from an RRR interval as a noisy sample of an underlying function. The function is approximated by means of a linear combination of a fixed number of suitable basis functions. The coefficients of the linear expansion can be easily computed by solving a linear least-squares problem, and constitute the extracted features of the RRR interval. The transformed patterns, i.e. the coefficients of the linear expansions, become the input signals of a feed-forward neural network classifier, which provides an output of zero for the normal and one for the ischemic case. In particular, the approximating function

y: $\mathbb{R} \to \mathbb{R}$ can defined as

$$y(t) = \sum_{j=1}^{m} a_j \phi_j(t);$$

where $\phi_j(t)$ are smooth functions from a suitable family, and the coefficients of the linear expansion $a_1, ...a_m$; have to be determined using the ECG values corresponding to the given RRR interval. In this report was offered approach for each patient, the ECG lead used and the statistics relative to the empirical distribution of ten thousand sensitivity (Se) and specificity (Sp) values from artificial test sets.It can be appreciated that sensitivity and specificity are both very high for all evaluated

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

patients,medians lying always above 96%. As a general measure of the reliability of the proposed method in the studied patient population, the grand averages of sensitivity and specificity are respectively 98.64% and 99.23%. [3]

Research 2. The second serious research on ischemia detection was held by a team from Greece including Costas Papaloukas, Dimitrios I. Fotiadis, Aristidis Likas and Lampros K. Michalis from University of Ioannina, Ioannina, Greece. They were using data sets from the European Society of Cardiology (ESC) ST-T database. The raw ECG signal containing the ST segment and the T wave of each beat were the inputs to the beat classification system and the output was the classification of the beat. [2] Principal component analysis(PCA) was used to provide dimensionality reduction of the input. The team has created ANN with performance of 90% sensitivity (Se) and 90% specificity (Sp) in beat classification. The ANN-based beat classifier was a part of four-stage ischemic episode detection procedure. The whole system was evaluated on the ESC ST-T database. The results of analyzer are sensitivity of 90% and positive predictive accuracy (PPA) of 89% when aggregate gross statistics was collected and sensitivity of 86% and positive predictive accuracy (PPA) of 87% when aggregate average statistics was collected.

Research 3. Third research on cardiac pathologies was held by a team from Italy including Carlo Marchesi and Rosaria Silipo. The part of this complex research was dedicated to ischemia detection. This research is very interesting, because it contains comparison between four different approaches to the problem of ischemia detection. The team has used ischemia-induced changes of ECG waveforms (developed through a project supported by the European Union and by the European Society of Cardiology, lead by the CNR Institute of Clinical Physiology, Pisa, Italy). [4] Compared approaches to ischemia detection are:

- Traditional approach, which uses empirical thresholds (sensitivity of 77% and positive predictive accuracy (PPA) of 82%).
- Improved version of approach, which uses static ANN trained with BP with using ST elevations, ST depressions,normal patterns and artifacts at preprocessing (sensitivity of 77% and positive predictive accuracy (PPA) of 83%).
- Approach, that uses principal component analysis (PCA) to dimension-reduction and a two-layer feed forward static ANN, that was trained using the BP (sensitivity of 77% and positive predictive accuracy (PPA) of 86%).
- Approach, that uses principal component analysis (PCA) to dimension-reduction and a two-layer feed forward static ANN, that was trained using the BP (sensitivity of 77% and positive predictive accuracy (PPA) of 86%).

Conclusions. All of reviewed approaches were been used to detect and classify heart diseases using ANN-based beat classifier. The results, obtained by all teams are better then other reported. Such results show promise that further implementation of such diagnostic systems in high-end cardiology may be possible.

References. **1.** World Health organization," Cardiovascular diseases", Fact sheet N°317, February 2007. http://www.who.int/cardiovascular_diseases.

2. C. Papaloukas, D.I. Fotiadis, A. Likas, L.K. Michalis, An ischemia detection method based on artificial neural networks, Artificial Intelligence in Medicine 24 (2002) 167–178.

3. A. De Gaetano , S. Panunzi , F. Rinaldi , A. Risi , M. Sciandrone , A patient adaptable ECG beat classifier based on neural networks, Applied Mathematics and Computation 213 (2009) 243–249

4. Rosaria Silipo and Carlo Marchesi, Artificial Neural Networks for Automatic ECG Analysis , IEEE transactions on signal processing, Vol. 46, No. 5, MAY 1998 1420-1422

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Ivchenko D.A., Bondarenko A.S., Mazur V.G.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Big Data Analysis in Health Care: Perspectives and Potential

Healthcare industry historically has produced vast amounts of data from moving accounting and patient care. While most of the data is stored in hard copy, the modern trend is in rapid digitization of large amounts of data. Following the mandatory claims and the potential to increase the quality of medical care in order to reduce costs, these massive amounts of data promise to support a wide range of medical and health services, including but not limited to support clinical decision-making, surveillance, population management and health care [1].

The Potential of large data analysis in health care will lead to improved outcomes exists in many scenarios, such as: analyzing patient characteristics and treatment outcomes and costs to recognize the most clinically and cost-effective treatments and offer analysis and tools, thereby influencing provider behavior.

Benefits for Health. By digitizing, merging and efficient using of large data health organization, from a physician offices and groups interested suppliers of large hospital networks and organizations accountable care stand to realize significant benefits. Potential benefits include detecting the disease at an early stage when they can be handled more easily and effectively.

Patient Analyst Profile: apply advanced analytics profiles for patients to recognize individuals who would suffered from disease (e. g. diabetes, blood pressure, asthma). Remote device monitoring and analyzing in instant amounts of fast moving data are in hospital and home security monitoring devices for unfavorable prognosis. Meanwhile, in many medical situations, continuous data in real time can bring into effect of difference between life and death.

The data structure that we have to work with. Structured data are the data that can be easily recalled, saved, analyzed and manipulated on any device. In health care, structured and semi-structured data includes readings and data derived from the ongoing conversion of paper documents into electronic health and medical accounting. Historically, the point of departure is generated unstructured data, medical records office, handwritten notes nurse and physician, hospital admission and discharge records, X-ray films. Structured data have to include familiar input field records, as patient name, date of birth, address, the name and address of the physician, treatment reimbursement codes.

Sources and types of data include:.

1. Biometric data: handwriting, blood pressure, X-rays, heart rate, fingerprints, retina scans.
2. Web data and social media: with interaction to Facebook, Vk, Twitter, LinkedIn, blogs and more. It may also include health plan websites or smartphone applications.
3. Machine to machine data: taking from distant sensors or other significant sign devices [2].

Method that are used for big data analyses. For large data analysis, these data should be combined. The second component is data in "raw" state and must be processed or transformed. Service-oriented architecture access combined with web services is one possibility.

Another approach is a data warehouse where data from different sources are combined and prepared for handling, although the data are not accessible in real time. Via steps extraction, transformation and loading, data from different sources become prepared and cleared. Depending on unstructured or structured data, more data types can be represented on a large data analysis platform.

The typical large data analysis in health include inquiries, records, OLAP and data mining. It should aggregate big amount of unstructured data, so we can use "NoSQL" database, like MongoDB or Postgresql [3].

Nowadays, Cloud Computing plays a significant role in increasing the data storage and processing capacity required to handle such an amount of data, ensuring its accessibility anytime and anywhere. The European Commission cloud computing strategy aims to facilitate faster adoption of secure cloud solutions in Europe, which should support secure storage of health data over the Internet.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Already existed solutions of mHealth. The European Parliament on the eHealth Action Plan 2012-2020 underlines the potential of mobile health and wellbeing apps for patients and the need to have a clear legal framework to ensure their development and safe adoption. SUSTAINS – Support Users To Access Information and Services – is a three-year project aimed at developing and deploying a basket of services in 11 European regions providing patients' access to Electronic Health Records (EHR). Medical results: Progress in treatment especially of chronic diseases need efficient and continuous contact between the patient and the professionals in order to achieve optimal medical result. Efficiency and economy: with the new treatments available, and the growing demand from patients/public, there is a need for improved efficiency and economy [4].

The mHealth project that already implemented. One more example is a new app for diabetes called One Drop, which just launched in the App Store few weeks ago in the USA. Users of the app can enter their personal insulin levels using a glucose meter, log the effectiveness of their medication, what they recently ate and what types of activities might affect their blood sugar. The food component will let users snap a picture, rather than searching for and adding in the food and nutrition information. It also comes with a journal and diabetic meal planner embedded in the app. App will help patients log glucose data using the MeterSync Blue glucose meter device. The app starts to keep a running log of how the body reacts to these different data points, which is an important feature in helping both a diabetic person and their doctor determine a more effective treatment plan [5].

Challenges. Big data analysis platform in healthcare must support key functions required for execution. Evaluation criteria may include platform accessibility, usability, ability to manage different levels of detail, privacy and security solutions, as well as quality control [6]. In addition, while most platforms available are now open source, typical advantages and disadvantages platform open source is used. To succeed, big data analysis in health care must packed so is user-friendly and transparent. Instant analysis of large data is a key claimed in the health sector. The gap between data collection and executing should be addressed. Dynamic presence of numerous analytical algorithms, models and methods in type pull-down menu is also required for large-scale adoption.

Conclusions. Big data analysis has the potential to change the way doctors use sophisticated technologies to store clinical data and make decisions. In the near future we will see a rapid, widespread adoption of large data analysis for medical industry and healthcare.

References. **1.** Dembosky A: "Data Prescription for Better Healthcare." Financial Times, December 12, 2012, p. 19. 2012.http://goo.gl/SYCQND **2.** IHTT Transforming HealthCare through Big Data Strategies for leveraging big data in the health care industry. 2013.http://goo.gl/6NQ7fE **3.** Borkar V., Carey MJ, Chen L: Big data platforms: what's next? ACM Crossroads 2012, 19(1):44-49. **4.** Support Users To Access Information and Services - http://www.sustainsproject.eu/ **5.** Razorfish's Jeff Dachis Launches One Drop, An App For Diabetics To Log And Share Info With Each Other, 2015 -http://goo.gl/CAv5SI **6.** Feldman B, Martin EM, Skotnes T: "Big Data in Healthcare Hype and Hope." October 2012. Dr. Bonnie 360. 2012. http://goo.gl/OuVrOZ

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Korshunova I.A., Kopychko S.M.
National Technical University of Ukraine "Kyiv Polytechnic Institute", Kyiv, Ukraine

Detecting epileptic seizures from raw EEG data using convolutional neural networks

Epilepsy is currently one of the most commonly diagnosed neurological disorders, which is characterized by the occurrence of spontaneous seizures. About 1 out of 3 patients has a drug-resistant epilepsy, which means that seizures cannot be controlled by medication. For their caregivers alarm systems have been built, which are often known for their low accuracy [1]. Therefore, a reliable seizure detection system can be an important tool for patients and caregivers. Existing seizure detection algorithms most commonly utilize hand-crafted features based on electroencephalogram (EEG) [2]. Our goal was to build a patient specific classifier based on the raw EEG signal, which were processed with convolutional neural networks (convnets) [3].

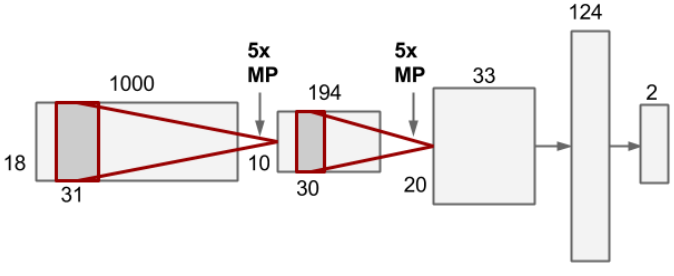

Figure 1. A schematic of a convnet architecture

The proposed architecture is shown on fig. 1. The input represents 10 second long sections of normalized EEG sampled at 100 Hz from 18 channels. The network has two convolutional layers each followed by a max-pooling (MP) layer [3]. The upper layer is a softmax preceded by a hidden layer, where we used dropout as one of the ways to prevent overfitting [4]. The hyper-parameters of the model (number of feature maps per layer and their shapes, max pooling shape, number of units in the hidden layer, learning rate decay) were chosen among reasonable settings on a validation dataset of several patients from CHB-MIT Scalp EEG database [5]. Ideally, if we are building a patient specific system, parameters should be tuned for every patient separately, however, as it involves a lot of computation time, we exploited one set of hyperparameters.

The performance of the model was determined with leave-one-hour-out cross-validation. For each patient from the CHB-MIT Scalp EEG database, we calculated the number of false positives and false negatives with respect to the number of seizures. The results are comparable to the current state-of-the-art solutions [1]. However, the detection delay, which is crucial for seizure detection systems, has to be determined to fully evaluate the proposed method. This work is, therefore, not finished yet and can be improved upon in various ways. For example, to take into account the dependence between EEG sections, one can use a recurrent neural network with features from convnet as its input. Moreover, this work is just the first step towards general seizure detector that is (possibly automatically) adaptable for each patient.

References. **1.** P. Buteneers, "Detection of epileptic seizures: the reservoir computing approach", Ph. D. dissertation, ELIS, Ghent University, Ghent, Belgium 2012 **2.** A. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment", PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, 2009 **3.** Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-Based Learning Applied to Document Recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov, 1998 **4.** N. Srivastava, "Improving neural networks with dropout", Masters thesis, U. Toronto, 2013 **5.** A. Goldberger, L. Amaral, L. Glass, J. Hausdor, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, H. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals", *Circulation*, vol. 101, no. 23. pp.215-220, 2000

***Malinovskiy A.D.*[1] *Nikolaiev S.S.*[2]**
[1]*National Technical University of Ukraine "Kyiv Polytechnic Institute", Faculty of Informatics and Computer Science, Kyiv, Ukraine;* [2]*Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine*

Classification of the ECG features for automatic heart diseases detection

Abstract. This paper describes an overview of different ECG features classifications aimed at automatic ECG annotation techniques and heart pathology detection with machine learning algorithms. Each class represents special features set that is used in different machine learning methods to classify pathologies and predict heart malfunctions. Partitioning features into classes is carried out by mathematical methods and used data types. Classification provided in this paper allows to select the feature class with required properties and quality metrics for specific studies.

Relevance. In modern cardiology automatic detection and prediction of heart pathologies is based on methods that analyze ECG (Electrocardiogram). These methods require extraction of specific ECG features for machine learning algorithms to make correct decisions on the patients health. All of these algorithms have different quality of ECG abnormalities detection and usage of the right set of features can greatly increase quality metrics. That is why it is very important to know what classes of features exist and how informative a given feature set is.

Description of research. The purpose of this work is to provide classification of the features that are used for automatic pathologies prediction and detection. The classification of feature types is the following:

- Timing interval-based features
- Morphological-based features
- Wavelet-based features
- Frequency-domain features (Using Fast Fourier Transform)
- Shape-based features(shape matching)

Timing interval-based features. This class works only with five peaks of ECG signal complexes: P, Q, R, S, T. It measures values like height of P wave and length of different parts of the heartbeat (PR segment, PR interval, QRS complex, ST segment, T wave, ST interval, RR interval). The obtained values are compared to thresholds and depending of the comparison results conclusion is made whether it is normal or abnormal complex (heartbeat). [1]

Morphological-based features. This class contains features that depend on the form of ECG signal. They can be divided into eight subclasses (groups). All of them contain amplitude values of the ECG signal. The major distinction between the groups is the method used for sampling the signals. First four groups (segmented ECG morphology features contained 19 features each) use a sampling rate determined by the heartbeat segmentation information (sampling windows determined by the QRS onset and offset and the T-wave offset points). Last four groups (fixed interval ECG morphology features) use a fixed-interval sampling rate and the heartbeat fiducial point (FP) located in the sampling windows. [1]

Wavelet-based features. The wavelet transformation is based on a set of analyzing wavelets allowing the decomposition of ECG signal in a set of coefficients. Each analyzing wavelet has its own time duration, time location and frequency band. The wavelet coefficient resulting from the wavelet transformation corresponds to a measurement of the ECG components in this time segment and frequency band. The wavelet transform is a convolution of the wavelet function ψ (t) with the signal x(t). Orthonormal dyadic discrete wavelets are associated with scaling functions φ (t). The scaling function could be convolved with the signal to produce approximation coefficients S. One of the crucial steps in the ECG analysis is accurately detect the different waves forming the entire cardiac cycle. [2]

Frequency-domain features. The main idea of frequency-domain features is to analyses spectrum of ECG signal instead of ECG signal itself. Frequency spectra are computed by the function Fast Fourier Transform. The DFT (Discrete Fourier transform) found by taking the n-point FFT. The

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

FFT returns a two-sided spectrum in complex form (Real and Imaginary Parts), which can scale and convert to polar form to obtain amplitude and phase. The spectrum varies with different cardiac arrhythmia's, and power spectra are observed in the frequency range from 1 to 20 Hz. [3]

Shape-based features (shape matching). When one heartbeat interval from the original ECG signal is selected, there is an initial translation bias depending on the choice of the interval starting point. The signal as a result may need to be circularly shifted to perform an initial registration. The translation required is usually much larger than the one is allowed during DTW (Dynamic time warping) alignment, so it is extracted separately. Since the signals are periodic, a circular shift is performed. Once the pair of signals are initially registered, the DTW alignment is performed. The alignment transform then is used to project one signal into the other and the residual error is evaluated. This pair-wise matching of a single heartbeat interval is repeated over multiple heartbeat segments over the available data and the average residual error is used to rank the matches. Finally, channel information from each of the leads combined through a winner take-all mechanism to retain the best matching ECGs as those that match in the highest number of channels with a score exceeding a chosen threshold. [4]

Conclusions. In this paper, an extensive classification of features for the use in machine learning algorithms for automated heart pathologies detection and prediction was offered. Main classes of features have been identified. The separation into classes was carried out by mathematical methods as well as the nature of the objects (segments spectra shape, etc.). Given classification allows selecting the required feature class with desired attributes and quality metrics for specific studies.

References.

1. Philip de Chazal, Maria O'Dwyer, and Richard B. Reilly : Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features // Transactions on biometrical engineering, Vol. 51, NO. 7, JULY 2004

2. Pooja Sabherwal: Wavelet Transform As Method for ECG Signal Analysis // International Journal of Emerging Science and Engineering (IJESE) ISSN: 2319–6378, Volume-2, Issue-1, November 2013

3. Chia-Hung Lin: Frequency-domain features for ECG beat discrimination using grey relational analysis-based classifier // Department of Electrical Engineering, Kao-Yuan University, Lu-Chu Hsiang, Kaohsiung 821, Taiwan Received 1 August 2006; received in revised form 4 April 2007; accepted 10 April 2007

4. Tanveer Syeda-Mahmood, David Beymer, and Fei Wang: Shape-based Matching of ECG Recordings // Proceedings of the 29th Annual International Conference of the IEEE EMBS Cité Internationale, Lyon, France August 23-26, 2007.

Sokolov P.Ye., Bidyuk P.I.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Model selection and combining forecast estimates

A methodology is proposed for model selection while analyzing processes in economy and finances. A combining technique is considered to reach necessary quality of forecasts.

All modern automatic control systems, the systems for business management, weather forecast systems, personal, technical and business oriented decision support systems (DSS) and many others directions of human activities require quality forecasts. One of functional elements of modern DSS directed towards solution of this problem is combining of forecast estimates generated by alternative techniques. According to [1] the simplest definition of DSS is as follows: DSS is a computer based system that provides any kind of help for a decision making user: in data collecting and its preliminary processing, model constructing and forecast estimation, computational process visualization and handy presentation of the final result.Existing results on the properties and performance of forecast combinations have been derived in the context of mean squared error loss. Under this loss function empirical studies have generally found that equally-weighted combined forecasts lead to better performance than estimates of optimal forecast combination weights which in turn outperform the best individual predictions The idea of combining forecasts is illustrated in Fig. 1. The forecasting methods used in the system proposed are as follows: regression analysis (linear and nonlinear models), the group method for data handling (GMDH – Kolmogorov-Gabor polynomials), fuzzy GMDH, fuzzy logic, appropriate versions of Kalman filter, neural nets, support vector regression, nearest neighbor and probabilistic type techniques (Bayesian networks, Bayesian regression). The set of techniques used covers linear and nonlinear nonstationary process. The nearest neighbor technique can be hired for generating long term forecasts in a case of availability long data samples (better with repeating patterns). Practically all the techniques are implemented in adaptive versions what makes the system more flexible and ready for newly coming data, and to fight some types of uncertainties. On the stage of model structure estimation an appropriate principal component analysis technique is applied when necessary. The information of a model structure provides a possibility for selection of parameters estimation techniques for candidate models among which are: ordinary LS and nonlinear LS (NLS), recursive LS (RLS), maximum likelihood (ML), and some suitable for parameter estimation versions of Markov Chain Monte Carlo (MCMC) techniques. Some special optimization techniques are applied in a case of estimating fuzzy GMDH structures. The candidate models estimated are tested with a set of statistical quality criteria some of which are as follows: combined criterion, determination coefficient, Durbin-Watson statistic, F-statistic, t-statistic etc. One or more acceptable models are used for computing forecasts that are also tested for quality with another set of statistical criteria. Usually three or more forecasting techniques are used for the same dataset to get a possibility for combining the forecasts so that to further improve the final combined estimate. It is known that combination of forecasts results in improvement of combined forecast in the case when variances of errors for separate forecasts do not differ substantially [2]. In the system proposed the forecasts generated by models can be computed (at least) with six selected techniques as shown in Fig. 1. Regression model (autoregression (AR) or AR with moving average (ARMA)) is used for generating forecasts as well as it is transformed into state space (SS) form for further use by the Kalman filter (KF). Adaptive version of KF provides a possibility for forecasting and on-line/off-line estimation of state disturbances and measurement noise covariances. The group method of data handling method (GMDH) is well-known technique that generates models in the general form of Kolmogorov-Gabor polynomial, and the last three methods are related to the intellectual data analysis techniques. The best results of combining the forecasts are achieved when variances of forecasting errors for different forecasting techniques do not differ substantially.

To reach high quality of forecasting model initial statistical (experimental) data is processed to perform normalization, imputation of missing observations, processing of outliers, digital filtering

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

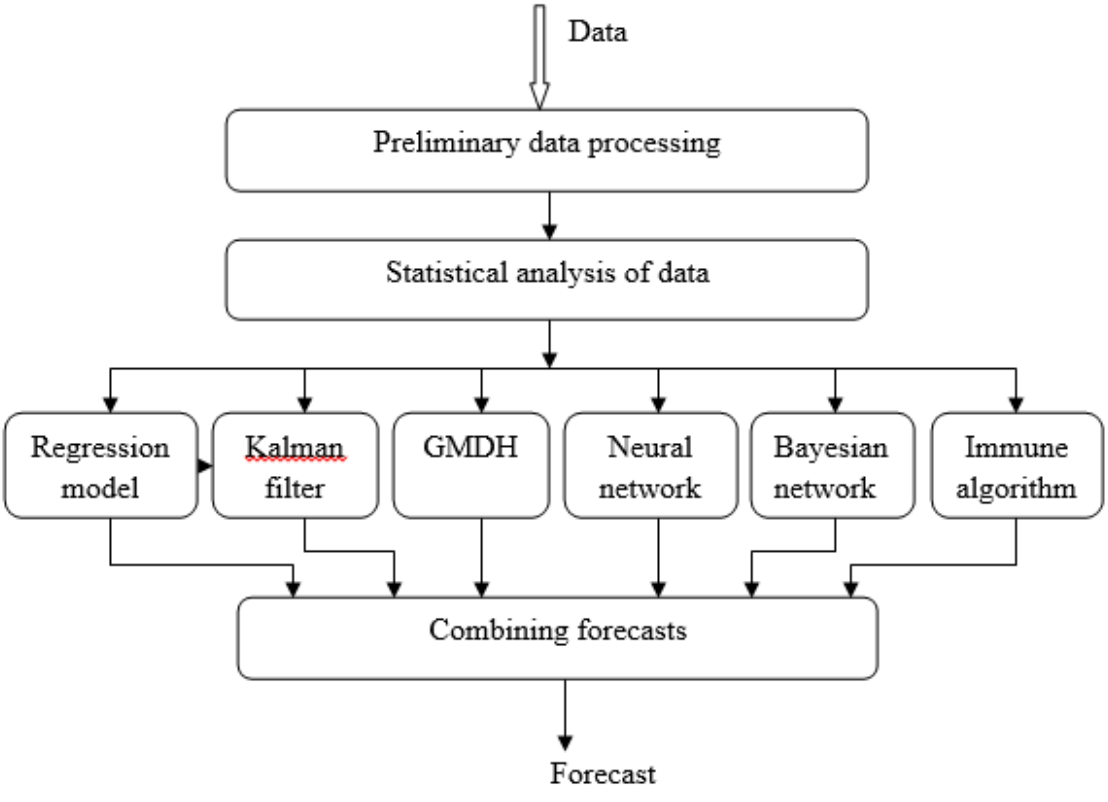

Figure 1. Illustration of principle for combining forecasts

and smoothing (when necessary). In the process of model structure and parameters estimation possible uncertainties are taken into consideration, among which are the following: uncertainty of model order; time delay; type of nonlinearity; type of stochastic external disturbances which is represented by respective type of distribution and its parameters; inconsistency of model parameters estimates; nonstationary errors of forecasting. The uncertainties mentioned are estimated and taken into consideration in the model construction process thanks to hiring alternative model order estimation procedures, optimal Kalman filter (estimation of covariances for stochastic external disturbances and measurement errors), special algorithms for estimation of delay time [3]. A substantial reduction of influence of probabilistic type uncertainties is reached with application of modern Bayesian techniques such as static and dynamic Bayesian networks (BN) the purpose of which is in generating alternative probabilistic forecasts.

References. **1.** Бідюк П.І., Романенко В.Д., Тимощук О.Л. Аналіз часових рядів. – Київ: НТУУ «КПІ», 2013. – 600 с. **2.** Tsay R.S. Analysis of financial time series. – New York: John Wiley and Sons, Inc., 2010. –715p. **3.** Бидюк П.И., Меняйленко А.С., Половцев О.В. Методы прогнозирования. — Луганск: Альма Матер, 2008. – 608 с. **4.** Gibbs B.P. Advanced Kalman filtering, least squares and modeling. – Hoboken: John Wiley and Sons, Inc., 2011. – 627 p.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Tkachenko D.A.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Automated university course timetabling using genetic algorithm

University course timetabling problem (UCTP) is a combinatorial problem, in which a set of lessons are scheduled into time slots in a way that the maximum number of constraints are satisfied. Large search space in this problem leads to many man-hours spent by employees, involved in solving it manually. Besides, NP-completeness [1] of this task limits the selection of algorithms suitable for its automated solution. For large universities complexity of UCTP impedes not only the usage of global optimization methods, but also the application of heuristic methods (for example, approaches based on graph coloring). This paper describes the metaheuristic approach to solving this problem that employs the genetic algorithm. [2, 3]

Data representation. The algorithm's input is the object of class *Schedule* that contains the list of objects of class *Lesson*. Each *Lesson* contains lists of objects of *Professor*, *Auditorium* and *Group* classes as well as objects of *Course* and *CourseType* classes, and integer *TimeSlot* (describes on which day of week a lesson takes place and when does it start). This is presented in the fig. 1. Input objects of class *Lesson* can have assigned *TimeSlot* (the time of some of the lessons may be fixed and does not need to be optimized).

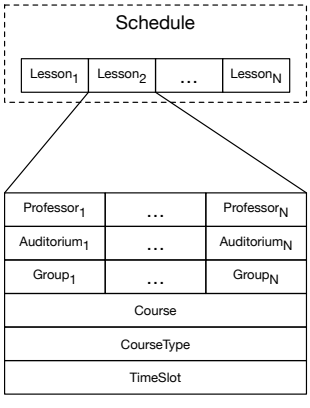

Figure 1. Data representation

Optimization objective. Assign each lesson from the set of lessons $\{l_1, l_2, ..., l_m\}$ to time slot (if it is not fixed) from the set of time slots $\{ts_1, ts_2, \ldots, ts_n\}$, so that the value of fitness function $f(S)$ (S stands for *Schedule*) is as close to 0 as possible. The number of ways to assign m lessons to n time slots can be calculated as n^m.

Constraints. There are two types of constraints: the *hard* and the *soft* ones. Hard constraints describe the situations that must never occur. Specifically, at any given moment of time:

- Professor must not be assigned to more than one lesson.
- Only a single lesson must be conducted in a specific auditorium.
- Group must not attend more than one lesson.

Satisfaction of the soft constraints makes attending lessons more convenient. Those include:

- Reducing the number of gaps between lessons for groups and professors.
- Decreasing the number of working days for professors.
- Assigning lessons to the time slots preferred by professors.

Fitness function. Value of the fitness function is defined by the number of violated constraints and is calculated as:

$$f(S) = C_H N_{HCV} + C_S S_{SCV},$$

where N_{HCV} — number of violated hard constraints, S_{SCV} — score of violated soft constraints, C_H and C_S — "weights" of hard and soft constraints respectively, chosen in a way, so that for any expected value of N_{HCV} and S_{SCV} the following condition holds: $\mathcal{O}(C_H N_{HCV}) > \mathcal{O}(C_S S_{SCV})$ (i.e., the order of magnitude of $C_H N_{HCV}$ exceed the order of magnitude of $C_S S_{SCV}$).

The value N_{HCV} is computed by counting the number of duplicates in the list of time slots associated with every professor, group and auditorium: each repetition increases the value of N_{HCV} by 1.

S_{SCV} is calculated using the following expression:

$$S_{SCV} = \sum_i W^i_{PC}(C_{EWD} N^i_{EWD} + C_{DTS} N^i_{DTS}) + C_{PG} N_{PG} + C_{GG} N_{GG},$$

where W^i_{PC} — "weight" of preferences of i-th professor; N^i_{EWD} — number of excessive working days of i-th professor; N^i_{DTS} — number of lessons taught by i-th professor assigned to disfavored time

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

slots; N_{PG} — number of gaps between lessons for professors; N_{GG} — number of gaps between lessons for groups; C_{PG}, C_{GG}, C_{EWD}, C_{DTS} — coefficients, that reflect undesirability of corresponding events.

Experiment. Software implementation of the described approach was used to generate the course timetable for SD department of ESC "IASA".

Input data. Attributes of the input data are presented in the table 1. Algorithm was set to minimize the number of working days for 14 professors. Lists of preferred time slots were defined for 4 professors: 2 of those lists consisted of 10 slots, another 2 had 5 slots. Algorithm was also set to minimize the number of gaps between lessons for all groups and professors.

Table 1. Input schedule attributes

Lessons:	
total	124
with fixed time slot	52
Time slots	25
Groups	12
Auditoriums	25
Professors	39

Algorithm parameters and coefficients. Genetic algorithm parameters used are shown in the table 2. Termination condition: no improvement in fitness is observed within 300 generations. The following coefficients were used for fitness function calculations: $C_H = 1000000$, $C_{EWD} = 100$, $C_{GG} = C_{DTS} = 10$, $C_S = C_{PG} = 1$, $W^i_{PC} = 1\ \forall i$.

Table 2. Genetic algorithm parameters

Population size	250
Crossover points	2
Mutation probability	2%
Elitism	5%

Results. Optimization procedure was repeated 10 times. Fitness of the best result: $f(S) = 388$. In the optimized schedule: 10 groups out of 12 and 30 professors out of 39 did not have gaps between lessons; 11 professors out of 14 had lessons spread across the minimal number of working days; 8 out of 11 lessons were assigned to the time slots preferred by professors. Plots of the fitness of the best individual $f(S)$ against a generation number are displayed in the fig. 2 and 3. Plots show that the fitness $f(S)$ decreased or remained the same in each subsequent generation. This indicates that each successive generation either produced the better schedule or the fittest individual was carried over to the next generation unaltered (when close to the local optimum).

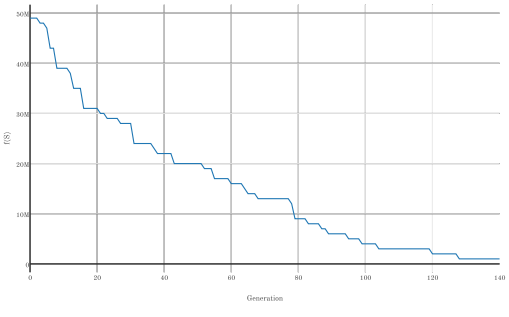

Figure 2. Hard constraints satisfaction

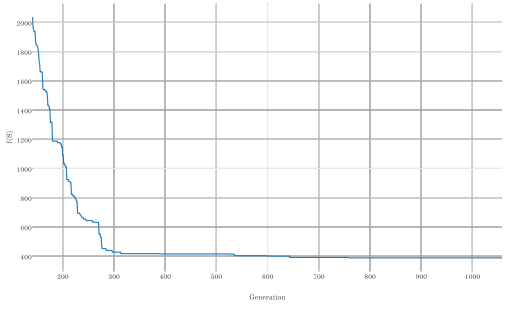

Figure 3. Soft constraints satisfaction

Conclusion. Genetic algorithm was used to solve university course timetabling problem. In particular, constraints that conform to the typical needs of the universities were formalized. Appropriate fitness function was created. Experiment proved the efficiency of the described approach: algorithm converged to near-optimal solution. Only about 1000 generations were produced for optimization, allowing the algorithm to perform well on any conventional PC.

References. **1.** M.M. Garey, D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York, NY: W. H. Freeman & Co., 1979. **2.** M. Mitchell. *An Introduction to Genetic Algorithms*. Cambridge, MA: A Bradford Book, 1998. **3.** A.E. Eiben and J.E. Smith. *Introduction to Evolutionary Computing*. Berlin and Heidelberg, Germany: Springer-Verlag, 2008.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

***Zaychenko Yu.P.*[1], *Ovi Nafas Aghaei Agh Ghamish*[2]**
[1]*Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine;* [2]*National Technical University of Ukraine "Kyiv Polytechnic Institute", Kyiv, Ukraine*

Banks bankruptcy risk forecasting with application of fuzzy methods

Introduction In a modern economy problem of banks financial state analysis and possible bankruptcy forecasting plays the significant role. The timely detection of features of coming bankruptcy enables top bank managers to adopt urgent measures for stabilizing financial bank state and prevent the bankruptcy. Up to the date there are a lot of methods and techniques of banks state analysis and determination of bank rating - WEBMoney , CAMEL, Moody's S&P etc. [1,2]. But their common drawback is that all of them work with complete and reliable data. and cannot give correct results in case of incomplete and unreliable input data. This is especially actual for Ukrainian bank system where bank managers often provide the incorrect information about their financial indices that to obtain new credits and loans.

Due to aforesaid it's very important to develop new approach and technique for banks bankruptcy risk forecasting under uncertainty. The main goal of this paper is to develop and investigate novel methods of bank financial state analysis and bankruptcy forecasting under uncertainty and compare with classical conventional methods and approaches.

For our research the quarterly accountancy bank reports were used obtained from National bank of Ukraine site [1]. For analysis the financial indices of 170 Ukrainian banks were taken to the date 01.01. 2008 and 01.07.2009, that is about two years before crises and just before the start of crises.

I. The application of fuzzy neural networks for financial state forecasting. For forecasting of banks bankruptcy risk it was suggested the application of fuzzy neural networks (FNN) ANFIS and TSK [3]. The application of FNN was determined by the following advantages :

1. the possibility to work with incomplete and unreliable information under uncertainty;
2. the possibility to use expert information in the form of fuzzy inference rules.

In the first series of experiments input data at the period of January 2008 were used (that is for two years before possible bankruptcy) and possible banks bankruptcy was analyzed at the beginning of 2010 year.

Experiment №1. Training sample- 120 Ukrainian banks,
Test sample- 50 banks; Number of rules = 5,10
Input data – financial indices (taken from bank accountant reports)

- Assets; Capital; Cash (liquid assets); Households deposits; Liabilities.

The comparative analysis of forecasting results versus the number of rules is presented in the table 1.

Table 1. Comparative analysis of FNN ANFIS and TSK in dependence on rules number

Network / number of rules	Total amount of errors	% of errors	Amount of 1 type errors	Amount of 2 type errors
Anfis 5	6	12%	0	6
Anfis 10	7	14%	1	6
TSK 5	5	10%	0	5
TSK 10	6	12%	1	5

Experiment №2. Training sample - 120 Ukrainian banks; Test sample - 70 banks; Number of rules = 10. Input data – financial indices the same as in experiment №1.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

After analysis of the experimental results the following conclusions were made:

1. FNN TSK gives more accurate forecasting than FNN ANFIS.
2. The variation of the number of rules in the training and test samples doesn't influence on the accuracy of forecasting.
3. The next series of experiments was aimed at determining optimal input data (financial indices) for forecasting. The period of input data was January 2008. While experiments the adequate input financial indices were detected with which the best forecasting results for Ukrainian banks.

II. The application of fuzzy gmdh for financial state forecasting. In the process of investigations Fuzzy Group Method of Data Handling (FGMDH) [3,4] was also applied for financial state of Ukrainian banks forecasting.

In the table 2 the corresponding results are presented in dependence on input data period.

III. The generalised analysis of crisp and fuzzy forecasting methods. In the concluding experiments the comparative analysis of application of all the considered fuzzy and crisp methods were carried out. The following methods were considered:

- fuzzy neural network ANFIS; fuzzy neural network TSK; crisp Kromonov's method; Byelorussian bank association method (crisp).

As input data were entered the financial indices of Ukrainian banks on July 2007 year. The results of application of all methods for bankruptcy risk analysis are presented in the Table 3.

Table 2. Comparative results of forecasting using method FGMDH in dependence on period of input data collection

Method/ period	Total amount of errors	%% of errors	1 type errors	2 type errors
ANFIS	7	10%	1	6
TSK	5	7%	0	5
Kromonov's method	10	15%	5	5
BBA method	10	15%	2	8

IV. Conclusions.

1. After experiments the most adequate set of financial indices was detected for banks bankruptcy risk forecasing.
2. It was established that FNN TSK gives much more accurate results than FNN ANFIS. The variation of rules number in training sample doesn't influence on forecasting results. While comparing the fuzzy forecasting methods it was established FNN give more accurate forecasting with more fresh data, that is are preferable for short-term forecasting. The fuzzy GMDH gives better results using older data that is, more preferable for long-term forecasting (two or more years).
3. In general, the comparative analysis had shown that fuzzy forecasting methods and techniques give beter results than conventional crisp methods while forecasting bank financial state and bankruptcy risk.

References. **1.** Rating estimates for commercial banks http://www.nbuv.gov.ua/Soc_Gum/Ekpr/2009_25/mescheryakov/htm. **2.** Building bank rating using the method of calculation multilevel integrated index of bank financial state. http://www.credit-rating.ua/ru/analytics/analytical-articles/. **3.** Zgurovsky M.Z., Zaychenko Yu.P. Decision-making Models and Methods under fuzzy conditions. – Kiev: Publ. House «Naukova Dumka»,. 2011.-275 p.(rus). **4.** Zaychenko Yu.P. Fuzzy models and Methods in Intelligent Systems. Kiev.: Publishing House«Slovo»,2008.-344.

Zyma D.S., Bidyuk P.I.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Information system for forecasting nonlinear nonstationary processes

Computer based information system is proposed for modeling nonlinear nonstationary processes in economy and finances.

The systemic approach to forecasting is based on the following system analysis principles: – functional completeness that is providing all functions necessary for preliminary data processing, model structure and parameters estimation, forecasting functions constructing, and selection of best alternative using a set of statistical criteria. A concept of the information system (IS) is based on modular approach to system architecture design that provides for a possibility of extending and/or modifying its functional possibilities. Separate functional blocks of the systems are combined into functional subsystems using the principle of functional similarity. The IS layout is given in Fig. 1.

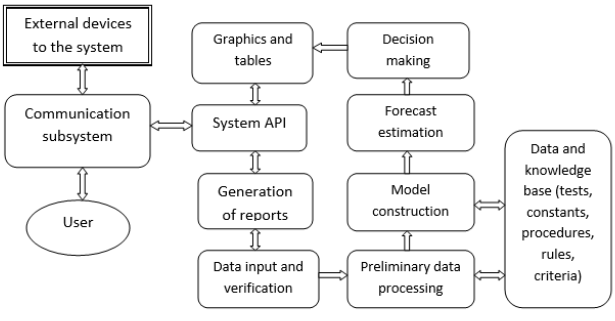

Figure 1. General layout of information system for forecasting

As the most of modern information systems directed towards data processing the IS proposed includes two basic components: the system and instrumental components. The system component has the following basic functions: – collection, modification, and renewal of experimental (statistical) data and the data used for system adjustment; – communication with other (external) systems, say, decision support systems (DSS) for managers, local and other nets; – means for the system re-configuration; – operational systems etc. The instrumental component contains a set of means for solving all tasks relevant to modeling and forecasting of nonlinear nonstationary processes (NNP) selected for investigation. To reach high quality of forecasting model initial statistical (experimental) data is processed to perform normalization, imputation of missing observations, processing of outliers, digital filtering and smoothing. The uncertainties mentioned are estimated and taken into consideration in the model construction process thanks to hiring alternative model order estimation procedures, optimal Kalman filter (estimation of covariances for stochastic external disturbances and measurement errors), special delay time estimation algorithms when necessary. A substantial reduction of influence of probabilistic type uncertainties is reached with application of modern Bayesian techniques such as static and dynamic Bayesian networks (BN) the purpose of which is in generating alternative probabilistic forecasts [3]. There is also available a wide set of alternative techniques for performing inference over the model.

References. **1.** Бідюк П.І., Романенко В.Д., Тимощук О.Л. Аналіз часових рядів. – Київ: НТУУ «КПІ», 2013. – 600 с. **2.** Tsay R.S. Analysis of financial time series. – New York: John Wiley and Sons, Inc., 2010. –715p. **3.** Бидюк П.И., Меняйленко А.С., Половцев О.В. Методы прогнозирования. — Луганск: Альма Матер, 2008. – 608 с.

Акимов В.С., Сергеєв-Горчинський О.О.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Визначення оптимального об'єму даних для побудови моделі класифікації

Постановка задачі. При побудові моделі класифікації зазвичай виникають проблеми пов'язанні з обмеженим об'ємом навчальних даних. В даній роботі запропоновано алгоритм оцінки оптимального об'єму даних для побудови моделі класифікації з наперед заданою точністю класифікації. У порівнянні з існуючими методами запропонований алгоритм дозволяє проводити оцінку оптимального об'єму даних лише по одному атрибуту, що значно зменшує час пошуку та прогнозування оптимального рішення.

Існуючі алгоритми.

1. Метод довірчих інтервалів [1].
2. Метод оцінки об'єму вибірки, запропонований Демиденко [2].
3. Метод ковзаючого контролю [3].
4. Метод порівняння емпіричних розподілів ймовірностей випадкових величин (ВВ) за допомогою відстані Кульбака-Лейблера [4].

Опис запропонованого алгоритму. Нехай маємо навчальну вибірку об'ємом n_0. Припустимо, що після побудови моделі класифікації було визначено, що точність класифікації моделі $M[0]$ є незадовільною за кількістю коректних класифікацій. Для покращення результату класифікації запропоновано алгоритм розрахунку оптимального об'єму вибірки по критерію відносної наближеності усіх значень атрибута з найбільшою ентропією до нормального розподілу ймовірностей [5]. Вхідними параметрами алгоритму є мінімальне необхідне значення приросту точності класифікації ε та кількість генерованих даних Δn на кожній ітерації пошуку. Метою алгоритму є визначення такого об'єму вибірки N_{onm}, починаючи з якого буде виконуватися нерівність $M[k] - M[0] \geqslant \varepsilon$ (1), де ε - приріст точності класифікації, $M[k]$ - точність класифікації поточної моделі, $M[0]$ - точність класифікації початкової моделі. Першим кроком є вибір атрибуту A, що має найбільше значення ентропії (атрибут з найбільшої множиною можливих значень). Оскільки в багатьох методах класифікації шукається атрибут з максимальною ентропією, загальна точності класифікації може бути підвищена при кількості даних N_{onm} для яких значення атрибута A відповідають нормальному розподілу. На кожній ітерації пошуку поточна кількість даних n' збільшується на Δn, після чого перевіряється умова досягнення нормального розподілу. Для перевірки умови нормального розподілу значень атрибуту A (для $n_0 + m \cdot \Delta n$ даних, де m - номер ітерації) розраховується середнє статистичне значення μ, вибіркове середньоквадратичне відхилення σ^2 та робиться припущення H_0, що множина значень атрибуту A для усіх даних має розподіл $N(\mu, \sigma^2)$. Для перевірки припущення розраховується критерій Пірсона. Якщо для поточних даних $n + m \cdot \Delta n$ припущення щодо нормального розподілу відхиляється, пошук продовжується. При виконанні умови (1) та наявності нормального розподілу значень атрибута A пошук зупиняється, а поточний об'єм даних приймається як оптимальний N_{onm}.

Література. **1.** Реброва О. Ю. Статистический анализ медицинских данных. Применение прикладного пакета Statistica. М.: Медиасфера, 2002. 312 с **2.** Demidenko E. Sample size determination for logistic regression revisited // Statist. Med. 2007; 26:3385–3397 **3.** Bos S. How to partition examples between cross-validation set and fraining set? / Saitama, Japan: Laboratory for information representation RIKEN. 1995. 4 p. **4.** Perez-Cruz F. Kullback-Leibler divergence estimation of continuous distributions // IEEE International Symposium on Information Theory, 2008. **5.** http://goo.gl/YASBBu

Амосов О.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Проблема вибору базисного вейвлета в задачі ідентифікації музичних нот звукового сигналу

В основі неперервного вейвлет-перетворення лежить обчислення взаємної кореляційної функції сигналу $f(t)$ і вейвлета $w_{s\tau}(t)$. Взаємна кореляційна функція різних сигналів описує як ступінь подібності форми двох сигналів, так і їх взаємне розташування один щодо одного. Значення цієї функції, а, отже $W(\tau, s)$ тим більше, чим більш схожі поведінки сигналу $f(t)$ і вейвлета $w_{s\tau}(t)$ на інтегрованій ділянці. Отже, вейвлет $w(t)$, схожий в амплітудно-часовому поданні з досліджуваним сигналом $f(t)$, дасть великі значення в масиві $W(\tau, s)$, на відміну від вейвлета, що не має подібностей в часовому поданні з сигналом.

У загальному випадку вибір базисного вейвлета, (його порядку і значення параметрів), як правило, проводиться емпіричним шляхом для кожної окремої задачі. Однак відомо, що в задачах виявлення простих складових сигналу найкращі результати дають вейвлети, з частотним спектром схожим зі спектром елементарних складових сигналу [1].

Для більш вузької задачі - ідентифікації нот музичного інструменту, материнський вейвлет можна сформувати з локальної ділянки функції сигналу базової ноти цього інструменту.

Такі вейвлети при обчисленні $Wf_s(\tau)$ з сигналами однойменних музичних інструментів дають найбільші результуючі значення. Крім того, властивість автомодельності сигналів музичних інструментів дозволяє сформувати і застосовувати тільки один вейвлет $w(t)$ для ідентифікації будь-якої ноти одного музичного інструменту. Це пояснюється однотипністю масштабування базисного вейвлета $w_0(t)$ при формуванні сімейства вейвлетів і базової функції $n_0(t)$ музичного інструменту при отриманні будь-якої іншої ноти $n_i(t)$ цього музичного інструменту.

При формуванні базисних вейвлет-функцій на основі квазіперіодичних сигналів, у тому числі сигналів базових нот музичних інструментів $n_0(t)$, необхідна перевірка умов, що характеризують основні властивості вейвлетів:

1. Нульове середнє. Для формування вейвлета, на основі сигналу $n_0(t)$ необхідно використовувати періодичну ділянку сигналу $n_0(t)$ з нульовими початковим $n_0(t_0)$ і кінцевим $n_0(T + t_0)$ моментами, такими, щоб $\int\limits_{t_0}^{T+t_0} n_0(t)dt = 0$.
2. Обмеженість енергії. Сигнал музичного інструменту $n_0(t)$ не містить розривів першого і другого роду, тому дана умова завжди буде виконуватися для отриманого вейвлета.
3. Автомодельність. Наявність властивості автомодельності у вейвлет-функції забезпечується наявністю цієї властивості у сигналів базових нот $n_0(t)$ і самим методом масштабування базисних вейвлетов, закладеним у процедуру НВП.
4. Кускова неперервність. Функція базової ноти $n_0(t)$ на всьому інтервалі існування є неперервною, так як вона описує коливання фізичного тіла з кінцевою масою в часі і за визначенням не може мати розривів.

Таким чином, дотримуючись всіх чотирьох умов, висунутих до вейвлета, можна отримати функцію базисного вейвлета, придатну для застосування в НВП. На сьогоднішній день вейвлет-перетворення є досить потужним інструментом дослідження структури нестаціонарних даних [2], що дозволяє одночасно аналізувати структуру сигналів в різних діапазонах масштабів спостереження.

Література. **1.** Фадеев, А. С. Выбор вейвлет-функций для анализа музыкальной информации / А. С. Фадеев, Е. А. Кочегурова // Современные техника и технологии: Труды XII Междунар. научно-практ. конф. студентов и молодых ученых — г. Томск, 26—30 мар. 2007 г. — Томск: Изд-во ТПУ, 2006 — Т. 2. — С. 194—196.. **2.** Strang, G. Wavelet transforms versus Fourier Transforms / G. Strang // Bull. Am. Math. Soc. — 1993. — V. 28. — P. 288—305.

Базалий О.Ю., Бидюк П.И.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Интелектуальные методы анализа данных в оценивании кредитоспособности физических лиц

Как правило, для оценки риска кредитования заемщика используется скоринговая модель оценивания кредитоспособности физических лиц. Ее сущность состоит в том, что каждый фактор, характеризующий заемщика, имеет свою количественную оценку. Суммируя полученные баллы, можно получить оценку его кредитоспособности.

Основными недостатками скоринговой модели являются ее слабая адаптация под текущую ситуацию и большая вероятность появления ошибки при определении кредитоспособности потенциального заемщика, обусловленная субъективным мнением банковского специалиста. Поэтому в настоящее время широко используются интеллектуальные методы и технологии анализа и обработки данных, обладающие способностью адаптации, т.е. способностью автоматического учета вновь поступающих данных и подстройки модели под них.

Обсуждаются следующие интеллектуальные адаптивные механизмы, часто используемые на практике:

- логистическая регрессия, т.е. нелинейная модель для задач с бинарным результирующим полем, таким, как "выдать"/"отказать". Обычно на ее основе строятся скоринговые карты, позволяющие подобрать оптимальный и экономически обоснованный порог отсечения;
- деревья решений, т.е. способ представления принятия решений в древовидной структуре, каждый объект которой определяет бинарный результирующий переход, например, "выдать"/"отказать". К его достоинствам относятся простота интерпретации полученных результатов и автоматический выбор наиболее значимых факторов;
- нейронные сети, т.е. самообучающаяся "сетевая" технология, автоматически определяющая вклад каждого фактора в конечный результат. При этом она учитывает взаимное влияние факторов и позволяет принимать во внимание имеющуюся информацию и определять сложные, нетривиальные зависимости;
- высококачественным современным инструментом анализа кредитного риска являются Байесовские сети (БС). Это вероятностно-статистические модели в форме направленных ациклических графов, вершинами которых являются выбранные переменные, характеризующие заемщика, а дуги указывают на существующие причинно-следственные связи между вершинами. Модели в форме БС – это логическое дополнение к нелинейным регрессионным моделям, поскольку они дают возможность получить альтернативный результат оценивания кредитоспособности в вероятностной форме.

В данном исследовании решается задача сравнения характеристик различных методов и обсуждаются области их применения. Так, оказывается, что самым "грубым" методом оценивания является логистическая регрессия, и далее, в направлении повышения "точности", идут деревья решений, самоорганизующиеся карты, нейронные и БС, которые дают возможность достичь самых точных результатов. Тот же самый порядок сохраняется при направлении от "простоты" к "сложности". Т.е., получается, что самым простым и грубым методом является логистическая регрессия, а самый точный результат, но с максимальной степенью сложности обработки информации, дают нейронные сети. Таким образом, на практике нужно искать разумный компромисс в оценках "грубо-точно" и "просто-сложно" для того чтобы для решения задачи оценки кредитоспособности физического лица выбрать метод, адекватно решающий поставленную задачу и являющийся финансово выгодным с точки зрения суммарных затрат на его применение. Для комплексного решения задачи оценивания кредитоспособности заемщиков кредитов целесообразно разработать специализированную систему поддержки принятия решений, построенную на основе современных принципов системного анализа.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Баклан В.С.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Анализ инвестиционной привлекательности акций аукционных домов с применением нечетких нейронных сетей Мамдани

Введение. В современной экономике вопрос инвестиций играет немаловажную роль. В поисках максимальной выгоды предприниматели ищут наиболее перспективные для инвестирования сферы. В условиях нестабильной экономической ситуации, кризиса имеет смысл не зацикливаться на традиционных нишах, а исследовать новые направления для вложений. Среди таких уже давно оказался арт-рынок. Считается, что инвестиции в предметы искусства – картины, скульптуры, старинные книги и т.д. – удел истинных ценителей и знатоков. Для тех же, кто не готов самолично вкладывать деньги и время в объекты коллекционирования, существуют более простые и надежные способы извлечь прибыль из нового инвестиционного направления, - приобрести акции аукционных домов.

В связи с тем, что 90% рынка аукционных домов занимают два крупнейших мировых представителя Christie's и Sotheby's, которые, к тому же, в последнее время показывают невиданные ранее успехи – 2013-2014 год стал рекордно успешным для арт-рынка [3] – имеет смысл в качестве примера и наглядного пособия рассматривать именно их.

Цель работы. Рассмотреть рынок современного и классического искусства на предмет его инвестиционной привлекательности, оценить возможную выгоду от вложений в акции самых известных аукционных домов мира, сравнить возможную выгоду или убытки с инвестициями в другие сферы, дать прогноз и составить наиболее эффективный алгоритм, который позволит инвестору получить максимальную прибыть в краткосрочной или долгосрочной перспективе, если такой возможен.

Методы. В качестве инструмента для исследования различных областей инвестиций применимо использование систем нечеткого логического вывода (систем НЛВ) и построенных на их основе нечетких нейронных сетей (ННС).

Выбор именно нейронных сетей обусловлен рядом их преимуществ [2] перед традиционными вычислительными системами, они:

- позволяют решать задачи при неизвестных закономерностях
- устойчивы к шумам во входных данных
- обладают потенциальным сверхвысоким быстродействием

Таким образом апарат НС позволяет решать многочисленные задачи принятия решений в условиях неопределенности и неполноты информации, в частности – прогнозирования [1].

Результаты. В данной работе рассмотрены и использованы нечеткие нейронные сети Мамдани для функции принадлежности Гаусса. Приведен алгоритм построения и работы метода, рассмотрены его преимущества перед другими алгоритмами и основные недостатки, с помощью тестовых значений с заранее известными реальными результатами дана оценка его эффективности и оценена погрешность. Также проведена оценка инвестиционной привлекательности рынка современного искусства в ближайшей и долгосрочной перспективе на основе имеющихся на текущий период данных. На основе полученных результатов сделаны выводы о том, являются ли акции аукционных домов Christie's и Sothebey's инвестиционно привлекательными как сами по себе, так и по сравнению с другими сферами инвестиционной деятельности.

Литература. **1.** Зайченко Ю.П. Нечеткие модели и методы в интеллектуальных системах. - К.: Изд. «Слово»,-2008. **2.** Преимущества нейронных сетей [Электронный ресурс]. – Режим доступа: http://www.aiportal.ru/articles/neural-networks/advantages.html. **3.** Арт-рынок в 2014 году: рекордный рост после рекордного падения [Электронный ресурс]. – Режим доступа: http://www.artguide.com/posts/663.

Барсученко В.В.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Применение нейронной сети Back Propogation для стабилизации изображения на смартфонах в условиях нестационарной динамики изменения их положения

Работа посвящена описанию и исследованию применения нейронной сети Back Propagation для стабилизации изображения на смартфоне в условиях нестационарной динамики изменения его положения и нахождения оптимальных значений параметров НС.

В данный момент наблюдается повышенный спрос на мобильные устройства, такие как смартфоны, планшеты, электронные книги. Это привело к тому, что традиционная бумажная литература, которая читалась в транспорте или при ходьбе, стала активно заменяться чтением электронных устройств. Однако изображение на мобильных устройствах при передвижении имеет нестационарную динамику часто с сильными амплитудами. Для решения данной проблемы в работе рассматривается подход к стабилизации изображений на смартфонах в условиях нестационарной динамики изменения их положения, который основывается на прогнозировании положения смартфона с применением нейронной сети Back Propagation [1, 2].

Стабилизация изображения. Для стабилизации изображения используется гироскоп, как основной, и акселерометр - как дополнительный. Гироскоп показывает угловую скорость оборота смартфона вокруг осей, а акселерометр – ускорение вдоль оси. И так как вектор скорости при изменении устройства в пространстве менее чувствителен по направлению чем вектор ускорения, данный датчик является более устойчивым к шумам и является более точным. С помощью угловой скорости можно найти угол, на который было повернуто гаджет. Соответственно стабилизация заключается в изменении положения изображения с отрицательной зависимостью относительно угла.

Так как человек во время ходьбы в основном держит устройство одной рукой и за нижнюю часть, вращение будет происходить только вокруг оси Oz. А динамика вдоль оси Oy будет определена с помощью акселерометра. Основной проблемой стабилизации изображения с помощью сенсоров является то, что снятие данных с датчиков происходит после изменения положения, то есть сдвиг изображения будет происходит с некоторой задержкой, учитывая время выполнения программного кода, что есть недопустимо при большой динамике. Поэтому целесообразно использовать прогноз на $(k+1)$ или $(k+m)$ шаг.

Существует множество методов прогнозирования рядов. Но в нашем случае нужно не только прогнозировать на основе предыдущих данных, но и обучать систему на данной выборке. Потребность в обучении обусловлена тем, что необходима информация о величине динамики изменения гаджета в пространстве. Поэтому целесообразно использовать нейронную сеть Back Propogation.

Предлагаемая математическая модель. Поскольку используется только с одна ось Oz, то пусть $x(t)$ - координата, которая определяется гироскопом в момент времени t. Следовательно на вход Нейронной сети Back Propogation подается ряд $x(t), x(t-1), ..., x(0)$. А на выходе соответственно будет $x(t+1)$

Для анализа и прогнозирования ВР необходимо выявить его полезные составляющие и максимально исключить случайные(погрешность аппаратуры, ошибки измерений и вычислений, субъективные ошибки участников процесса измерений и анализа данных).

В случае, когда распределение экспериментальных данных не поддается описанию каких либо явных закономерностей может оказаться эффективным моделирование ВР случайным процессом.

В качестве базовой модели обычно принимается нормальный белый шум [3] $a(t)$ - случайный процесс, значение которого в различные моменты времени статистически независимы, а значение в любой фиксированный момент времени распределены одинаково по нормальному

закону:

$$p(a) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(a-M)^2}{e^{2\sigma^2}}} \quad (1)$$

где σ^2 - дисперсия

Для полного описания нормального случайного процесса (т.е. процесса, для которого одномерное распределение в любой момент времени является нормальным) достаточно задать только его среднее значение, дисперсию и автокорреляционную функцию. Поскольку значения автокорреляционной функции при любой ненулевой задержке равны нулю, то для построения модели в виде нормального белого шума по данным нужно оценить только его дисперсия и среднее значение [3]

Следовательно нужно найти прогнозируемое $M(t+1)$ и $\sigma(t+1)$ на основе рядов $M(t), M(t-1), ..., M(0)$, $\sigma(t), \sigma(t-1), ..., \sigma(0)$ соответственно для нахождения требуемого $x(t+1)$

Прогноза на 1 шаг вперед может быть недостаточно в связи с задержкой выполнения программного кода и считывания данных с сенсорного датчика поэтому для минимизации задержки между зрительным контактом пользователя и прогнозируемым значением далее рассматривается прогнозирование на n шагов вперед.

Известно, что любая нейронная сеть есть функцией. Таким образом, координаты гироскопа на текущем шаге зависят от координат на n шагов назад и еще несколько шагов, поэтому следующая зависимость имеет вид:

$$x(t+n) = f(x(t), x(t-1), ..., x(t-n)) \quad (2)$$

А критерий, который необходимо оптимизировать имеет следующий вид:

$$E = \sum_{t=n+1}^{T} (x(t+n) - f(t,n))^2 \quad (3)$$

Веса сети настраиваются с целью минимизации значения данного критерия. Следует отметить, что - это количество периодов времени, от которых, зависит прогноз.

Параметры экспериментов. В состав параметров экспериментов входят:1) параметры нейронной сети (количество входящих нейронов, количество скрытых нейронов, количество скрытых слоев, скорость обучения, количество итераций отведенных под поиск минимального критерия); 2) шаг прогнозирования НС $(k+1, k+2, \ldots, k+n)$; 3) частота снятий показаний с сенсорных датчиков.

Инструментарий сравнительного анализа результатов экспериментов. Эффективность реализации параметров экспериментов состоит из: 1) минимального числа итераций, необходимых для определения оптимального значения прогноза;2) минимальной частоте снятий показаний сенсорных датчиков; 3) минимальному отклонению от заданной траектории устройства в пространстве, оцениваемую экспертом

Содержание доклада. В предлагаемом докладе будут представлены результаты практических исследований предложенной математической модели, а также сравнительный анализ различных параметров экспериментов.

Литература. **1.** Зайченко Ю.П. Нечёткие модели и методы в интеллектуальных системах. Учебное пособие для студентов высших учебных заведений. - К.: «Издательский Дом «Слово»», 2008. - 344 с.

2. Michael A. Nielsen, “Neural Networks and Deep Learning”, Determination Press, 2015.

3. Безрученко Б.П., Смирнов Д.А. Статистическое моделирование по временным рядам [Электронный ресурс] Сарат. отд-ние Ин-та радиотехники и электроники РАН. - Электрон. дан. - Саратов, 2000. - Режим доступа: http://window.edu.ru/window/_catalog/files/r29971/sgu025.pdf, свободный. - Загл. с экрана.

Бідюк П.І., Устимчик М.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Використання регресійних моделей для розв'язання задач медіапланування

Медіапланування – це вибір оптимальних каналів розміщення реклами, що проводиться з метою досягнення максимальної ефективності рекламної кампанії.

На жаль, на сьогодні, в процесі медіапланування дуже рідко використовують дані, які були отримані в процесі проведення аналогічних/схожих активностей (реклами) в минулому. В своєму дослідженні візьмемо, як приклад, компанію X та виконаємо повний аналіз поточного стану і планування активностей на наступний рік. Таким чином, досліджується діяльність компанії, що продає продукт Y, а для стимулювання продажів використовується ряд активностей, які будемо позначати через $k(i)$, де i - показник активності. Прикладами активностей можуть бути такі:підсвітка продукту на пункті продажів, залучення хостес до реклами продукту, проведення вечірок у підтримку продукту.

Нехай для підтримки продукту Y компанія використовує шість активностей $i=6$. Для дослідження маємо вибірки даних за деякий період:

- продажі компанії продукту Y за якийсь період (в тому числі період, де не проводились підримки продукту) у кількості та вартості;
- ряд активностей у розрізі вартості

Таким чином, необхідно виконувати обробку даних у формі часових рядів. В даній роботі необхідно визначити оптимальний розподіл матеріальних та часових ресурсів на підтримку активностей з метою отримання максимального прибутку. Отже, потрібно визначити час запуску, тип активності та час її закінчення для вибраних регіонів України.

Вибір методу прогнозування – це досить непроста задача, яка часто потребує одночасного застосування декількох альтернативних методів і вибору кращого з них на основі аналізу отриманого результату.

Найпопулярнішими на сьогодні методами прогнозування розвитку процесів довільної природи є такі: методи регресійного аналізу, нечітка логіка, ймовірнісні методи, метод групового врахування аргументів (МГВА)нейронні мережі, методи на основі м'яких обчислень, метод подібних траєкторій та деякі інші.

Для розв'язання задачі прогнозування скористаємось такими лінійними моделями: авторегресією (АР), авторегресією з ковзним середнім (АРКС), змішаною регресією (АРКС плюс додаткові регресори). Також застосуємо нелінійну регресію типу логіт, яка надає можливість планувати події у випадку, коли основна змінна у лівій частині приймає два значення, наприклад, високий та низький рівні продаж. Залежною змінною будуть продажі компанії для продукту Y, а незалежними змінними буде ряд активностей $k(i)$.

Результати прогнозування, отримані за лінійними та нелінійними моделями, будуть використані для побудови плану реалізації активностей на наступний період часу. Таким чином можна сформувати альтернативні плани (з використанням оцінок прогнозів за різними типами моделей) та вибрати кращу альтернативу за допомогою допоміжних критеріїв. Для практичного розв'язання поставленої задачі медіа планування проектується та реалізується програмно система підтримки прийняття рішень (СППР), яка забезпечить належну обробку даних, побудову прогнозуючих математичних моделей та формування альтернативних планів. СППР будується на основі принципів системного аналізу, які передбачають ієрархічність структури системи, наявність контролю обчислювального процесу на кожному етапі обробки даних та врахування можливих невизначеностей за допомогою методів інтелектуального аналізу даних.

Література. **1.** http://www.markint.ru/mediaplanirovanie-v-reklame

Виклюк Я.І., Леко Т.Б.
ПВНЗ "Буковинський університет", Чернівці, Україна

Розробка simulink моделі для вивчення залежності між появою сонячної активності та виникненням ураганів

На сьогодні однією з актуальних тем є дослідження причин виникнення ураганів, оскільки для ефективного подолання наслідків цих природних явищ, необхідно впроваджувати нові методи їх прогнозування. Вони базуються на розробці нових математичних моделей, які дозволяють передбачити моменти виникнення ураганів. Проте існуючі на сьогодні методи не пов'язують рух повітряних мас із сонячною активністю. І, відповідно, не враховують активність сонця, як чинник, що впливає на появу ураганів. В своїх наукових працях професор Радовановіч запропонував «геліоцентричну» гіпотезу, згідно якої зародження значної кількості ураганів може бути обумовлене спалахами на сонці [1].

Тому метою нашої роботи була перевірка цієї гіпотези на основі статистичних даних про виникнення ураганів в Атлантичному та Тихому океанах методами DataMining та розробка інформаційної системи на основі Simulink. Навчальна вибірка містила характеристики активності сонця на період з травня по жовтень від 1999 до 2013 роки та число ураганів в Атлантичному, західному та східному узбережжях Тихого океану. В процесі експерименту враховувалась часова затримка між вхідними та вихідними факторами (lag=0-3).

Так, як загальна кількість моделей, які необхідно побудувати, сягає декількох тисяч, для реалізації вищезазначених експериментів у середовищі SIMULINK була розроблена програма, що автоматизує всі кроки експериментів. Ця програма складається з трьох частин. Перша частина імпортує дані в робочу область Matlab за допомогою окремого скрипта, реалізує кореляційний та R/S аналіз (correlation, my_R/S блоки відповідно) та формує навчальну вибірку без пропусків, сезонності та трансформовану з врахуванням лагової залежності (блоки my_interp, my_season_matrix). Результати передаються в робочу область та зберігаються в окремій структурі, яка експортується на жорсткий диск за допомогою останнього блоку my_save (рис. 1 a).

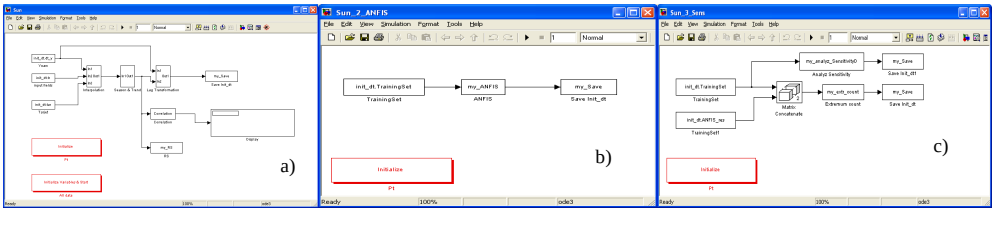

Рис. 1. Програма розрахунку в середовищі Matlab&Simulink a) попередня обробка даних, b) навчання ANFIS, c) аналіз чутливості

В другій частині програми, що є най тривалішою в часі, блок my_ANFIS відповідає за навчання нейронної мережі чи пошук ключових факторів. Цьому блоку відповідає окрема розроблена функція в Matlab. В цій програмі дані експортуються та імпортуються у файл даних, що був створений попередньою програмою (рис. 1 b).

Третя частина відповідає за аналіз чутливості. Це дало змогу встановити функціональну залежність кількості ураганів від ключових факторів (рис. 1 c).

Побудована засобами імітаційного моделювання, програма дозволяє візуалізувати процес аналізу, і, відповідно, автоматизувати всі етапи не залежно від початкових даних.

Література. **1.** Radovanović M, Pecelj M. 2012. The Heliocentric Hypothesis on the Origin of Forest Fires. International Scientific Conference "Forestry Scienceand Practice for the Purpose of Sustainable Development of Forestry" 1-4th November, 2012, Banja Luka, Republic of Srpska/B&H, Abstractbook, p. 10.

Вініченко А.М.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Семантичні композитні веб-додатки в наукових дослідженнях

Типовими компонентами роботи дослідника, що займається аналізом предметної області, є збір фактів, побудова моделі предметної області, співвіднесення власної моделі з моделями інших дослідників. Вимоги до збору та організації доступу до масивів даних, труднощі розвитку та роботи з комплексними моделями даних, необхідність обчислень (що здійснюються за допомогою комп'ютерних систем високої продуктивності або за рахунок розподіленої мережі обчислень) і розуміння того, що рішення багатьох наукових проблем знаходиться на стику наукових дисциплін, ставить завдання розвитку інформаційних технологій в науці.

Одне з таких завдань — це знаходження нових фактів та інформації (оптимальних рішень) на основі існуючих даних. Використання семантичних веб-додатків може допомогти вирішити дану проблему. Протягом дослідження проаналізовані кілька сценаріїв застосування семантичних додатків в наукових дослідженнях. В основі таких додатків лежать бази знань, які в свою чергу реалізовуються за допомогою онтології. Бази знань наукових досліджень будуються на основі різних зв'язків. Підтримка семантичних зв'язків явним чином призводить до утворення багатошарової семантичної структури контенту електронної бібліотеки, кожен шар якої відповідає деякій властивості семантичних зв'язків. Така структура може формуватися на основі онтології зв'язків і може служити джерелом інформації для проведення якісно нових наукометричних вимірювань, для дослідження структурних властивостей корпусу знань у різних галузях науки [1].

В останні роки на основі бібліографічних посилань у публікаціях, що випускаються в авторитетних періодичних виданнях, почали створюватися індекси цитування, які формують бібліометричну статистику. Зв'язки цитування в текстових публікаціях зазвичай представляються неструктурованим чином у вигляді списку використаної літератури.

Поряд зі зв'язками цитування між документами наукових електронних бібліотек існують різноманітні інші семантичні зв'язки. Наприклад, зв'язок може вказувати, що її цільовий документ містить наукові результати, що базуються на результатах, описаних у вихідному документі зв'язку, або що у вихідному документі зв'язку спростовується результат, викладений у її цільовому документі. Зв'язок може також вказувати, що вихідний її документ є новою редакцією цільового документа або що він являє собою його складову частину, наприклад, анотацію. Існує велика різноманітність семантичних зв'язків, які можна при необхідності підтримувати між документами бібліотеки [2].

В результаті виконаних досліджень проаналізовано процеси пов'язані з науковою діяльністю, що спричиняють виникнення потреби формування семантичних зв'язків між публікаціями, а в середовищі електронної бібліотеки — між представленими в ній інформаційними об'єктами. Аналізуються також категорії семантичних зв'язків, які при цьому породжуються, розглянуто властивості структури семантичних зв'язків, її багатошаровий характер, семантика зв'язків, питання їх класифікації, відомі онтології, створені для цих цілей, і онтологія, що використовується в прикладі.

Використання онтології грунтується на таких її характеристиках як:

- децентралізація — можливість роботи з базою знань декількох сайтів на одній онтології;
- логічний висновок — можливість породження нових фактів;
- природне наслідування і вкладеність;
- опис предметної області, метадані.

У підсумку, такого роду семантичний додаток і одержувана його засобами інформація відкривають нові можливості для розвитку наукових досліджень, забезпечують нові технології для наукової та науково-організаційної діяльності.

Література. **1.** D. Shotton. (2010, October 14). *Introduction the Semantic Publishing and Referencing (SPAR) Ontologies* [Online]. Available: https://opencitations.wordpress.com/2010/10/14/introducing-the-semantic-publishing-and-referencing-spar-ontologies/. **2.** D. Shotton. (2011, January 19). *Semantic annotation of publication entities using the SPAR (Semantic Publishing and Referencing) Ontologies* [Online]. Available: http://imageweb.zoo.ox.ac.uk/pub/2011/presentations/Shotton&Peroni_Semantic_Annotation_Using_SPAR_Ontologies.pdf.

Гоцко У.І.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Використання "fingerprinting" методу для ідентифікації сигналів звукового оточення людини

В повсякденному житті нас оточує дуже велика кількість різноманітних звуків: слова, музика, кроки, стук у двері, шум води, свист чайника, сигнал машини і т.д. Кожен звук несе в собі якусь інформацію, тому важливим є правильно та своєчасно інтерпретувати почуте. Ця проблема є особливо актуальною для людей з вадами слуху, а також в задачах створення систем безпеки, в основі яких лежить аналіз звукового оточення.

В даній статті розглядається алгоритм ідентифікації немовленнєвих звукових сигналів з їх подальшою інтерпретацією в зрозумілому для користувача вигляді. Алгоритм базується на так званому процесі створення "відбитку" ("fingerprinting" - процес, який дозволяє отримати з аудіозразка хеш-маркери) і полягає в наступному: користувач записує базу звуків - зразків, які зазвичай зустрічаються в його повсякденному житті. Формується база "відбитків" зразків. Дальше в режимі реального часу аналізуються оточуючі звуки, відділяється тишина/шум від інформативних звуків. До відрізку інформативного сигналу застосовується "fingerprinting". Після цього хеш інформативного відрізку порівнюється із базою зразків.

Розглянемо "fingerprinting" метод докладніше. Відрізок звуку можна представити у вигляді частотно-часового графіку - спектрограми. На одній осі відкладають час, на іншій - частоту, на третій - інтенсивність. Кожна точка на графіку представляє інтенсивність конкретної частоти в даний момент часу. Розташовуючи часову шкалу на осі x, а частотну - на осі y, одержуємо горизонтальну лінію, яка являє собою безперервний чистий тон і вертикальну лінію, яка ілюструє миттєвий сплеск білого шуму. "Fingerprinting" створює відбиток звукового зразка шляхом сторенння трьохвимірного графіку та виявлення на ньому частоти "піку інтенсивності". Для кожного з таких пікових значень фіксується частота і проміжок часу відносно початку звукового запису. Зберігати відбитки можна, наприклад, у вигляді хеш-таблиці, в якій роль ключа відіграє частота [1].

Процес порівняння відрізка звукового запису з базою хешів відбувається наступним чином: після отримання "відбитку", як описано вище, використовується перший ключ для пошуку відповідного запису - частота. Якщо у відрізку звукового запису було виявлено декілька співпадінь, то перевіряється відповідність частот у часі. Тобто, якщо зразок і невідомий запис мають багато точок співпадання частот, але вони не співпадають у часі, то відповідності між звуковими записами немає. Якщо ж частоти співпадають в один і той самий момент часу, то відповідність між записами присутня.

Варто відмітити, що "fingerprinting" метод має безліч застосувань в області звукового аналізу, як, наприклад, ідентифікація музичних пісень (програмне забезпечення розроблене компаніями Shazam, SoundHound), виправлення некоректних музичних метаданих (програмне забезпечення розроблене компанією Last.fm) та ін. [2].

Література. **1.** Avery Li-Chun Wang "An Industrial-Strength Audio Search Algorithm" [Електронний ресурс]. - Режим доступу: https://www.ee.columbia.edu/dpwe/papers/Wang03-shazam.pdf **2.** Nick Palmer "Audio Fingerprinting" [Електронний ресурс]. - Режим доступу: http://www.palmnet.me.uk/uni/FYP/Audio%20Fingerprinting.pdf

Даниляк О.С., Бідюк П.І.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Проектування системи підтримки прийняття рішень для оцінювання і прогнозування стану підприємств

Системи підтримки прийняття рішень (СППР) набули широкого застосування в економіках передових країн світу і кількість їх впроваджень у практику постійно зростає. На рівні стратегічного управління СППР використовуються, зокрема для довго-, середньо- і короткострокового прогнозування, а також для фінансового планування, бухгалтерського обліку, розв'язання задач логістики і т. ін. СППР, орієнтовані на оперативне управління, застосовують у галузях маркетингу (прогнозування та аналіз збуту, дослідження доступних сегментів ринку і цін), при виконанні науково-дослідних та конструкторських робіт, в управлінні кадрами. Оперативно-інформаційні застосування пов'язані з оптимізацією виробничих процесів, придбанням та обліком товарно-матеріальних запасів, їх фізичним розподілом і т. ін. [1].

Створення СППР для прогнозування динаміки розвитку підприємств розглядається як послідовність задач, які включають розробку функціональних та вимог користувача до системи, архітектури і функціональної структури та інтерфейсу користувача, вибору (побудови) та реалізації методів обробки даних і представлення результатів. В системі програмно реалізовані методи обробки даних та критерії якості (адекватності) моделей і прогнозів, які виконують такі функції:

1. Попередня обробка даних;
2. Ідентифікація типу часового ряду;
3. Кореляційний аналіз даних;
4. Оцінювання параметрів математичних і статистичних моделей;
5. Критерії вибору кращих моделей з множини оцінених кандидатів;
6. Критерії оцінювання якості прогнозу;
7. Правила прийняття рішень щодо вибору кращого прогнозу.

Відмінною рисою запропонованої СППР є контроль обчислювальних процесів на кожному рівні її ієрархії: – попередня обробка даних (критерії якості даних); – оцінювання структури і параметрів математичних і статистичних моделей (критерії адекватності моделей); – оцінювання якості прогнозів (статистичні критерії якості прогнозів); – критерії якості альтернативних рішень, що формуються на основі прогнозів. Крім того, у системі передбачається ідентифікація, поглиблений аналіз, математичний опис та врахування можливих невизначеностей у процесі обчислень, згаданих вище. Для опису та обробки невизначеностей (тобто зменшення їх впливу на якість проміжних та остаточних результатів) застосовано, зокрема, такі інструменти: – методи заповнення пропусків даних, обробки екстремальних значень та цифрової фільтрації; – адаптивний фільтр Калмана для врахування зовнішніх випадкових збурень на досліджувані процеси; – функціонально повні процедури оцінювання структури моделей для звуження множини допустимих моделей-кандидатів; – множина альтернативних методів оцінювання параметрів математичних моделей, яка забезпечує отримання незміщених консистентних оцінок параметрів; – множина альтернативних методів обчислення оцінок прогнозів та їх комбінування.

Твердження. Таким чином, можна стверджувати, що запрограмована функціональність СППР із врахуванням можливостей обробки невизначеностей процесу моделювання і прогнозування, а також контролю обчислювальних процесів за різними множинами критеріїв якості забезпечує умови отримання оцінок прогнозів прийнятної якості.

За допомогою розробленої СППР вже розв'язано ряд прикладних задач аналізу стаціонарних та нестаціонарних фінансово-економічних процесів, оцінювання та прогнозування стану виробничих підприємств на основі статистичних даних стосовно динаміки їх розвитку. Для розв'язання таких задач у системі передбачено побудову та використання нелінійних моделей у

формі логістичних функцій, лінійних моделей регресійного типу (моделі Альтмана, Спрінгейта, Ліса, Тоффлера, Чессера та інші), моделі на основі правил нечіткої логіки, байєсівські мережі та регресійні прогнозуючі моделі типу авторегресії (AP(p)) і авторегресії з ковзним середнім (APRC(p, q)). Останні забезпечують можливість отримання багатокрокових прогнозів без використання проміжних оцінок [3]. Альтернативні (ймовірнісні) оцінки станів обчислюються за допомогою динамічних байєсівських мереж. Функція комбінування оцінок прогнозів дає можливість підвищувати якість прогнозування в умовах застосування ідеологічно різних прогнозуючих моделей [4,5]. Цей метод ефективний якщо дисперсії похибок окремих прогнозів суттєво не відрізняються. Серед розв'язаних прикладних задач також є математичне моделювання та прогнозування процесів ціноутворення на нафтопродукти на ринку України, аналіз демографічних процесів, прогнозування цін деяких активів на біржі. Для обчислення оцінок прогнозів будуються функції прогнозування (їх можна табулювати), які дають можливість знаходити коректні значення короткострокових та середньострокових прогнозів [3,4]. Встановлено, що оцінки прогнозів, отримані за допомогою функцій прогнозування, побудованих на основі лінійних моделей, мають високі характеристики якості: – нульове математичне сподівання похибок оцінок прогнозів не залежно від кількості кроків прогнозування (тобто відсутність зміщеності оцінок); – обмеженість дисперсії похибок оцінок прогнозів деякою константою.

Виконано порівняння якості прогнозів, отриманих на основі моделей авторегресії з ковзним середнім різних порядків, із прогнозами, оціненими за методами експоненційного згладжування, групового врахування аргументів (чіткий та нечіткий), фільтра Калмана, методу опорних векторів та використаних ймовірнісних моделей. Всі вказані методи використовувалися для оцінювання одно- та багатокрокового прогнозів вибраних змінних, що характеризують поточний стан досліджуваних підприємств. Виконуються роботи з розробки та реалізації нових методів оцінювання моделей на основі процедур Монте Карло для марковських ланцюгів (МКМЛ). Методи цього класу характеризуються універсальністю застосування та можливістю отримання прийнятних за якістю оцінок параметрів математичних моделей. Аналізується можливість введення в систему модуля автоматичної класифікації рядів статистичних даних та вибору належного класу математичних моделей для їх опису і подальшого використання для розв'язання задач прогнозування і та прийняття управлінських рішень. Також розглядається задача оцінювання довгострокових прогнозів на основі моделей трендів та застосування методу подібних траєкторій.

Література. **1.** Олексюк О.С. Системи підтримки прийняття фінансових рішень на мікрорівні. – Київ: Наукова думка, 1998. – 508 с. **2.** Бідюк П.І., Баклан І.В., Савенков О.І. Часові ряди: моделювання і прогнозування. – Київ: ЕКМО, 2003. – 143 с. **3.** Бідюк П.І., Романенко В.Д., Тимощук О.Л. Аналіз часових рядів. – Київ: НТУУ «КПІ», 2013. – 600 с. **4.** Tsay S.A. Analysis of financial time series. – New York: John Wiley & Sons, 2010. – 713 p. **5.** Zou H., Yang Y. Combining time series models for forecasting // Int. J. of Forecasting, 2004, Vol. 20, pp. 69 – 84.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Дембовский О.Ю.
Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины, Киев, Украина

Проблема операционализации понятия устойчивого развития в системах поддержки решений по планированию развития городов (регионов)

Введение. Изучение вопросов устойчивого развития продолжаются со времени первого доклада комиссии Брунтланд 1987 года [1] , на уровне ООН и международных организаций, исследовательских проектов, и программ ЕС (например, SPARTACUS и PROPOLIS), местных властей, но многие стратегические документы все еще отстают от целостного понимания принципов устойчивого развития в аспекте их практического учета [2]. Украинская "Концепция устойчивого развития насленных пунктов" [3] в этом смысле тоже не лишена недостатков.

Несмотря на внушительное число публикаций по теме многокритериальной поддержки решений в сложных системах, разнообразие методов (одна из попыток систематизации которых для условий определенности показана на рис.1) и ряд новаций, включая активное применение нечеткой математики, а также массу готовых компьютерных приложений, проблемы построения и корректного применения методологического базиса принятия решений, адекватного условиям конкретной задачи, остаются в целом нерешенными.

1. Системный контекст проблемы. Рассматривая проблему принятий решений в области планирования будущего для города (региона), необходимо учитывать различные типы *неопределенности*, а именно:

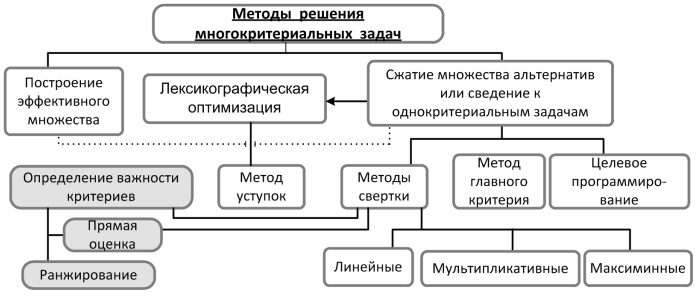

Рис. 1. Многокритериальные методы и процедуры в ПР (на основе [4])

- *общесистемную*, отражающую динамическую сложность причинно-следственных связей;

- *политическую*, вследствие политических изменений (объема властных полномочий ЛПР (лица, принимающего решение), доминирующих элит, состава и действий акторов и стейкхолдеров);

- *управленческую*, связанную с изменением приоритетов отдельных целей управления по достижении целевого оптимума значений некоторых показателей, изменением существующей миссии, новых приоритетов развития местного сообщества.

2. Подходы к трактовке понятия устойчивости. В определении смысла, вкладываемого в понятие устойчивого развития, наметились следующие отличающиеся подходы:

- *капитальный подход* (экономисты теоретки) <-> *интегрированные подходы* (прагматики);

- *сильная* <-> *слабая устойчивость* (акцент на консервацию природных ресурсов или на возможность их замещения, что примыкает к капитальному подходу);

- *приоритет потребностей будущих поколений* <-> *справедливость* в удовлетворении потребностей между нынешним и будущими поколениями (равновесие).

3. Подходы к решению проблемы. В терминах исследования операций, предполагая, что конечное множество альтернатив и ограничения уже заданы, выбирают малое количество приемлемых вариантов, на чем решение задачи оканчивается.

В постановке задачи многокритериального выбора в условиях определенности сначала ищется множество ээфективное множество недоминируемых альтернатив (Эджворта-Парето); далее действуют в зависимости от размера полученого множества и с учетом конкретного

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

контекста задачи. Ключевым моментом является корректное построение системы критериальных ограничений и системы параметров, которых объективно может быть несколько десятков. В определении группы лучших решений после сужения его до Парето-множества в условиях определенности могли бы далее применяться известные методы выбора (рис.1). Однако в силу указанной выше неопределенности (п.1) применение методов ранжирования или прямой оценки критериев, как и методы свертки, в чистом виде часто некорректны.

Ряд исследователей прямо критикует известный метод анализа иерархий (МАИ) Т.Саати (см., например, В. Подиновский [5] и дальнейшие его публикации относительно МАИ). Также следует учитывать известные в литературе проявления различных психологических эффектов и индивидуальных предпочтений у ЛПР и экспертов в процессах ПР.

Сведение проблемы к задаче ПР в условиях неопределенности (с использованием одного из вероятностных критериев) тоже не выглядит обоснованным - формально, любые предположения о распределении вероятностей будущего состояния объекта задачи ПР некорректны, поскольку в значительной мере являются результатом действий ЛПР, акторов и внешней среды на горизонте планирования (10-20 лет и больше).

Для уровня планирования города предлагается следующая интерактивная последовательность действий в многокритериальной задаче ПР:

1) доформальный анализ и выбор ограничений (ресурсные, деятельности, отходов), набора ключевых показателей (индикаторов); 2) генерирование множества альтернатив; 3) сжатие их до Парето-множества; 4) анализ состава решений по достижению плановыми действиями критериальных результатов (на временном горизонте примерно 3-5 лет), дополнительных условий (бюджетных ограничений, отношений издержки/выгоды и т.п.); 5) выбор известными методами небольшого числа “лучших” решений; 6) мониторинг индикаторов, с интервалом в 1 год и повторением шагов 2)-5) и корректировкой плановых действий. Прикладное интерес имело бы применение в конкретной задаче нескольких известных методов ПР (на шаге 5).

3. Выводы и направления дальнейших исследований. Рассмотрена проблема операционной постановки задач ПР в планировании разаития городов (регионов), предложены подходы к ее решению, дальнейшие преспективы имеют исследования в области методологии применения существующих подходов. Аспекты “справедливости между поколениями” и уровня устойчивости[1], как и капитального подхода (п.2) могут учитываться при разработке экономико-математических моделей межотраслевого типа и финансовой математики (при преодолении проблемы монетарной оценки, капитала, ресурсов и потребления). Перспективной выглядит также аналитическая формализация трактовки устойчивого развития в виде “займа у будущих поколений” с использованием соответствующего понятийного и математического аппарата.

Литература. **1.** Наше общее будущее: докл. Междунар. комиссии по окружающей среде и развитию (МКОСР)//пер. с англ. – Нью-Йорк, 1987. – 412 С. http://www.un.org/ru/ga/pdf/brundtland.pdf (дата доступа - 15.08.2014)

2. Дембовский О. Ю. Устойчивое развитие: анализ определений, источники, глобальный и местный контексты // Моделювання та інформаційні технології. Збірник наукових праць. Інст. пробл. модел. в енергетиці. – К.: 2013. – Вип. 69. – с. 85-99

3. Про Концепцію сталого розвитку населених пунктів. Верховна Рада України; Постанова від 24.12.1999 № 1359-XIV, http://zakon4.rada.gov.ua/laws/show/1359-xiv (дата доступа - 12.02.2015)

4. Уткин Л.В. Многокритериальное принятие решений/[Электронный ресурс]. http://www.levvu.narod.ru/Papers/Multicrit.pdf (дата доступа - 18.03.2015).

5. Подиновский В.В., Подиновская О.В.О некорректности метода анализа иерархий// Проблемы управления.– М.: 2011. – № 1. – с. 8-13

[1]Вопрос в рамках обычного горизонта планирования (10-25) подлежит учету в моделях ПР установкой жестких сроков достижения индикаторов, лимитирующих потребления (расход) природных ресурсов, норм переработки отходов

Карпенко Д.І.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Дослідження методів згладжування даних зі збереженням фази

Двовимірні зашумлені масиви даних часто представлені зображеннями у відтінках сірого, що надає можливість візуально оцінити якість позбавлення від шуму.

Позбавлення зображень шуму, як правило, здійснюється наступним способом: зображення перетворюється в якійсь області, де шумова складова більш легко ідентефікована, потім застосовується операція порога для видалення шуму, і, після цього, відбувається перетворення в інвертоване реконструйоване і, сподіваємося, вільне від шуму зображення. Вейвлет-перетворення виявилося дуже успішним у розрізненні сигналу і шуму: енергія з сигналу направляється в обмежену кількість коефіцієнтів які «відділяються» від шуму. Після цього відбувається шумозаглушення вейвлет-стиснення, що полягає у визначенні величини вейвлет-коефіцієнтів, яку можна очікувати від шуму (визначення порогу), а потім зменшення величини всіх коефіцієнтів на цю суму та застосування оберненого перетворення задля реконструювання оцінки сигналу [1, 2].

Проблема з пониженням шуму вейвлет стиснення заключається в тому, що дискретне вейвлет-перетворення не інваріантне при трансляції, адже у вейвлет-представленні немає надмірності. Таким чином, форма відновленого сигналу після вейвлет стиснення і оберненого перетворення буде залежати від трансляції сигналу - ясно, що це не є задовільним. Щоб подолати проблему, було розроблено звільнення від шуму, що є інваріантним до трансляції [1]. Воно включає усереднення результату шумозаглушення вейвлет стиснення по всіх можливих трансляціях сигналу. Це дає досить прийнятні результати.

У той час як використання середньоквадратичних помилок у відновленні одновимірних сигналів може бути розумним, їх використання для порівняння зображень (двовимірних сигналів) піддається критиці [2, 3], адже часто зображення існують виключно задля зорової системи людини. Поки не розроблено метрики, яка безпосередньо враховує візуальне сприйняття людини. Тим не менш, одна вимірювана величина, здається, дуже важлива при людському сприйнятті образів – це фаза.

Класичний демонстрація важливості фази була розроблена Оппенгеймом і Лімом [4]: вони беруть перетворення Фур'є двох зображень, використовують фазову інформацію з одного та амплітудну інформацію з іншого, після чого проводять обернене перетворення Фур'є для отримання нового зображення. На такому зображенні можна побачити особливості зображення, з якого отримують фазову інформацію. Особливості другого зображення майже не помітні.

Для того, щоб зберегти дані про фазу в зображенні, ми повинні спочатку отримати фазову і амплітудну інформації в кожній точці зображення. Найбільш важливим параметром в процесі видалення шуму є поріг. Для вирішення вказаних проблем наведено схему автоматичного визначення порогу, алгоритм шумозаглушення, заснований на розкладанні сигналу з використанням складних вейвлетів. Цей алгоритм зберігає фазову інформацію, що є найважливішою частиною сигналу та продемонстровано різницю із алгоритмом, що не зберігає фазу.

Отже, враховано необхідне обмеження: підвищення якості зображення (наприклад, зменшення шуму) не повинне пошкоджувати фазову інформацію зображення.

Література. **1.** R. R. Coifman Time-in variant wavelet denoising / R. R. Coifman D. Donoho// Springer-Verlag. – 1995. – Vol. 103. – pp. 125–150.

2. D. L. Donoho Denoising by soft-thresholding / D. L. Donoho // Transactions on Information Theory. – 1995. – Vol 41. – pp. 613–627.

3. D. Wilson A new metric for grey-scale image comparison / D. Wilson, A. Baddeley, R. Owens // International Journal of Computer V ision. – 1997. – Vol. 24. – pp. 5–17.

4. A. V . Oppenheim The importance of phase in signals / A. V . Oppenheim J. S. Lim // In Proceedings of The IEEE. – 1981. – Vol. 69. – pp. 529–541.

Клунько Д.А.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Технология проектирования интеллектуальных систем обучения

Работа посвящена проектированию интеллектуальных систем обучения. Предлагаемая технология позволяет формализировать предметную область, сформировать когнитивную модель и оперировать в ней в целях обучения.

Дистанционное образование – это форма обучения, равноценная очной, вечерней, заочной и экстернатам, реализуемая, в основном, по технологиям дистанционного обучения. Технологии дистанционного обучения состоят из педагогических и информационных технологий дистанционного обучения. Информационные технологии дистанционного обучения – это технологии создания, передачи и сохранения учебных материалов, организации и сопровождения учебного процесса дистанционного обучения с помощью телекоммуникационной связи.

Характерными особенностями дистанционного образования являются: гибкость, модульность, экономическая эффективность. Предлагаемая в настоящей работе технология обучения, базируется на наборе инструментальных средств, выполняющих следующие функции: получение эталонной, формальной модели предметной области, создание базы знаний в рамках полученной модели, генерация тестовых вопросов, формирование когнитивной модели обучаемого подобной эталонной, диагностика пробелов в знаниях обучаемого.

Анализ формальных понятий. Для построения структурной логической модели предметной области используется метод анализа формальных понятий (АФП или Formal Concept Analysis). Этот метод был предложен R. Wille и успешно применяется в настоящее время в задачах анализа данных и машинного обучения. Он базируется на предположении о том, что единицей мышления являются понятия. В данном контексте понятие это категория для обозначения совокупности объектов обладающих определенными свойствами. Для формализации на множестве объектов V и признаков A определено отношение $I \subseteq V \times A$, такое, что pIa, где $p \in V$, $a \in A$ тогда и только тогда, когда a есть признак объекта p. Тогда тройка $K = (V, A, I)$ называется формальным контекстом. Формальный контекст, представим в виде бинарной матрицы, строки которой помечены именами объектов, а столбцы - значениями признаков. Формальным понятием (концептом) формального контекста называется пара (P, G), таких, что каждый объект $p_i \in P$ обладает всеми признаками $g_j \in G$, и таких, то каждый признак $g_j \in G$ присущ всем объектам $p_i \in P$. Множество объектов P составляет объем понятия, а множество всех свойств G, которыми они обладают, – его содержание.

Множество формальных понятий является частично упорядоченным по отношению включения и образует полную решетку $L(K)$, которая называется концептуальной решеткой формального контекста K [1].Решетка концептов представима в виде линейной диаграммы (диаграммы Хассе), в которой каждому узлу соответствует концепт из данного контекста. Дуальный изоморфизм на решетке концептов отображает фундаментальное свойство понятий – чем больше объем, тем меньше содержание(обратное отношение).

Свойством концептуальных диаграмм является отношение порядка на множестве формальных понятий, которое позволяет находить зависимости на признаках. Пусть $K = (V, A, I)$ есть формальный контекст и пусть $X \subseteq A$, $Y \subseteq A$. Тогда $X \to Y$, т.е. XY, или множество признаков Y от множества признаков X, если все объекты из $P \subseteq V$, имеющие признаки из X, обладают также всеми признаками из Y, т.е. $X' \subseteq Y'$ (или $Y \supseteq X''$). В этом случае формальное понятие, содержащее X, находится непосредственно ниже, содержащего признаки Y в концептуальной решетке. Такие зависимости образуют набор импликаций формального контекста. В подчиненных контекстах тоже могут возникать подобные свойства. В результате может быть построена иерархия контекстов, которую можно назвать .

Таким образом, в результате применения метода анализа формальных понятий мы получаем модель предметной области в виде иерархии формальных контекстов, каждому из которых соответствует список формальных понятий, список импликаций и концептуальная решетка

формального контекста. Суммарно эта информация хранится в виде исходной бинарной матрицы, а также в текстовом виде.

Использование результатов обработки контекста. Импликации, выявляемые из контекста, можно разбить на три группы.

1. Импликации, не несущие смысловой нагрузки. Это импликации, посылка которых ложна на множестве всех объектов контекста. Такие импликации исключаются из рассмотрения.

2. Правильные (истинные) импликации, которые адекватно отражают зависимости предметной области. Это зависимости, в которых объем посылки, совпадает с объемом заключения. Точность таких импликаций равна 100%.

3. Правдоподобные импликации (ассоциации), которые адекватно отражают зависимости между свойствами элементов контекста, но выполняются не на всей предметной области, а для отдельных объектов, т.е. объем посылки больше объема заключения.

Импликации групп 2 и 3 подаются поочередно на суд эксперта. Если какая либо из них неадекватно описывает предметную область, эксперту предлагается привести объект являющийся контр примером. Таким образом, импликации используются для пополнения модели предметной области и для диагностики знаний ученика. По импликациям строятся тестовые вопросно-ответные соответствия.

Список формальных понятий используются для построения тестов. И дает дополнительную возможность эксперту структурировать предметную область, поскольку некоторые формальные понятия имеют собственные обозначения. Концептуальная решетка используется при построении тестовых вопросов, определении стратегии их выдачи обучаемому. Линейная диаграмма во многих случаях хорошо иллюстрирует зависимости предметной области и используется в качестве наглядного пособия.

Атоматическая генерация тестов. Для формализации вопроса мы используем интеррогатив [2]. Тестовые вопросно-ответные соответствия, предполагаются состоящими из субъекта и предпосылки. Субъект определяет (явно или неявно) множество альтернативных ответов. Предпосылка вопроса дополняет субъект инструкцией, согласно которой из списка альтернатив необходимо построить правильный ответ. По числу возможных альтернатив вопросы делятся на 2 класса с небольшим (по крайней мере конечным) числом альтернативных ответов: -вопросы и с большим числом – -вопросы. Здесь идет речь не о количестве альтернативных ответов, а о способе их задания: в одном случае это явное перечисление, в другом – условие или предложение в котором на местах имен стоят переменные. Неправильные, альтернативные ответы называются дистракторами. Исходным материалом для построения вопросно-ответных соответствий являются множество формальных понятий и импликации, извлекаемые из контекста.

Тесты, составленные по импликациям, позволяют проверить, как ученик усвоил закономерности и правила данного фрагмента знаний. В вопросах, составленных по импликациям, используются правила удаления конъюнкции. Перечень свойств из которых можно выбрать задается. Концептуальная решетка дает возможность автоматически выбрать дистракторы для «ли»-вопросов.

Содержание доклада. В предлагаемом докладе будут представлены результаты исследований предложенной технологии проектирования интеллектуальных систем обучения.

Литература. **1.** Белнап Н., Стил Т. Логика вопросов и ответов . - М.: «Издательский Дом «Прогресс»», 1981. - 288 с.

2. Таран Т.А. Основы дискретной математики. - К.: «Видавництво «Просвіта»», 2003. – 288 с

Кондратенко Н.Р., Снігур О.О.
Вінницький національний технічний університет, Вінниця, Україна

Нечітка логічна система типу 2 для оцінювання перспективності артезіанської свердловини

В Україні з кожним роком зростає частка підземних вод у загальному балансі господарсько-питного водоспоживання в зв'язку зі скороченням обсягів придатних для споживання поверхневих вод. Для оцінки артезіанської свердловини на предмет перспективності її використання необхідно проаналізувати її за понад 300 параметрами. Дані за цими параметрами надходять поступово в процесі геологорозвідувальних робіт. Частину необхідних значень можливо отримати до початку робіт або ж на початковому їх етапі, інші - лише після проведення низки спеціальних заходів. В цілому тривалість оцінки запасів підземних вод може становити від 3 до 10 місяців [1].

Очевидно, що для остаточного прийняття позитивного рішення щодо перспектив розробки свердловини необхідно дочекатися повної інформації за всіма необхідними ознаками. Втім, можливість проміжної, хоча й наближеної, оцінки дала б змогу робити висновки про доцільність подальших досліджень протягом усієї тривалості геологорозвідувальних робіт, та уникнути зайвих витрат матеріальних та людських ресурсів там, де подальша розробка не видається перспективною.

Пропонується побудувати систему підтримки прийняття рішень, що може працювати з підмножиною параметрів свердловини, доступних на даний момент, у будь-якій проміжній точці розвідки.

На вхід моделі подається вектор ознак $V_1 \subset V_2 \subset V_3 = V$, доступних на певний момент перебігу гідрогеологічного дослідження. Серед них вибирається така комбінація параметрів, що найбільш адекватно відображає стан свердловини та водночас не містять надлишкових ознак [2].

Отриманий вектор вхідних ознак подається на вхід нечіткої логічної системи-класифікатора, що ставить у відповідність вхідному вектору значень інтервал із області визначення вихідного параметру. За вихідний параметр приймемо перспективність свердловини $D \in [0; 10]$. Початкову форму функцій належності вибрано довільно, спираючись на дані аналізу кількох свердловин Вінницької області. Базу правил побудови нечіткого логічного висновку також синтезовано на основі експериментальних даних. Проведено оптимізацію параметрів функцій належності отриманої нечіткої логічної системи типу 1 за допомогою генетичного алгоритму. Функції належності типу 1, отримані після оптимізації, перетворюються на інтервальні функції належності типу 2 з невизначеним середнім за методом [3]. Побудова інтервального нечіткого логічного висновку здійснюється за алгоритмом Карніка-Менделя [4]. Вихідна змінна приймає значення $[y_l; y_r] \in D(Y)$, y_r та y_l - права та ліва границі інтервалу. Його ширина свідчить про невизначеність рішення, прийнятого системою. Невеликий розкид значень говорить про те, що введених даних достатньо для прийняття рішення; значна ж ширина інтервалу свідчить про те, що інформації, накопиченої на поточний момент часу, недостатньо. В такому випадку рекомендовані подальші дослідження з метою отримання необхідних даних.

Література. **1.** Боревский Б.В. Оценка запасов подземных вод. / Б.В. Боревский, Н.И. Дробноход, Л.С. Язвин – 2-е изд., перераб. и доп. - К.: Выща шк. Головное изд-во, 1989. – 407 с.: ил. **2.** Кондратенко Н.Р. Нечітка модель оцінки запасів підземних вод / Н.Р. Кондратенко, О.О. Снігур // Інформаційні технології та комп'ютерна інженерія. Тези доповідей Міжнародної науково-практичної конференції. - Вінниця: ВНТУ.- 2014. **3.** Кондратенко Н.Р. Діагностика гіпотиреозу на основі нечіткої логіки з використанням інтервальних функцій належності / Н.Р. Кондратенко, Н.Б. Зелінська, С.М. Куземко // Наукові вісті національного технічного університету України „Київський політехнічний інститут". – 2003. №4. **4.** Karnik N. Type-2 Fuzzy Logic Systems / N. Karnik, J. Mendel, Q. Liang // IEEE Trans. On Fuzzy Systems, Vol. 7, No. 6. - 1999. - pp. 643-658.

Копичко С.М., Коровай О.В.
Національний технічний університет України "КПІ", факультет прикладної математики, Київ, Україна

Аналіз методів агрегації кардинальних експертних оцінок для СППР

Вступ. У системах підтримки прийняття рішень (СППР) в якості однієї з основних складових використовують результати експертного оцінювання. Дуже важливо, щоб інформація, яку надали експерти, була максимально достовірною. Для підвищення достовірності експертного оцінювання альтернатив часто використовують метод парних порівнянь та застосовують алгоритми агрегації оцінок[2].

Огляд основних методів агрегації оцінок. На теперішній час розроблено велику кількість методів агрегації експертних оцінок, отриманих при парних порівняннях. Більшість методів дають прийнятні результаті у разі високих ступенів внутрішньої узгодженості матриці парних порівнянь (МПП). Методи можуть використовувати зворотній зв'язок із експертом, але тоді з'являється велика кількість запитань експертові у випадках, коли узгодженість вхідної МПП не висока, або/та у випадках, коли у вхідній матриці зустрічаються протиріччя в ранжуванні деяких альтернатив.

Деякі з найбільш вживаних методів агрегації оцінок[2]:

1. Метод сум елементів рядків (Sums by Rows);
2. Середнє геометричне (Geometric average);
3. Метод ступеня (Powers method);
4. Нормалізація елементів стовпців (Columns normalizing);
5. Метод власного вектора (Eigen vector method);
6. Метод «Квадрат» («Square» method);
7. Комбінаторний метод (Combinatorial method).

Перші чотири методи відзначаються відносною простотою та застосовуються у випадку високої узгодженості МПП[3]. Метод «Квадрат» доцільно застосовувати при відносних відхиленнях не більших за 30%. Якщо розглядати усі ці методи з точки зору виконання ними свого основного призначення — підвищення достовірності, то виявиться, що комбінаторний метод, при відносних помилках експертних парних порівнянь менших 75%, спостерігається значна перевага ефективності цього методу у порівнянні з іншими методами[2].

Комбінаторний метод, використовуючи зворотній зв'язок з експертом, має на меті зведення запитань до експерта до мінімуму. Але для подальшого скорочення кількості запитань доцільно ввести деяку величину відхилення, яка б при найнеуспішніших комбінаціях все ж залишала б узгодженість всієї матриці у допустимих межах[1].

Висновки. Зважаючи на широке використання СППР у наші дні, важливо, щоб методи обробки інформації, які у них застосовуються були максимально ефективними. Зокрема задача удосконалення старих методів агрегації експертних оцінок та створення нових методів є актуальною та важливою. У роботі були розглянуті методи агрегації, проаналізовані результати їх роботи з точки зору ефективності цих методів та доцільності їх використання за різної кількості помилок у вихідних даних. Була виказана пропозиція внесення модифікацій у комбінаторний метод для підвищення його дієвості, а також виконана програмна реалізація комбінаторного методу та проведено тестування програмного продукту .

Література. **1.** Циганок В.В. Комбінаторний алгоритм парних порівнянь зі зворотним зв'язком з експертом / В.В. Циганок / Реєстрація, зберігання і обробка даних. — 2000. — Т. 2, № 2. — С. 92–102. **2.** Циганок В.В. Визначення ефективності методів агрегації експертних оцінок при використанні парних порівнянь/ В.В. Циганок / Реєстрація, зберігання і обробка даних. — 2009. — Т. 11, № 2. — С. 83–89. **3.** Тоценко В.Г. Методы и системы поддержки принятия решений. Алгоритмический ас-пект / В.Г. Тоценко; ИПРИ НАНУ. — К.: Наук. думка, 2002. — 382 с

Кравець І.О.
Чорноморський державний університет ім. П.Могили, Миколаїв, Україна

Використання нейронних мереж для ідентифікації та керування слабоформалізованими об'єктами

Багато об'єктів керування є специфічними і для них не можливо використання стандартних методів керування, які використовуються для автоматичного керування лінійних технічних об'єктів. Для таких об'єктів характерні таки особливості: математичний опис моделі відсутній, або частково відсутній, не стаціонарність моделі, необхідності швидкої адаптації до змін зовнішнього середовища у процесі експлуатації Для ідентифікації та керування такими об'єктами використовуються системи штучного інтелекту, насамперед нейронні мережі завдяки їх можливості навчання та пристосування к змінам зовнішнього середовища.

Розглянуті економічний приклад та приклад керування технічним об'єктом. У якості прикладу технічного об'єкта розглянуто задачу паркування вантажівки, навчальна вибірка якої наведена у [2]. Але на місті об'єкту курування може бути будь-який робототехнічний об'єкт. У якості прикладу економічного об'єкта розглянуто задачу керування фірмою, де керуючою змінною є обсяг виробництва [4].

Для нейрорегуляторів на базі чітких нейронних мереж зручно використовувати рекурентні нейронні мережі типу NARX з затримкою у часі між виходом мережі і входом та з часовою затримкою вхідних даних [3]. Так як модель об'єкта невідома, то у системі керування використовується адаптивне керування. Для нейрорегуляторів на базі чітких нейронних мереж використані регулятор з передбаченням та регулятор з моделлю процесу у вигляді авторегресії та ковзаючого середнього. На етапі ідентифікації розробляється модель керованого процесу у вигляді нейронної мережі, яка на етапі синтезу використовується для синтезу регулятора. Регулятор з передбаченням використовує модель нелінійного керованого процесу у вигляді нейронної мережі для того, щоб передбачити його майбутню поведінку. Крім того, регулятор обчислює сигнал управління, який оптимізує поведінку об'єкта на заданому інтервалі часу. Нейромережевий регулятор з моделлю процесу у вигляді авторегресії та ковзаючого середнього використовує в якості моделі керованого процесу модель нелінійної авторегресії із ковзким середнім

Для нейрорегуляторів на базі нечітких нейронних мереж використано нечітку нейронну мережу з логічним висновком Сугено [1].

В результаті дослідження нейромережевих регуляторів слабоформалізованими об'єктами було встановлено:

- для об'єктів, які можуть бути представлені рівнянням стану (диференціальними або різницевими з невідомими коефіцієнтами), можливе використання нейромережевих регуляторів як на базі чітких рекурентних нейронних мереж, так і нечітких нейронних мереж. Серед нейрорегуляторів на базі нечітких нейронних мереж найкращі результати показала нечітка нейронна мережа з логічним висновком Сугено;
- для об'єктів, які можуть бути представлені тільки базою правил, краще використання нечітких нейронних мереж. Серед нейрорегуляторів на базі нечітких нейронних мереж найкращі результати показала нечітка нейронна мережа з логічним висновком Сугено та попередня кластеризація вхідних данних, яка дозволяє значно зменшити базу правил.

Література. **1.** Зайченко Ю.П. Нечеткие модели и методи в интеллектуальних системах.- К., „Слово", 2008.- 333 с. **2.** Рутковская Д., Пилинський М., Рутковский Л.- Нейронные сети, генетические алгоритмы: Пер. с польск. И.Д. Рудинского.- М.:Горячая линя Телеком, 2006.-452 с.:ил. **3.** Хайкин Саймон.- Нейронные сети: полный курс, 2-е издание. : Пер. с англ. – М. : Издательский дом «Вильямс», 2006. – 1104 с. : ил. – Парал. тит. Англ. – ISBN 5-8459-0890-6. **4.** О классификациях стратегий компаний [Текст] / Ю. А Маленков // "Эмитент. Существенные факты, события, действия. Единое информационно-аналитическое обеспечение промышленности и предпринимательства Северо-Западного региона РФ". – N42(173) – 2006.

Ліщук К.І.
Національний технічний університет України "КПІ", Київ, Україна

Застосування метолу аналізу ієрархій для оцінки якості програмних продуктів

Одним з найбільш актуальних питань сучасної індустрії програмного забезпечення є оцінка якості. Достатньо часто при оцінці програмного забезпечення (ПЗ) як на стадії розробки, так і на стадії експлуатації, ми стикаємось з наступними питаннями: чи визначені вимоги до ПЗ в повному обсязі, чи всі можливі аспекти відображені, чи відповідає розроблене ПЗ всім визначеним вимогам та чи задовольняє потреби всіх зацікавлених сторін (розробників, архітекторів, замовників, кінцевих користувачів тощо).

Якість програмних продуктів (ПП) є комплексною характеристикою з точки зору різних учасників процесу. Модель якості визначається стандартом ISO/IEC 9126 та являє собою сукупність взаємопов'язаних характеристик, котрі являють собою базис для специфікації вимог до якості, котрі задовольняють потреби всіх зацікавлених осіб [1]. Стандарт ISO 9126 пропонує використовувати для опису моделі якості ПЗ багаторівневу модель. На верхньому рівні виділено 6 основних характеристик якості ПО. Для якісної оцінки кожної з характеристик, вона описується за допомогою кількох атрибутів. Для кожного атрибута визначається набір метрик, що дозволяють його оцінити. Для забезпечення якості в процесі проектування та розробки ПЗ рекомендується використовувати стандарт ISO/IEC 90003:2004. Він визначає набір технік і процедур, які рекомендується застосовувати для контролю й забезпечення якості ПП. Тобто, для того, щоб задовольнити потреби всіх сторін, необхідно враховувати всю сукупність характеристик ПП, які є важливими для всіх зацікавлених осіб.

Необхідність враховувати для оцінки якості ПП на різних етапах життєвого циклу велику кількість характеристик, критеріїв, показників та мір якості при оцінці тих чи інших властивостей ПП, їх неоднозначність та наявність конфліктуючих показників, наявність крім кількісних, так і якісних показників, призвело до появи великої кількості методів аналізу ПЗ. Сам процес оцінки якості ПП може розглядатись як наступна послідовність дій: спочатку визначення того, до чого ми прагнемо, далі – як цього найбільш ефективно досягти. Тобто більшість подібних задач можуть розглядатися як задачі багатокритеріального вибору. Для розв'язання задач багатокритеріального вибору існує велика кількість методів, котрі відрізняються обчислювальною складністю, необхідністю наявності експертних оцінок тощо.

Для аналізу та оцінки характеристик якості ПП по сукупності критеріїв пропонується використовувати метод аналізу ієрархій (МАІ) Т.Сааті [2]. Необхідно відмітити, що кількість можливих характеристик та критеріїв, за якими виконується оцінка, достатньо велика, тому використання класичного методу аналізу ієрархій (МАІ) є неможливим. Тому для розв'язання задачі пропонується використовувати одну з модифікації МАІ, запропоновані Павловим О.А., Ліщук К.І. [3,4].

Запропонований підхід для оцінки якості ПП дозволяє розв'язати поставлену задачу багатокритеріального вибору та відкриває нові можливості для побудови інформаційних систем оцінки якості ПП.

Література. **1.** Андон Ф.И. Основы инженерии качества программных систем [Текст]/Андон Ф.И., Коваль Г.И., Коротун Т.М., Лаврищева Е.М., Суслов В.Ю. - К.: Академпериодика, 2007. – 672 с. **2.** Саати Т. Принятие решений. Метод анализа иерархий: пер. с англ. [Текст] / Т. Саати; пер. с англ. Р.Г. Вачнадзе. – М.: Радио и связь, 1993. – 315 с. **3.** Павлов А.А. Многокритериальный выбор в задаче обработки данных матрицы парных сравнений [Текст] / А.А. Павлов, Е.И. Лищук, В.И. Кут // Вісник НТУУ "КПІ". Інформатика, управління та обчислювальна техніка. К.: "ВЕК+", 2007. – №46. – С. 84–88. **4.** Павлов О.А. Модифікований метод аналізу ієрархій (версія 1,2) [Текст] / О.А. Павлов, К.І. Ліщук, О.С.Штанькевич, Г.А.Іванова, О.П. Федотов // Вісник НТУУ "КПІ". – К.: "ВЕК+", 2009 – № 51.

Ляхевич С.В., Тимошенко Ю.О.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Архітектура системи моделювання поведінки коптера

Системи імітаційного моделювання широко розповсюджені та відомі як ефективні засоби первинної апробації моделей без необхідності їх повної імплементації на апробаційних стендах. Пропонується архітектура організації такого імітаційного моделювання взаємодії між компонентами безпілотних літальних апаратів та обробки вхідної інформації для процесу керування.

На поточний момент в Україні існує гостра необхідність у спеціалізованих засобах доставки на малі дистанції різних технічних засобів, таких як камери, сонари, тепловізори та інші вантажі. Більшість з цих завдань вимагає не просто доставку, але утримання вантажу в повітрі на заданій позиції з високим ступенем точності. Таким засобом можуть бути різні мультироторні літальні апарати (коптери), які дозволяють вирішити завдання спостереження та моніторингу різних локацій (лісові пожежі, дорожній трафік, огляд важкодоступних об'єктів)

Проте, не дивлячись на зростаючі потреби використання коптерів у різних сферах (природоохорона, відеоспостереження, поштові послуги) існуючі технологічні рішення або мають закриту платформу, що не дозволяє запрограмувати пристрій на рішення потрібної задачі, або вартість їх використання виходить за рамки рентабельності. Таким чином, виникає необхідність у вітчизняному виробництві мультироторних літальних апаратів, яке буде економічно рентабельним. Істотно знизити витрати на виробництво та апробацію можливо за рахунок використання системи моделювання керування [1] та поведінки коптерів в цілому при мультиагентній взаєиодії [2], розробка якої і є метою представленого дослідження. Дана модель взаємодій реалізована за використанням платформи Unity у зв'язку з великою кількістю наявних бібліотек, що описують різні фізичні процеси і моделі взаємодій.

На рисунку 1 представлена архітектура запропонованої системи моделювання:

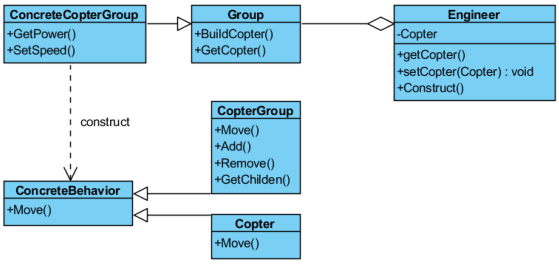

Рис. 1. Архітектура системи моделювання

Запропонована система моделювання дозволяє вирішувати наступні завдання: апробація моделі автоматичного польоту за наявності перешкод (та їх виявлення інфрачервоними дальномірами), моделей автоматичної посадки та пошуку шляху, що дає можливість проводити дослідження на імітаційному стенді, і істотно скоротити витрати на апробацію поведінкових моделей БПЛА.

Література. **1.** Sutton. R. "Integrated Architecture For Learning, Planning, and Reacting based on Approximating Dynamic Programming". in Machine Learning: Proccedings of Seventh International Workshop. pp. 216-224, San Francisco: Morgan Kaufmann, 1990. **2.** P. Hart, N. Nilsson and B. Raphael, "A Formal Basis for the Heuristic Determination of Minimum Cost Paths," in IEEE Transactions on Systems Science and Cybernetics, 1968, Vol. 4, No. 2, pp. 100-107 **3.** http://docs.unity3d.com/Manual/index.html

Олефір О.С., Савчук О.Б.
Національний технічний університет України "КПІ", факультет прикладної математики, Київ, Україна

Порівняння впливу різних факторів на очікувану тривалість життя населення України

Показник очікуваної тривалості життя населення є одним з основних критеріїв, що характеризують якість життя та рівень здоров'я нації. Статистика фіксує склад померлих за різними демографічними і соціальними ознаками, а це дає змогу проводити факторний аналіз показників смертності, оскільки в наш час гостро постає проблема визначення впливу того чи іншого фактору на нього. Особливо це стосується факторів смертності, яких можна уникнути або покращити (виліковні захворювання, алкогольна залежність і т.д.).

На основі статистичних даних про смертність на території України за період з 2000 по 2013 рр. [1] зроблено порівняльний аналіз та визначено міри впливу на очікувану тривалість життя наступних факторів:

1. Коефіцієнт народжуваності(народжені на 1000 осіб);
2. Сальдо міграції;
3. Показник безробіття населення у віці 15-70 років;
4. Показники інфляції;
5. Середній рівень доходу населення;
6. Кількість зареєстрованих випадків захворювання на новоутворення;
7. Кількість зареєстрованих випадків захворювання на хвороби системи кровообігу.

Для дослідження використовувався метод лінійної множинної регресії в стандартизованому масштабі, оскільки він часто використовується для прогнозування демографічних показників, що залежать від різних соціальних та економічних явищ, а також для визначення впливу кожного з цих факторів окремо та в сукупності на моделюємий показник. До того ж цей метод дає можливість ранжувати фактори за мірою їх впливу [2].

Були розраховані наступні коефіцієнти парної кореляції між показником смертності та факторами впливу:

$$r_{(yx1)} = -0.34; r_{(yx2)} = -0.32; r_{(yx3)} = -0.24; r_{(yx4)} = 0.54;$$
$$r_{(yx5)} = -0.59; r_{(yx6)} = -0.48; r_{(yx7)} = 0.79.$$

Використовуючи ці коефіцієнти отримано рівняння множинної регресії в стандартизованому вигляді:

$$(t_y) = -0.64t_{x_1} + 0.08t_{x_2} - 0.77t_{x_3} + 0.24t_{x_4} + 0.14t_{x_5} + 0.04t_{x_6} + 0.25t_{x_7}.$$

Порівнюючи розраховані коефіцієнти регресії, можна оцінювались міри впливу факторів на показник смертності.

Відповідно до значень парної кореляції та коефіцієнтів рівняння регресії в стандартизованому вигляді було зроблено висновок, що серед розглянутих факторів найбільш впливовими є захворювання на хвороби системи кровообігу, середній рівень доходу населення та інфляція.

У подальшій роботі планується розробити засоби для прогнозування смертності населення України в залежності від обраних факторів.

Література. **1.** Населення України. 2013. Демографічний щорічник. – К. : Держкомстат України, 2013. – 400 с. **2.** Костромин А. В. Конспект лекций по курсу «Эконометрика». Часть2. Казанский государственный финансово – экономический институт. – Казань 2004. – 47 с.

Павленко А.І.
Інститут кібернетики ім. В.М.Глушкова НАН України, Київ, Україна

Пошук найкоротшого шляху у динамічному мультиграфі методами мурашиних колоній з одним поколінням

Задача пошуку найкоротшого шляху є класичною задачею комбінаторної оптимізації, яка полягає у мінімізації суми ваг ребер, що складають маршрут між двома заданими вершинами на зваженому орієнтованому графі. Широке застосування задачі у різних галузях людської діяльності сприяє її активному вивченню і дослідженню нових алгоритмів та підходів. Розглядається підхід до побудови математичної моделі проблеми пошуку оптимального маршруту між вказаними містами (аеропортами) з додатковими умовами: обов'язковим включенням (виключенням) деяких міст до маршруту, задання часового інтервалу подорожі, її тривалість і вартість тощо.

Дано зважений орієнтований мультиграф, в якому між двома вершинами (містами або аеропортами) може існувати декілька ребер, що являють собою авіасполучення між аеропортами: $G = (V, A)$, де $V = \{v_1, v_2, \ldots, v_n\}$ - множина n вершин; A - множина дуг. Нехай $A_{ij} = \{a_{ij_1}, a_{ij_2}, \ldots, a_{ij_q}\} \subseteq A$ - множина дуг між вершинами i і j, q - кількість дуг. Для спрощення покладемо $\exists a_k = (i, j) \in A$. За умовами задачі необхідно знайти оптимальний шлях з початкової вершини $s \in V$ до цільової $d \in V$ в заданий проміжок часу $T = [t_{0\min}, t_{0\max}] \subseteq T$. Нехай $t_0 \in T$ - час відправлення з вершини s.

Послідовність вершин $P(s, d, t_0) = \{a_1, a_2, \ldots, a_w\}$ вважається шляхом з s в d, який починається в час t_0, якщо початкова вершина кожної дуги є кінцевою вершиної попередньої дуги $\forall a_k = (i_k, \cdot) \in A : a_k \notin \{(s, \cdot)\} \exists a_{k-1}(\cdot, i_k)$.

Якщо рейс $a_k \in A$ виконується в час t_{k-1}, тоді прибуття в аеропорт j відбувається в $t_k = t_{k-1} + \lambda_{a_k}(t)$, де $\lambda_{a_k}(t)$ - невід'ємна залежна від часу функція, яка визначає тривалість перельоту рейсом a_k з i в час t_{k-1}. $t \in T$ - змінна часу, яка визначає час виходу з вершини i. Час подорожі - різниця часу прибуття і часу початку подорожі.

$$t(P) = \sum_{k=1\ldots w} (t(a_k) + g(a_{k-1})) - t_0,$$

де $t(a_k)$ - час транзиту вздовж дуги a_k; $t(P)$ - час проходження повного шляху; $g(a_{k-1})$ - час очікування у початковій вершині дуги a_{k-1}.

$c_{a_k}(t)$ - невід'ємна залежна віж часу функція, яка являє собою загальну вартість перельоту a_k. Вартість шляху P визначається як $c(P) = \sum_{k=1\ldots w} c_{a_k}(t)$, де w - кількість дуг у шляху.

Метою є знаходження оптимального шляху (або доступних шляхів з врахуванням додатковіих умов) $P^*(s, d, [t_{0_min}, t_{0_max}])$. $c(P^*) \leqslant (P) \forall P(s, d, [t_{0_min}, t_{0_max}])$. Додатково можуть накладатись умови: задаватись обов'язкові для відвідування вершини $V_{mandatory} \subseteq P$, заборонені вершини $\forall v \in V_{prohibited} : v \notin P$, максимальна вартість повного шляху $(P) \leqslant c_{\max}$, максимальна кількість вершин у маршруті $\left|P\right| \leqslant w_{\max}$, тривалість маршруту у часі $t_{\min} \leqslant t(P) \leqslant t_{\max}$. Варто зазначити, що змінювати міста (вершини) кожного дня необов'язково.

При розв'язанні задачі важливим критерієм є мінімізація часу розврахунку. Пошук оптимального шляху виконується алгоритмом мурашиних колоній [1] на реальних даних перельотів авіаліній Європи за один тиждень. Реалізовано алгоритм мовою C#, який працює в рамках одного покоління з наперед заданим розміром колонії. Розмір групи мурах, які можуть одночасно бути на полюванні, може бути заданим параметром або визначатись автоматично з врахуванням завантаження машини (кількості вільних потоків системи за замовчуванням). Коли в пулі потоків з'являється вільний потік, він обирається для створення нової мурахи і моделювання її роботи. Коли мураха знаходить маршрут або він перестає бути допустимим, мураха зникає, а її потік повертається в загальний пул. Така система дозволяє оптимізувати час та ресурси за рахунок можливості використовувати існуючі потоки замість їх постійного створення та видалення. Схема дозволяє зменшити час простою процесора. Оновлення феромонів виконується тільки у випадку успішної побудови маршруту. Для проведення експерименту

були використані такі значення параметрів $\alpha = 1.5$, $\beta = 0.5$, $w_{\max} = 10$, загальна кількість мурах в обох схемах складає 1000, граничні значення феромонів min-max схеми оновлення феромонів - 0.1 і 0.7, коефіцієнт випаровування $\rho = 0.9$. Граф містить 15000 дуг (авіарейсів). Операція випаровування виконується після завершення проходу кожної 20ї мурахи. Розмір групи мурах на полюванні (або розмір популяції) для класичної схеми становить 20 мурах. На Рис. 1 наведено порівняння тривалості роботи алгоритму з одним поколінням і кількома (класична схема) для побудови маршруту з аеропорту "Бориспіль". На Рис.2 наведено порівня-

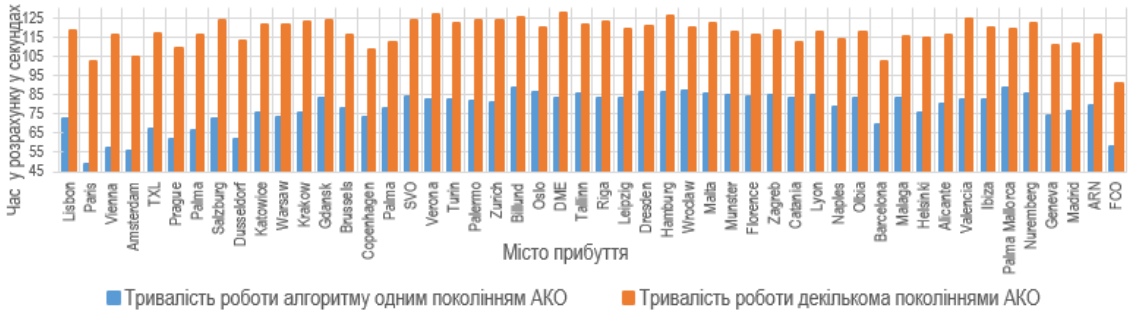

Рис. 1. Порівняння роботи алгоритму АКО з одним і декількома поколіннями за тривалістю роботи

ння вартості оптимальних маршрутів з аеропорту "Бориспіль" до 50 міст Європи мурашиним алгоритмом з одним і кількома поколіннями (класична схема). Були проведені експерименти

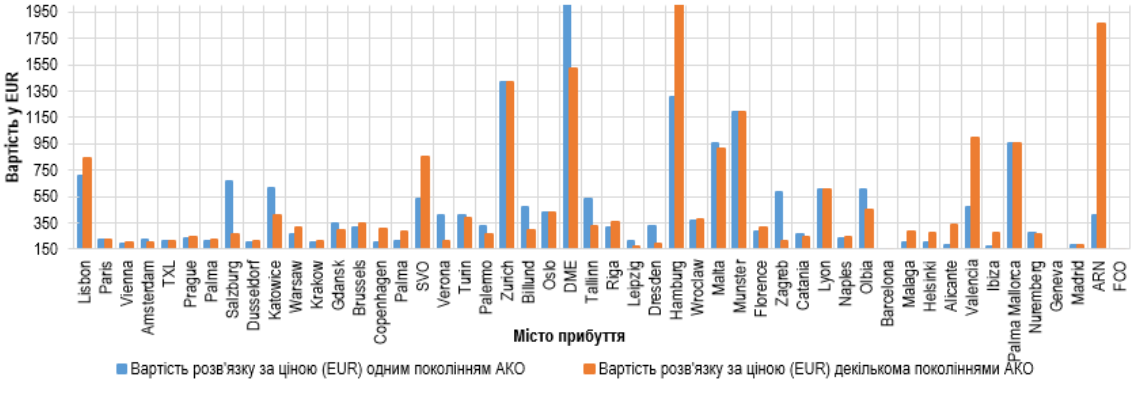

Рис. 2. Порівняння роботи алгоритму АКО з одним і декількома поколіннями за ціною фінального маршруту

на більшій вибірці (муршрути з Борисполя до 113 міст Європи). Для 10ти з них алгоритм з одним поколінням не зміг знайти допустимих розв'язків, алгоритм за класичною схемою не знайшов розв'язки для трьох міст. Разом з тим, для знайдених маршрутів запропонований алгоритм дає кращі результати у середньому на 16% по всій вибірці.

Висновки. Досліджено застосування алгоритму мурашиних колоній для пошуку шляху у динамічному мультиграфі і описаний підхід для зменшення часу пошуку. Наведене відносне порівняння якості алгоритму між класичною схемою і запропонованою. Додаткового дослідження потребує вивчення і оптимізація параметрів алгоритму.

Література. **1.** Dorigo B., Stutzle T., "Ant colony optimization" in MIT Press, Cambridge Massachusetts, 2004.

Папуч Т.Ю.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Аналіз кредитоспроможності юридичних осіб за методами інтелектуального аналізу даних

Вступ. Актуальність задачі. Кредитний ризик протягом довгого часу є важливою та популярною темою в фінансовій області, що пояснюється значним попитом у сфері кредитування. Кредитний ризик - втрати, які може понести фінансова установа у результаті несплати позичальником кредиту та відсотків по ньому. Існуючі моделі та методи недостатньо ефективно працюють у країнах з перехідною економікою. Саме тому так гостро постає проблема вдосконалення і адаптації цих моделей до української економіки.

Невід'ємною частиною аналізу кредитоспроможності позичальника є оцінка ризику його банкрутства. У роботі розглядаються та порівнюються існуючі моделі оцінки банкрутства, а саме: чіткі, класичні, статистичні моделі аналізу та моделі нечіткого логічного висновку. Статистичні моделі включають модель Альтмана, Ліса, Давидової-Бєлікова, Спрингейта. Також розглядаються нечітко-множинний підхід та модель нечіткого логічного висновку Мамдані. Основні методи для вимірювання кредитного ризику та кредитоспроможності позичальників, що використовуються у роботі, є кредитний скоринг, моделі нечіткого логічного висновку та штучні нейронні мережі. Інтеграція методів аналізу банкрутства та моделей аналізу кредитоспроможності дозволить покращити точність кредитного прогнозу та зменшити відсоток помилок 1-го і 2-го роду.

Теоретичні відомості. Фінансовим установам необхідно прогнозувати величину ризику банкрутства потенціального позичальника перед наданням йому кредиту. Статистичні методи поділяються на методи логістичної регресії та дискримінантного аналізу. Модель Альтмана використовує класичний метод багатовимірного дискримінантного аналізу і має наступний вигляд:

$$Z = 1.2 * K_1 + 1.4 * K_2 + 3.3 * K_3 + 0.6 * K_4 + 1.0 * K_5, \quad (1)$$

де K_1 = власний оборотний капітал/сума активів;
K_2 = нерозподілений прибуток/сума активів;
K_3 = прибуток до сплати відсотків/сума активів;
K_4 = ринкова вартість власного капіталу/вартість позикового капіталу;
K_5 = об'єм продажів/сума активів.
В результаті після підрахунку Z-показника можна зробити наступний висновок:
якщо Z$<$1.81 - дуже висока ймовірність банкрутства;
якщо $1.81 \leqslant Z \leqslant 2.7$ - висока ймовірнсть банкрутства;
якщо $2.7 \leqslant Z \leqslant 2.99$ - можливе банкрутство;
якщо $Z \geqslant 3.0$ - ймовірність банкрутства дуже мала.

Підхід Альтмана на основі багатовимірного дискримінантного аналізу далі був розвинений іншими дослідниками. До числа відомих моделей прогнозування банкрутства відносяться:

- Модель Ліса має наступний вигляд:

$$Z = 0.063 * K_1 + 0.092 * K_2 + 0.057 * K_3 + 0.001 * K_4, \quad (2)$$

де K_1 = оборотний капітал/сума активів;
K_2 = прибуток від реалізації/сума активів;
K_3 = нерозподілений прибуток/сума активів;
K_4 = ринкова вартість власного капіталу/позиковий капітал;
При Z$<$0.037 - висока ймовірність банкрутства.

- Модель Давидової-Бєлікова має наступний вигляд:

$$R = 8.38 * K_1 + K_2 + 0.054 * K_3 + 0.63 * K_4, \quad (3)$$

де K_1 = оборотний капітал/сума активів;
K_2 = чистий прибуток/сума власного капіталу;

K_3 = об'єм продажів/сума активів;
K_4 = чистий прибуток/собівартість;
При R<0 - ймовірність банкрутства максимальна (90-100 %); 0<R<0.18 - ймовірність банкрутства висока (60-80 %); 0.18<R<0.32 - ймовірність банкрутства середня (35-50 %); 0.32<R<0.42 - ймовірність банкрутства низька (15-20 %); R>0.42 - ймовірність банкрутства мінімальна.

Таким чином, ці моделі використовують фінансові коефіцієнти та показники для оцінки ризику банкрутства позичальників.

У роботі також використовується матричний метод прогнозування банкрутства Нєдосєкіна, заснований на апараті нечітких множин. Цей метод складається з наступних етапів:

Етап 1. Визначення лінгвістичних змінних і нечітких множин.
Етап 2. Побудова набору окремих показників, що впливають на оцінку ризику банкрутства.
Етап 3. Визначення рівня значущості кожного показника.
Етап 4. Класифікація ступеня ризику.
Етап 5. Класифікація значень показників.
Етап 6. Оцінка рівня показників.
Етап 7. Класифікація рівня показників.
Етап 8. Оцінка ступеня ризику.
Етап 9. Лінгвістичне розпізнавання.

Аналіз ризику банкрутства за допомогою моделі нечіткого логічного висновку Мамдані складається з наступних етапів.

Етап 1. Визначення лінгвістичних змінних і нечітких підмножин: "'Стан підприємства", "'Ризик банкрутства", "'Рівень показника $_i$".
Етап 2. Побудова набору окремих показників X_i.
Етап 3. Формування бази правил системи нечіткого висновку.
Етап 4. Фазифікація вхідних параметрів.
Етап 5. Логічний висновок.
Етап 6. Композиція виходів.
Етап 7. Дефазифікація за центроїдним методом.

Модель нечіткого логічного висновку Мамдані показує високу ефективність і якість навіть при нечіткій інформації про позичальників.

Розв'язання задачі. У ході роботи побудована ІСППР, яка після визначення ймовірності банкрутства позичальника, аналізує його кредитоспроможність. Першим методом є кредитний скоринг. Кредитний скоринг може бути визначений як систематичний метод для оцінки кредитного ризику, що виконує послідовний аналіз факторів, які були визнаними такими, що впливають на рівень ризику. Процес кредитного скорингу починається саме з визначення цих факторів і їх врахування у модель. Такі фактори включають коефіцієнти аналізу, кредитні рейтинги, Z-оцінку Альтмана та кредитну оцінку.

Другим інструментом аналізу є модель нечіткого логічного висновку Мамдані, що використовує наступні змінні, які описують кредитоспроможність позичальника: кредитна історія, фінансово-економічний стан, коефіцієнт незалежності, коефіцієнт миттєвої ліквідності, період обертання запасів, ймовірність несвоєчасного погашення кредиту, ризик збитків, стан документації, банкрут (або ні), забезпечення, платоспроможність. У подальшому планується застосування штучних нейронних мереж при аналізі кредитного ризику.

Література. **1.** Зайченко Ю.П. Комплексний аналіз ризику банкрутства корпорацій в умовах невизначеності / Ю.П. Зайченко // Стаття. - International Journal 'Information Technologies & Knowledge', vol.6, № 2, 2012, p. 103-125. **2.** A. Samreen, F. Batul Zaidi, A. Sarwar. Design and Development of Credit Scoring Model for the Commercial Banks in Pakistan: Forecasting Creditworthiness of Corporate Borrowers. - International Journal of Business and Commerce, vol. 2, № 5, 2013, p. 1-26.

Письменний І.О.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Мультиагентні системи в моделюванні підприємств галузі

В даній роботі розглядається задача моделювання галузі екомоміки за допомогою мультиагентних систем. Завдяки правильній побудові моделей користувач зможе оцінити актуальність та прибутковість свого бізнесу, визначитися зі стратегіями його поведінки для різних кон'юктур.

Перед створенням нового бізнесу необхідно максимально правильно оцінити його актуальність та прибутковість, зпрогнозувати його роботу в різних кон'юктурах. Розв'язання задачі прогнозування процесів життєвого циклу підприємства на тому чи іншому етапі розвитку, у більшості випадків пов'язане з моделюванням залежностей результуючих характеристик від різноманітних вхідних факторів. [1,2] Вони можуть мати різноманітну природу та визначатися як кон'юктурою ринку, так і внутрішніми параметрами підприємства, в той же час динаміка зміни цих факторів також має велике значення, що значно ускладнює розрахунок потенціної ефективності компанії традиційними методами.

Для вивчення складних систем застосовується комп'ютерне моделювання. [3] Одним з найрозвинутіших програмних средовищ розробки мультиагентних систем є Java Agent Development Framework (JADE). Даний фреймворк значно прискорює розробку мультиагентної системи, надаючи користувачеві готовий контейнер та можливість реалізації агентів з життєвим циклом аідповідно до стандартів Foundation for Intelligent Physical Agents (FIPA). [4]

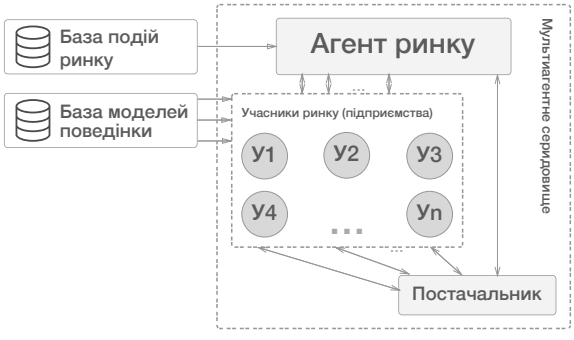

Рис. 1. Схема розроблюваної мультиагентної системи

В даній роботі галузь буде промоделювано для різних умов зовнішнього середовища (ріст/стагнація економіки, валютні коливання, спричинена цими факторами зміна цін на комплектуючі та робочу силу). Показником правильності побудови та регулювання системи буде наявність короткострокових економічних циклів Кітчина, котрі спостерігаються в реальному світі.

Розроблюваний сервіс являє собою мультиагентну систему, побудовану на платформі JADE (рис. 1), що дозволяє якісно та з мінімальною затратою ресурсів виконати поставлену задачу.

Подальшим напрямком дослідженння буде моделювання підприємств суміжних галузей, постачальників ресурсів, що дозволить спостерігати довгострокові економічні цикли та точніше визначати довгострокову стратегію бізнесу.

Література. **1.** Б. В. Мисник, Математичне моделювання та формалізація основних потоків виробничого підприємства. **2.** Б. В. Мисник, Особливості моделювання процесів функціонування виробничих підприємств на основі концепції «штучного життя» **3.** Л.А. Гуревич, А.Н. Вахитов Мультиагентные системы // Введение в Computer Science. – 2005. –с.116-139. **4.** Fabio Luigi Bellifemine, Giovanni Caire,Dominic Greenwood, Developing Multi-Agent Systems with JADE

Поворознюк А.І.
Національний технічний університет "Харківський політехнічний інститут", Харків, Україна

Мінімізація ризиків лікарських помилок у системах підтримки прийняття рішень при проведенні діагностично-лікувальних заходів

Формалізована задача комплексної оцінки етапів діагностично-лікувального процесу з метою мінімізації ризиків лікарських помилок. Розроблені методи кластеризації діагнозів у просторі фармакологічних дій і корекції порогів у діагностичному вирішальному правилі. Виконана програмна реалізація системи й тестування на реальних медичних даних.

Ключові слова: комп'ютерна система, діагностика, медикаментозне лікування, фармакологічна дія, лікарська помилка.

Постановка проблеми. Термін «лікарська помилка» визначає неправильну діагностику хвороби або неправильні лікарські дії (призначення терапевтичних процедур, ліків, хірургічне втручання й т.п.), які обумовлені сумлінною помилкою лікаря, якщо при цьому виключається недбалість і несумлінність при виконанні своїх обов'язків. Для більшості патологій у різних областях медицини лікарські дії виконуються медикаментозним шляхом. Причиною помилки діагностики є недостатній обсяг діагностичних даних (застосування застарілого обладнання або недостатня кваліфікація лікаря, який призначав список обстежень), або їх невірна інтерпретація (особливо при суб'єктивному аналізі якісних показників). Причиною неправильних лікарських дій при правильному діагнозі є недостатнє врахування індивідуальних особливостей пацієнта (алергічні реакції на певні препарати, список хвороб, якими вже хворів пацієнт, які ліки приймав за певний проміжок часу і т.д.). Прийняття неправильного рішення як на етапі діагностики, так і на етапі лікування може мати катастрофічні наслідки для здоров'я пацієнта.

У формалізованому виді задачею діагностики є класифікація стану i-го пацієнта D_i при аналізі вектора діагностичних ознак X_i [1]. При медикаментозному лікуванні кожному діагнозу D_i ставиться у відповідність множина необхідних фармакологічних дій (ФД) f_{D_i}, на підставі якої формується комплекс лікарських препаратів (ЛП), які забезпечують f_{D_i}, з врахуванням нестерпності i-го пацієнта до окремих препаратів, несумісності препаратів, багатокритеріального порівняння препаратів-аналогів [2]. ФД – це вплив активних компонентів ЛП на окремі органи людини й організм у цілому. На теперішній час є широкий спектр комп'ютерних діагностичних систем у різних предметних областях медицини [1], інформатизація етапу формування комплексу ЛП обмежується медичними довідниками фармацевта, у тому числі у вигляді інформаційно-пошукових систем [3], які представляють лікареві структурований список (класи, підкласи і т.д.) ЛП і текстовий опис їх властивостей.

При цьому задачі діагностики й лікувальних заходів розглядаються незалежно друг від друга, при діагностиці мінімізується ризик неправильної постановки діагнозу без врахування етапу лікувальних заходів, тому актуальною є задача мінімізації ризику лікарської помилки при комплексній оцінці всіх етапів діагностично-лікувальних заходів (ДЛЗ).

Метою роботи є розробка інформаційних технологій комплексної оцінки етапів ДЛЗ з метою підвищення їх ефективності й мінімізації ризику лікарських помилок.

Формалізація задачі комплексної оцінки етапів ДЛЗ. В [1] формалізовані наступні етапи перетворення інформації в системах підтримки прийняття рішень в медицині: структурна ідентифікація біосигналів $F1 : x(t) \to X$ і медичних зображень $F2 : x(j,k) \to X$; формалізація опису різнорідних діагностичних ознак і синтез ієрархічних структур діагностуємих станів $F3 : D \to S_D$ і діагностичних ознак $F4 : D \to S_z$; синтез діагностичних вирішальних правил (ВП) при взаємодії S_D і S_z $F5 : X_i \to D_i$; формування множини необхідних ФД $F6 : D_i \to f_{D_i}$ та комплексу ЛП $F7 : f_{D_i} \to Y_i$.

Для мінімізації ризиків лікарських помилок розглянемо більш докладно перетворення $F3$ і $F5$. Синтез S_D – бінарного дерева рішень виконується процедурою ієрархічної кластеризації множини діагностуємих станів $\{D_i\}_n$ за критерієм мінімуму помилки кластеризації в просторі ознак X (перетворення $F3$). Як наслідок, у ході такого процесу утворюється бінарне дерево,

коренем якого є повна множина діагнозів $\{D_i\}_n$ у заданій предметній області, у гілках розташовуються кластери діагнозів, близько розташованих друг до друга, а листами – окремі діагнози. Процес діагностики – рух по дереву рішень, у кожній k-й вершині якого виконується діагностика станів D_q і D_l, шляхом обчислення ВП і прийняття рішення на користь D_q або D_l. На кожному i-му етапі ВП аналізується чергова ознака x_i й обчислюється відношення правдоподібності $\Theta = \prod_i P(x_{ik}/D_q)/P(x_{ik}/D_l)$, яка порівняється з порогами $\Theta > A$, $\Theta < B$. При виконанні однієї з нерівностей ухвалюється рішення про D_q або D_l відповідно й виконується перехід на більш низький рівень ієрархії S_D з метою уточнення діагнозу. При невиконанні обох нерівностей додається наступна $i+1$ ознака й процедура повторюється.

Границі прийняття рішень у ВП A и B пов'язані з помилками класифікаціями α і β наступними співвідношеннями: $A = (1-\beta)/\alpha$, $B = (1-\alpha)/\beta$, де α і β – помилка першого й другого роду відповідно й визначаються розташуванням еліпсоїдів розсіювання об'єктів навчальної вибірки в просторі ознак без врахування їх впливу на етап вибору необхідних ФД і наступного призначення ЛП.

Для мінімізації ризику неправильних медичних заходів, які виникають при помилковій діагностиці, шукається залежність між помилкою при діагностиці (D_q замість D_l), і її наслідків при призначенні комплексу ЛП (Y_q замість Y_l). Так як $Y_q \to f_{D_q}$, а $Y_l \to f_{D_l}$, то ризик визначається розбіжністю компонентів множин f_{D_q} і f_{D_l}, і для його мінімізації виконується перехід від традиційного простору ознак X у простір фармакологічних дій F, компонентами якого $f_m \in F$ є бінарні змінні (0 – відсутня фармакологічна дія, 1 – присутня), а кожний стан D_i представляється i-ю вершиною гіперкуба.

Застосування ієрархічної кластеризації за критерієм мінімуму сумарному зв'язку (мінімальний розріз R) у просторі F забезпечує мінімум ризику прийняття неправильного рішення на етапі формування комплексу ЛП при синтезі дерева рішення S_D на етапі діагностики. Крім того, у роботі пропонується метод корекції порогів A і B у ВП, враховуючи помилки, які виникають на етапі призначення комплексу ЛП. Отримані залежності між α, β і мінімальним розрізом R : $\alpha = 0,5(1-R_{ql}), \beta = 0,5(1-R_{lq})$. Обчисленні в такий спосіб α і β задають пороги $A = (1-\beta)/\alpha$, $B = (1-\alpha)/\beta$ у ВП, що забезпечує врахування ризиків при призначенні комплексу ЛП у діагностичному ВП.

Виконана програмна реалізація системи і її тестування на реальних медичних даних з використанням навчальної вибірки з 400 пацієнтів. У якості основної платформи обґрунтований вибір Java, проектування архітектури виконувалося з використанням шаблонів проектування [4]. В архітектурі системи виділено три основні модулі: модуль взаємодії з користувачем, базу даних, що включає в себе базу знань і модуль побудови знань.

Висновки. Розроблені методи побудови нового класу систем підтримки прийняття рішення при проведенні ДЛЗ на основі формалізації етапів проведення ЛДЗ при їхній комплексній оцінці, що дозволяє мінімізувати ризики лікарських помилок, підвищити вірогідність і обґрунтованість рішень. Архітектура програмного забезпечення системи дозволяє легко адаптуватися до різних предметних областей медицини.

Література. **1.** Поворознюк А.И. Системы поддержки принятия решений в медицинской диагностике. Синтез структурированных моделей и решающих правил – Saarbrücken Germany: LAP LAMBERT Academic Publishing GmbH & Co. KG, 2011. – 314 с. **2.** Поворознюк О.А. Біотехнічна система призначення лікарських препаратів в дерматології. Автореферат дис. на здоб. вч. ступ. к. т. н.: спец. 05.11.17 "Біологічні та медичні прилади і системи" – Харків, 2010. – 21 с. **3.** Компендиум 2007 – лекарственные препараты /Под ред. В.Н. Коваленко, А.П. Викторова. [Електронный ресурс] http://www.compendium.com.ua. **4.** Бек К. Шаблоны реализации корпоративных приложений.: Пер. с англ.– М.: ООО "И.Д. Вильямс", 2008. – 176 с.

Попова М.В., Безносик О.Ю.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Застосування методів кластерізації для слабкоструктурованих задач

Задача, що розв'язується. В роботі розглядається задача аналізу рейтингів викладацького складу НТУУ "КПІ", пошук прихованих взаємозв'язків та видача рекомендацій щодо покращення показників підрозділів ВНЗ. Поставлена задача відноситься до слабкоструктурованих. Дані, що використовуються, є актуальними, вірними та такими, що походять з достовірних джерел.

Використовувані дані. Для аналізу доступні такі дані:

1. Рейтинг науково-педагогічних працівників, що включає в себе:
   - загальний рейтинг;
   - рейтинг з навчально-методичної роботи;
   - рейтинг з науково-інноваційної діяльності;
   - рейтинг за посадами та віком;
2. Дані з ректорського контролю.

Існуючі підходи до задачі. На даний момент існує аналіз кожної складової рейтингування науково-педагогічних працівників, що відбувається кожного року, і результати якого публікуються у навчальному виданні [1].

Результати ректорського контролю публікуються в періодичному виданні "Київський політехнік". Схеми обчислення і загальні висновки подаються у інформаційному виданні з матеріалами кожного туру комплексного моніторингу якості підготовки фахівців в НТУУ "КПІ" [2].

Наразі не встановлено взаємозв'язок та не простежені залежності між цими наборами даних.

Актуальність задачі. Зрозуміло, що рейтинг науково-педагогічних працівників з навчально-методичної роботи має бути пов'язаний з успішністю студентів; відповідно, студенти кафедр-лідерів з навчально-методичної роботи мають демонструвати вищі результати з ректорського контролю, ніж студенти менш успішних підрозділів. Однією з задач роботи є перевірка кореляції вищезазначених даних. Іншою задачею є побудова кластерів на основі аналізу вказаних вище даних для виявлення прихованих зв'язків, а також отримання "портрету" типового доцента, професора, викладача, науково-педагогічного працівника певного віку, тощо з точки зору методів кластерного аналізу.

Методи кластерізації. Можна виділити ряд груп методів (деякі методи можна віднести відразу до декількох груп, і тому пропонується розглядати дану типізацію як деяке наближення до реальної класифікації методів кластеризації):

1. Ймовірнісний підхід. Передбачається, що кожен об'єкт відноситься до одного з k класів.
2. Підходи на основі систем штучного інтелекту.
3. Логічний підхід. Побудова дендрограмми здійснюється за допомогою дерева рішень.
4. Теоретико-графовий підхід.
5. Ієрархічний підхід.
6. Інші методи, що не ввійшли до попередніх груп.

Результат роботи. Результатом роботи є видача рекомендацій щодо покращення показників підрозділів ВНЗ на основі розрахованих кореляцій і сформованих кластерів.

Література. **1.** Аналіз рейтингів науково-педагогічних працівників у 2013/2014 навчальному році / Уклад. В. П. Головенкін. – К.: НТУУ "КПІ", 2014. – 36с. **2.** Матеріали XIX туру комплексного моніторингу якості підготовки фахівців в НТУУ "КПІ" (V курс, осінь'2014)/Укл. М. М. Перестюк.- К.: ВПІ ВПК "Політехніка", 2015. - 28с.

Розпутний М.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Засоби здійснення логічних висновків над онтологіями

Онтологія - формалізоване представлення певної області знань. Концептуальна схема данної предметної області складається з визначених об'єктів (елементів) та взаємовідношенн між ними. Онтології використовуються в розробці програмних проуктів, моделюванні бізнес-процесівб симантичних павутин, штучного інтелекту. Обробка даних, вилучення інформації з неструктурованих, або погано структурованих даних, прийняття рішень на основі цієї інформації є метою використання онтологій.

Проблема прийняття рішення в умовах обмежених ресурсів є актуальною, тому зараз розробка методів основаних на онтологіях є вкрай корисною.

Існують велика кількість моделей, методів, методологій програмних засобів, які спрямовані на розв'язання заач розробки онтологій, та прийняття висновків, рішенн на їх основі. Активний дослідження в цьому напрямку відбвуваються разом з розвитком обчислювальної техніки. оскільки вона дає нові можливісті дослідження великих обємів даних, та їхньої обробки. Для автоматизації аналізу та синтезу даних створювалися різноманітні моделі процесів обробки, а також відповідні алгоритми та структури представлення даних [1].

Описові логіки, що дозволяють описувати поняття предметної області в недвозначному, формалізованому вигляді : Attributive language (AL), Extension language (EL) Frame based description language (FL). Для їх використання розроблено мови їх опису (OWL-DL, OWL-Lite, OWL-Full).

Найбільш широковідомі методи : метод оцінки адекватності розвитку онтології, метод прямого нечіткого висновку, пакетно-орінтованої описової логіки, семантико-синтаксичний алгоритм стиску та інші.

В даний час існують велика кількість програмних засобів для створення онтологій для побудови баз знать. Найбільш відомі : CYC, Protege, KAON2, OntoEdit, KADS22.

За допомогою онтологій вирішується широкий спектр задач, в яких необхідно оброблювати дані, що враховують їх семантику. Тому зростає інтерес до мови онтології. Особливий інтерес є в алгоритмах міркування над кількома розподіленими автономними модулями онтології. Наприклад в типових додатках інтеграції данних, де користувач відправляє запити, виражені з використанням термінів онтологій, що мають отримати відповіді від різних джерел даних. В таких умовах запит користувача має бути переведений із користувацької онтології в запит, виражений у термінах вихідних даних онтологій. Використання онтологій для цих цілей може істотно спростити а також відкрити нові можливості в розробці додатків, що вирішують завдання автоматизованої обробки і доступу до даних.

Одне з найбільш важливих завдань, яку можна вирішити, використовуючи онтології - це семантичний пошук. В даний час проблема пошуку інформації у великих масивах є надважливою. Ця проблема ускладнюється ще й тим, що існуючі пошукові механізми здійснюють пошук інформації без урахування семантики елементів, що входять до запиту, а також контексту, в якому вони використовуються [2].

Для прийняття рішень необідна розробка якісно нового методу до обробки бази знань і прийняття рішення на основі семантики даних що оброблюються є основною метою дослідження. Основні шляхи реалізації даних ідей : створення нового підходу до побудови і обробки онтологій, модернізація існуючих алгоритмів роботи з ними. В основу дослідження покладено досвід відчизняних і світових розобок в свері інтелектуальних систем прийняття рішень.

Література. **1.** Джозеф Джарратано. Экспертные системы принципы разработки и программирование / Джозеф Джарратано, Гари Райли — М. : Вильямс, 2006. — 1152 с. **2.** J. Bao, D. Caragea, and V. Honavar. "Modular ontologies - a formal investigation of semantics and expressivity". In R. Mizoguchi, Z. Shi, and F. Giunchiglia (Eds.): Asian Semantic Web Conference 2006, LNCS 4185, pages 616–631, 2006.

Сидорук І.А., Зайченко Ю.П.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Формування оптимального інвестиційного портфеля в умовах невизначеності

Інвестиції є ключовим фактором розвитку економіки країни, оскільки саме вони виступають найважливішим засобом забезпечення умов виходу з економічної кризи та надійним механізмом соціально-економічних перетворень, формують виробничий потенціал на новій науково-технічній основі, що неодмінно призводить до підвищення якісних показників господарської діяльності. Задача ефективного розміщення інвестиційних ресурсів та грамотного управління ними є особливо актуальною в сучасних умовах глибоких економічних змін та невизначеності, зумовлених фінансовою кризою економіки України. А боротьба з невизначеністю є однією з найважливіших задач розвитку ринку цінних паперів та інтенсивної інвестиційної діяльності. В зв'язку з цим в роботі застосовується нечітко-множинний підхід, де

- ризик портфеля – це не його волатильність, а можливість того, що очікувана прибутковість портфеля виявиться нижчою за деяку встановлену планову величину;
- кореляція активів в портфелі не розглядається і не враховується;
- прибутковість кожного активу – це невипадкове нечітке число (наприклад, трикутного або інтервального виду). Аналогічно, обмеження на дуже низький рівень прибутковості може бути як звичайним скалярним, так і нечітким числом довільного виду [1].

В роботі розглянуто пряму [2], двоїсту та багатокритеріальну [3] задачі оптимізації, також побудовано оптимальний портфель для заданого проміжку часу. Для отримання вхідних даних для системи оптимізації було використано НМГВА, який дозволив спрогнозувати прибутковість компонент портфеля у вигляді нечітких чисел, і на основі отриманих результатів, скориставшись методами оптимізації, побудувати оптимальний портфель. Вибір НМГВА для прогнозування був зумовлений його основними перевагами:

- метод не потребує задання моделі в явному вигляді, модель конструюється сама в процесі роботи алгоритму;
- метод працює на коротких вибірках;
- отримання прогнозу в вигляді нечітких чисел, що дозволяє врахувати невизначеність при прогнозуванні.

Завдяки використанню результатів прогнозування НМГВА, систему вдалося позбавити ризику суб'єктивності експертів. Адже саме експертна оцінка використовувалась раніше для визначення інтервалу прибутковості компонент портфеля, що виступає вхідними даними для системи оптимізації портфеля.

Таким чином, в роботі проведено дослідження в області управління інвестиційною діяльністю з використанням якісно нового нечітко-множинного підходу. Система оптимізації інвестиційного портфеля дозволяє автоматизувати процес пошуку оптимального рішення і надає можливість здійснювати науково-обгрунтоване управління своїми інвестиціями з можливістю відкидання планових збитків від володіння переоціненими або ризикованими активами, що підвищує ефективність бізнесу.

Література. **1.** Зайченко Ю.П. Нечеткие модели и методы в интеллектуальных системах: учебное пособие [для студентов высших учебных заведений] — К.: Слово, 2008. — С.341. — Библиогр. : с. 334—342.— ISBN 978-966-8407-79-6 **2.** Використання нечітко-множинного підходу з різними функціями приналежності для формування оптимального інвестиційного портфеля: матеріали міжнар. наук. конф. ISDMCI'2010, (Херсон, 17-21 трав. 2010 р.) / М-во освіти і науки, молоді та спорту України [та ін.]. —Х. : ХНТУ, 2010.— С.394–398. — Бібліогр. в кінці доп. **3.** Пряма та багатокритеріальна задачі оптимізації інвестиційного портфеля в умовах невизначеності: матеріали IV Всеукр. науково-практичної конф. ІТІА-2011, (Одеса, 12-14 жовтня 2011р.) / М-во освіти і науки, молоді та спорту України. — О.: ОНАПТ, 2011. — С.51. — Бібліогр. в кінці доп.

Тавров Д.Ю., Булах О.Р.
Національний технічний університет України "КПІ", Київ, Україна

Меметичний метод побудови системи нечіткого виведення для визначення ступеня кредитоспроможності особи

Вступ. Задача визначення ступеня кредитоспроможності особи (кредитного скоринга) виникає у банках та інших фінансових організаціях під час прийняття рішення про видачу кредитів клієнтам. Задача полягає в тому [1], щоб на основі інформації про клієнта обґрунтовано прийняти рішення, видавати йому кредит чи ні, і на яких умовах. Серед існуючих поширеними на практиці є класичні методи [2], які не враховують принципової нечіткості природи вхідних даних про клієнта. У літературі запропоновано використання нечітких систем для розв'язання цієї задачі [3], у яких знання про клієнта представлено у вигляді бази нечітких правил. Основною вадою такого роду систем є потреба в експертному визначенні відповідних правил. У даній роботі пропонується меметичний метод побудови бази правил, який не потребує залучення експерта, а базується на еволюційних обчисленнях.

Структура системи нечіткого виведення. Задачу визначення ступеня кредитоспроможності особи можна розглядати як задачу нечіткої класифікації, де кожну особу можна віднести до класу кредитоспроможних або некредитоспроможних осіб. Система нечіткого виведення (СНВ) для такої задачі складається з декількох вхідних лінгвістичних змінних X_i, $i = \overline{1,n}$, однієї вихідної змінної C — номеру класу, до якого потрібно віднести особу (1 — для кредитоспроможних осіб, 0 — для некредитоспроможних), модуля нечіткого виведення та бази нечітких правил S. Кожна вхідна змінна відповідає деякому атрибуту клієнта. Кожне правило $R_j \in S$, $j = \overline{1,s}$, в СНВ можна представити у вигляді:

$$\text{Якщо } x_1 - A_{j1}, \ldots, x_n - A_{jn}, \text{ то } C_j \text{ з } CF = CF_j,$$

де $A_{i1}, \ldots, A_{in}$ — значення вхідних змінних (нечіткі множини, визначені своїми функціями належності $\mu_{i1}, \ldots, \mu_{in}$), C_j — номер класу, CF_j — ступінь достовірності правила R_j.

Для побудови бази S шляхом застосування еволюційного алгоритму потрібно сформувати навчальну вибірку з m n-вимірних шаблонів $\mathrm{x}_p = (x_{p1}, x_{p2}, \ldots, x_{pn})$, $p = \overline{1,m}$, де x_{pi}, $i = \overline{1,n}$, — значення атрибутів p-го клієнта. Щодо кожного шаблону відомо, якому класу він належить. Тоді, клас та ступінь достовірності кожного нечіткого правила можна детерміновано визначити згідно з наступною процедурою [4]:

1. Обчислити ступінь сумісності кожного навчального шаблону $\mathbf{x}_p$ з правилом R_j:
$$\mu_j(\mathbf{x}_p) = \mu_{j1}(x_{p1}) \times \ldots \times \mu_{jn}(x_{pn}) \ . \quad (1)$$
2. Для кожного класу обчислити суму ступенів сумісності навчальних шаблонів із R_j:
$$\beta_{C_j}(R_j) = \sum_{\mathrm{x}_p \in C_j} \mu_j(\mathrm{x}_p) \ , \quad C_j = \overline{0,1} \ . \quad (2)$$
3. Пошук класу k, для якого досягнуто максимум $\beta_{C_j}(R_j)$. Якщо для обох класів досягнуто максимум $\beta_{C_j}(R_j)$, то визначити клас C_j неможливо, і CF_j покладають рівним 0.
4. Якщо можливо однозначно визначити максимальне значення $\beta_{C_j}(R_j)$ та клас k, який йому відповідає, то CF_j обчислюють за формулою
$$CF_j = \frac{\beta_k(R_j) - \beta_{1-k}(R_j)}{\beta_0(R_j) + \beta_1(R_j)} \ . \quad (3)$$

Модуль нечіткого виведення визначає клас шаблону $\mathbf{x}_p$ згідно з наступним підходом [4]. Для кожного шаблона модуль визначає єдине правило-переможець R_l за формулою

$$\mu_l(\mathbf{x}_p) \cdot CF_l = \max_{R_j \in S} (\mu_j(\mathbf{x}_p) \cdot CF_j) \ , \quad (4)$$

тобто правило, котре має максимальний добуток ступеня сумісності для шаблону $\mathbf{x}_p$ $\mu_j(\mathbf{x}_p)$ та ступеня достовірності CF_j. Якщо одночасно декілька правил є переможцями для одного шаблону, або ж $\mu_j(\mathbf{x}_p) = 0 \ \forall R_j \in S$, то вважають, що шаблон $\mathbf{x}_p$ не належить жодному класу.

Меметичний алгоритм побудови бази нечітких правил. У літературі виділяють два методи побудови бази нечітких правил систем нечіткого виведення (СНВ) за допомогою еволюційних обчислень. За Піттсбурзького підходу (Pittsburgh approach), кожний індивід у популяції в алгоритмі являє собою повну базу правил. За Мічиганського підходу (Michigan approach) кожний індивід є окремим правилом, а популяція являє собою повну базу правил. У даній роботі використовуватимемо другий підхід до побудови бази правил. Пропонований меметичний алгоритм є вдосконаленням еволюційного алгоритму, запропонованого в [4].

Кожний індивід (правило) у популяції являє собою вектор $\mathbf{r} = (r_1, \ldots, r_n)$, де r_i — номер функції належності, що відповідає i-ій лінгвістичній змінній. Якщо $r_i = 0$, то в якості функції належності для i-ої лінгвістичної змінної береться функція, тотожна одиниці.

Меметичний алгоритм побудови бази правил для СНВ передбачає виконання таких кроків:

1. Згенерувати початкову популяцію (базу нечітких правил) P з μ правил, вибираючи елементи правил випадковим чином за рівномірним законом.
2. Обчислити значення функції пристосованості $f(\mathbf{r})$ $\forall \mathbf{r} \in P$.
3. Вибрати λ пар батьківських індивідів.
4. Застосувати до кожної батьківської пари оператор схрещування з імовірністю p_c.
5. Застосувати до кожного отриманого нащадка оператор мутації з імовірністю p_m та помістити результат у множину P'.
6. Застосувати до множини P' оператор локального пошуку з імовірністю p_{mem}, який передбачає виконання таких кроків:
   - вибрати випадковим чином індекс q правила з P' та індекс елемента t у ньому;
   - шляхом перебору всіх можливих значень елемента t вибрати значення, котре дає найбільшу пристосованість правила q.
7. Замінити κ найменш пристосованих індивідів у P на індивідів із P'.
8. Перевірити виконання умови завершення: якщо вона виконується, зупинити алгоритм, або повернутися до кроку 2 в протилежному разі.

Функція пристосованості $f(\mathbf{r})$ в алгоритмі має вигляд

$$f(\mathbf{r}) = NCP(\mathbf{r}) \ , \tag{5}$$

де $NCP(\mathbf{r})$ — кількість тестових шаблонів, які правильно класифіковано за допомогою $\mathbf{r}$.

Практичні результати. Роботу розробленого меметичного алгоритму було протестовано на прикладі, описаному в [4, с. 609]. У прикладу використано навчальну вибірку розміром у 689 шаблонів із 15 атрибутами. Для коректності порівняння одержаних результатів параметри алгоритму було взято аналогічними описаним в [4]. Імовірність виконання локального пошуку було покладено рівною $p_{mem} = 0,1$. Середня частка правильно класифікованих шаблонів за 20 запусків еволюційного алгоритму в [4] складає 86,7%. Використання меметичного алгоритму з операцією локального пошуку, описаною вище, дає можливість одержати середню частку правильно класифікованих шаблонів, рівною 90,1%.

Література. **1.** Кочедыков Д. А. Система кредитного скоринга на основе логических алгоритмов класификации / Д. А. Кочедыков, А. А. Ивахненко, К. В. Воронцов [Електронний ресурс]. — 2003. — Режим доступу: www.ccas.ru/frc/papers/voron05credit.pdf. **2.** Nurlybayeva K. Algorithmic Scoring Models / K. Nurlybayeva, G. Balakayeva // Applied Mathematical Sciences. — 2013. — Vol. 7. — P. 571–585. **3.** Shang K. Applying Fuzzy Logic to Risk Assessment and Decision-Making / K. Shang, Z. Hossen. — Casualty Actuarial Society, Canadian Institute of Actuaries, Society of Actuaries, 2013. — 59 p. **4.** Ishibuchi H. Performance Evaluation of Fuzzy Classifier Systems for Multidimensional Pattern Classification Problem / H. Ishibuchi, T. Nakashima, T. Murata // IEEE Transactions on Systems, Man, and Cybernetics. — 1999. — Vol. 29, No. 5. — P. 601–618.

Филатова А.Е.
Национальный технический университет "Харьковский политехнический институт", Харьков, Украина

Программный модуль повышения качества визуализации рентгенологических изображений

Постановка проблемы и анализ литературы. На сегодняшний день основным методом скрининг диагностики молочных желез является рентгенологическая маммография [1]. В медицинской практике при проведении рентгенологических исследований стремятся максимально уменьшить энергию ионизирующего излучения, а также время экспозиции, что, в свою очередь, приводит к снижению качества визуализации исследуемых органов. Поэтому задача повышения качества визуализации маммограмм за счет цифровой обработки изображений является актуальной научно-практической задачей. Методы цифровой обработки изображений, позволяющие повысить качество визуализации, можно разделить на два больших класса: поэлементные преобразования (изменение яркости и контрастности) и оконные преобразования (линейные и нелинейные методы фильтрации) [2, 3]. Кроме перечисленных методов, в основу которых положена работа с одним изображением, можно выделить группу методов, позволяющих получить результирующее изображение на основе анализа двух других изображений (исходного и корректирующего) [4]. Такие методы называются режимами наложения и широко используются, например, в графическом редакторе Adobe Photoshop [5].

Цель работы – разработка метода повышения качества визуализации цифровых рентгенологических изображений на основе математических методов обработки полутоновых изображений и проверка работоспособности разработанного метода на примере обработки маммограмм.

Метод визуализации молочной железы на маммограмме. Для решения поставленной задачи в работе предлагается следующая математическая модель изображения молочной железы $I[x, y]$:

$$I[x, y] = D[x, y] + F[x, y] + R[x, y], \tag{1}$$

где $D[x, y]$ – изображение, содержащее структуру тканей; $F[x, y]$ – изображение, содержащее плотность тканей; $R[x, y]$ – шумовая составляющая; $(x, y) \in M$ – координаты пикселей цифрового изображения; M – множество точек, принадлежащих изображению молочной железы.

Т.к. характеристики шума неизвестны, а полезный сигнал является нестационарным, то для устранения шумовой составляющей предлагается использовать адаптивную винеровскую фильтрацию с квадратным окном размером $p_1 \times p_1$: $A[x, y] = wiener(I[x, y], p_1)$, где $A[x, y]$ – изображение после фильтрации; $wiener()$ – функция, реализующая адаптивную фильтрацию Винера; p_1 – линейный размер апертуры (например, $p_1 = 5$). Тогда для повышения качества визуализации маммограмм необходимо найти оценку изображения $\widetilde{D}[x, y]$, содержащего структуру тканей, и оценку изображения $\widetilde{F}[x, y]$, содержащего плотность тканей.

Для получения оценки $\widetilde{D}[x, y]$ в модели (1) необходимо удалить из изображения $A[x, y]$ неравномерность яркости, обусловленную различной толщиной тканей от края молочной железы к грудной клетке, т.е. выровнять интенсивность фона. Оценку фона можно получить, выполнив адаптивную винеровскую фильтрацию с большим размером апертуры $B[x, y] = wiener(I[x, y], p_2)$, где p_2 составляет порядка 2%–3% от размеров изображения. Тогда с помощью режима наложения Divide можно удалить полученную оценку фона $B[x, y]$ (корректирующее изображение) из отфильтрованного изображения $A[x, y]$. Исследования на реальных маммограммах показали, что для повышения контрастности деталей изображения после наложения Divide необходимо выполнить γ-коррекцию с параметром $\gamma > 1$: $\widetilde{D}[x, y] = \left(\frac{A[x,y]}{B[x,y]} mxI\right)^{\gamma}$, где mxI – максимальное значение яркости выходного изображения.

Оценка фона $B[x, y]$ содержит не только информацию о неравномерности яркости из-за

различной толщиной тканей от края молочной железы к грудной клетке, но и яркостную составляющую, которая отражает плотность различных тканей молочной железы. Тогда с помощью режима наложения Color Burn можно получить оценку $\widetilde{F}[x,y]$ модели (1), используя $B[x,y]$ в качестве корректирующего изображения: $\widetilde{F}[x,y] = mxI - \frac{mxI - A[x,y]}{B[x,y]} mxI$. С учетом предлагаемой математической модели (1) оценка изображения молочной железы может быть вычислена по следующему выражению (режим наложения Normal с параметром прозрачности 50%): $\widetilde{I}[x,y] = \frac{\widetilde{D}[x,y] + \widetilde{F}[x,y]}{2}$. С целью коррекции динамического диапазона изображения предлагается использовать нелинейное контрастирование с использованием s-образного закона следующего вида: $S[x,y] = \frac{mxI}{1+e^{-(J[x,y]-a)b}}$, где $J = [x,y] = \frac{\widetilde{I}[x,y] - mnI'}{mxI' - mnI'} \in [0,1]$ – нормированное изображение $\widetilde{I}[x,y]$; mnI', mxI' – минимальное и максимальное значения изображения $\widetilde{I}[x,y]$; a, b – константы, определяемые исходя их статистических характеристик изображения $\widetilde{I}[x,y]$.

Программный модуль IMRI-MAM. С целью ускорения математической обработки программный модуль IMRI-MAM написан на языке программирования Matlab R2014a. Т.к. программное обеспечение цифровых рентгенологических маммографических комплексов (ЦРМК) фирмы «Радмир» разработано в среде программирования Borland C++ Builder 6.0, то для интегрирования программного модуля IMRI-MAM в рассматриваемые комплексы необходимо создать одноименную библиотеку динамической компоновки (DLL) IMRI-MAM, содержащую функции реализации разработанного математического метода повышения качества визуализации маммограмм. Существует два основных способа подключения DLL-библиотеки: статический и динамический. При статическом связывании в Borland C++ Builder 6.0 необходимо к проекту подключить lib-файл. Однако при создании DLL-библиотеки с помощью Matlab R2014a lib-файл может иметь формат, используемый в среде программирования Microsoft Visual C++ Studio не старше 2008 года. Проблема заключается в том, что компании Microsoft и Borland используют различный формат lib-файлов. Эту проблему можно обойти следующим образом: либо с помощью соответствующих утилит создать lib-файл нужного формата, либо использовать динамическую компоновку DLL-библиотеки, для которой не требуется наличие lib-файла. В данной работе выбран динамический способ подключения разработанной DLL-библиотеки IMRI-MAM к программному обеспечению ЦРМК фирмы «Радмир». Кроме того, для работы библиотеки IMRI-MAM в составе ЦРМК фирмы «Радмир» необходимо установить среду MCR, обеспечивающую штатную работу компонентов Matlab.

Выводы. В работе предложена математическая модель изображения молочной железы на маммограмме, учитывающая свойства регистрируемых изображений. На основе предложенной математической модели и существующих методов цифровой обработки сигналов разработан метод повышения качества визуализации маммограмм. Рассмотрены вопросы разработки программного модуля повышения качества визуализации рентгенологических снимков IMRI-MAM и подключения разработанного программного модуля IMRI-MAM к отечественным цифровым рентгенологическим маммографическим комплексам MADIS и SYMA производства фирмы «Радмир» ДП АО НИИРИ (Украина, г. Харьков). Проверена работоспособность метода на реальных цифровых маммограммах.

Литература. **1.** Терновой С.К. Лучевая маммология / С.К. Терновой, А.Б. Абдураимов. – М.: ГЭОТАР-Медиа, 2007. – 128 с. **2.** Гонсалес Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М.: Техносфера, 2005. – 1072 с. **3.** Красильников Н.Н. Цифровая обработка 2D- и 3D-изображений: учеб. Пособие / Н.Н. Красильников. – СПб.: БХВ-Петербург, 2011. – 608 с. **4.** Бойко Д.А. Метод визуализации патологических структур на маммограммах с использованием послойного наложения / Д.А. Бойко, А.Е Филатова. – Вісник НТУ «ХПІ». Серія: Математичне моделювання в техніці та технологіях. – Х.: НТУ «ХПІ», 2014. – №6 (1049). – С. 29-34. **5.** Айсманн К. Ретуширование и обработка изображений в Photoshop (3-е издание) / К. Айсманн, У. Палмер – М.: Издательский дом «Вильямс», 2007. – 560 с.

Фіногенов О.Д., Штрик Р.В.
Національний технічний університет України "КПІ", Київ, Україна

Аналіз можливостей інтернет сервісів знаходження плагіату

Накопичення інформації призводить до повсякчасного її копіювання та повторення. Все більш гостро постає проблема визначення унікальності контенту та встановлення авторства [1]. Сучасні обсяги інформації вимагають наявності автоматизованих або автоматичних засобів перевірки.

В мережі Інтернет існують чимало сервісів для визначення факту плагіату або відсотку запозичень [2–6] та інші, що використовують різні алгоритми та/або бази для перевірки. Але, відповідно, також розробляються та вдосконалюються і засоби їх обходження.

Найбільш поширеними засобами приховування плагіату для електронних документів є: заміна символів (літери, знаки пунктуації в тексті замінюються іншими, схожими або такими, що не відрізняються за написанням, але які мають різні числові значення в кодах), використання «стоп слів» (слова, що не несуть самостійного смислового навантаження, тому їх можна видалити без спотворення змісту тексту) та «стоп символів», синонімізація тексту (заміна деяких слів на синоніми, в результаті текст має різне написання), зміна порядку тексту, використання іншомовних джерел з їх подальшим перекладом та інші.

Більшість алгоритмів є нестійкими до тих, чи інших модифікацій тексту. Це робить їх неефективними при відсутності додаткових засобів боротьби з методами приховування плагіату. Для боротьби з можливими модифікаціями при запозиченнях проводиться нормалізація тексту, при якій текст приводиться до нормальної форми, де можливі модифікації при запозиченнях змінюються таким чином, що вихідний та модифікований текст мають однакову нормальну форму.

Так як використання унікального тексту не дає змоги перевірити ефективність даних сервісів, то в якості тесту було використано неоригінальний текст, що становив собою фрагмент правил прийому, які опубліковані на сайті університету [7]. Початковий розмір тексту "Тесту 1" складав 5 абзаців (228 слів).

1. "Тест 1" - базовий фрагмент тексту [7];
2. "Тест 2" - текст "Тесту 1" з додаванням 5 "стоп слів", обраних зі словника;
3. "Тест 3" - текст "Тесту 1" з додаванням "стоп символів" (із заміною символу "а" (укр.) на "a" (лат.)). Із загальної кількості 1814 символів замінено 30, що становить менше 2 %;
4. "Тест 4" - текст "Тесту 1" із застосування перестановки абзаців;
5. "Тест 5" - текст "Тесту 1" перекладений російською мовою.

В якості результату за тестом наводиться найвищий відсоток подібності та, в деяких випадках, кількість посилань за цим відсотком. Для [6] результати надані у вигляді кількості запозичень, оскільки у відсотковому відношенні, результуючі дані сервіс не надає (табл. 1).

Табл. 1. Порівняльний аналіз сервісів для перевірки унікальності тексту

Сервіс	Тест 1	Тест 2	Тест 3	Тест 4	Тест 5
Антиплагиат [2]	44.97 %	43.74 %	42.31 %	44.24 %	23.56 %
Findcopy [3]	100 % (4)	92.8 %	100 % (3)	99.0 %	0 %
Text.ru [4]	100 % (9)	94 % (9)	83 % (9)	100 % (8)	81 %
ContentWatch [5]	>90 %	>90 %	>81.3 %	>61.3 %	>46.9 %
PR.CY [6] (кількість запозичень)	27	21	27	27	6

Результати "Тесту 1" показують, що початковий текст був знайдений усіма сервісами, але [2] не знайшов джерело початкового тексту на сайті НТУУ "КПІ". Велика кількість 100% результатів пояснюється тим, що текст правил прийому до кожного ВНЗ розробляється на основі єдиних Умов прийому до ВНЗ і містить багато спільних фрагментів.

Результати "Тесту 2" демонструють, що додавання навіть невеликої кількості "стоп-слів" (5 слів ≈ 2 %), значно впливають на значення відсотку унікальності більшості сервісів.

Після додавання стоп символів ("Тест 3"), результати роботи не погіршились тільки у [3] та [6]. Враховуючи те, що відсоток замінених символів є дуже малим, можна зробити висновок про значний вплив на роботу сервісів даного методу приховування факту запозичення.

Перестановка абзаців ("Тест 4") мала значний вплив лише на [5].

Знайдені посилання з іншомовним перекладом ("Тест 5") знаходяться на сайтах, що підтримують багатомовні версії і дійсно мають переклад Умов прийому. Серед знайдених джерел жоден сервіс не знайшов сайт НТУУ "КПІ". Сервіс [3] взагалі видав результат – 100 % унікальності.

Результати, що були отримані за допомогою сервісу [5], показують, що даний сервіс обробляє інформацію не лише зі сторінок сайтів, але й з розташованих на них документів. Так, наприклад, для даного тесту ("Тест 1"), були знайдені посилання на файли у форматах *.doc та *.pdf.

Дані, що отримані за допомогою сервісу [6] не тільки є найменш інформативними з розглянутих сервісів, але й при повторному завантаженні тесту видали дещо інші результати, які відрізнялися від початкових на 2-3 посилання.

Окремим питанням, що потрібно дослідити є цитування. Використання правильно оформленої цитати не може вважатися плагіатом згідно [1], і це потрібно враховувати. Серед обраних сервісів лише позиція [2] чітко прописана – цитати на власний розсуд перевіряючого. Інші сервіси не зазначають, як вони вирішують це питання.

Значною перевагою володіють сервіси, що вказують, які саме фрагменти тексту були запозичені. Найчастіше це реалізується підсвічуванням однакових частин. Це є корисним функціоналом, адже, при перевірці документів значного об'єму перевіряючому знадобиться на порядок більше часу, ніж то дозволено. Та і коректність роботи сервісу перевірити наглядно – значно простіше.

При перевірці відмічено особливість сервісу [6], який в результаті перевірки видав дуже велику кількість посилань, з яких було запозичено фрагменти тексту. Насправді, значна частка посилань, що була відмічена кількістю запозичень рівною 1, не є коректною.

Висновки. Не зважаючи на велику кількість засобів перевірки документів, більшість з них чутлива до типових методів приховування факту запозичення. Найбільш вразливим місцем сервісів аналізу запозичень є приховування плагіату при перекладі, але цей метод має ряд обмежень та потребує (на відміну від інших) більше зусиль.

Іншою проблемою є метод додавання "стоп символів". Потрібно враховувати, що кількість замін може бути достатньо великою, що призведе до значного погіршення роботи та результатів сервісу.

Найбільш достовірним з розглянутих сервісів виявився сервіс [4], але до його недоліків слід віднести недостатньо ефективну роботу з методами додавання "стоп слів" та "стоп символів", та незадовільний пошук і порівняння при перекладі текстів.

Для забезпечення ефективного пошуку плагіату, потрібно враховувати всі виявлені недоліки, а також переваги кожного із сервісів, наприклад, виділення запозичених фрагментів, пошук серед файлів та інше.

Література. **1.** Закон України "Про авторське право і суміжні права" № 3792-XII від 23.12.1993 р.: редакція від 05.12.2012. – Режим доступу: http://zakon4.rada.gov.ua/laws/show/3792-12/page1. – Дата доступу: 05.04.2015. **2.** Інтернет-сервіс "Антиплагиат". – Режим доступу: http://www.antiplagiat.ru/. – Дата доступу: 05.04.2015. **3.** Інтернет-сервіс "Findcop". - Режим доступу: http://findcopy.ru/. – Дата доступу: 05.04.2015. **4.** Інтернет-сервіс "Text.ru". - Режим доступу: http://text.ru/. – Дата доступу: 05.04.2015. **5.** Інтернет-сервіс "ContentWatch". - Режим доступу: http://www.content-watch.ru/text/. – Дата доступу: 05.04.2015. **6.** Інтернет-сервіс "PR.CY". - Режим доступу: http://pr-cy.ru/unique/. – Дата доступу: 05.04.2015. **7.** Правила прийому до НТУУ "КПІ" в 2015 році - Режим доступу: http://kpi.ua/rule/. – Дата доступу: 05.04.2015.

Чегренець О.С.
Національний технічний університет України "КПІ", Київ, Україна

Острівна модифікація алгоритму оптимізації мурашиними колоніями

Актуальність. Задачі комбінаторної оптимізації постійно виникають в різноманітних сферах господарської діяльності людини. Реальні задачі, зазвичай, відносяться до класу NP-складних задач великої розмірності. На практиці для розв'язання таких задач використовуються наближені алгоритми, що дозволяють за обмежений час гарантовано повернути варіант розв'язку [1]. Навіть незначне покращення знайденого розв'язку в кінцевому підсумку може зекономити значні кошти. Тому проблема розробки більш точних модифікацій існуючих наближених алгоритмів є завжди актуальною.

Оптимізація мурашиними колоніями. Останнім часом при оптимізації складних систем все частіше застосовуються природні механізми пошуку розв'язків. Одним з найвідоміших представників «природних обчислювань» є алгоритм оптимізації мурашиними колоніями (ОМК). Для обміну інформацією агенти використовують феромон - спеціальний секрет, що відкладається як слід при переміщенні мурахи. Феромонний слід в ОМК служить розподіленою чисельною інформацією, яка разом з еврестичною інформацією про задачу, використовується мурахами для недетермінованої побудови розв'язку [2].

Класичний алгоритм досить ефективно знаходить розв'язки лише для задач невеликої розмірності. Однак простота класичного мурашиного алгоритму залишала можливості для збільшення ефективності алгоритму. В основному, ці вдосконалення пов'язані з використанням історії пошуку, більш ретельним дослідженням областей навколо вже знайдених вдалих рішень, більш інтенсивного використання поточного найкращого розв'язку, вдосконалення логіки вибору наступного міста мурахою та механізму оновлення феромон [3]. Однією з найбільш ефективних модифікацій алгоритму є MAX-MIN оптимізація мурашиними колоніями з застосування локального пошуку.

Острівна модифікація алгоритму оптимізації мурашиними колоніями. Іншим підходом для модифікації алгоритму є застосування кооперативних метаалгоритмів, де в якості базового алгоритму може використовуватися одна з існуючих модифікацій мурашиного алгоритму.

Основна ідея кооперативних алгоритмів полягає в незалежному виконанні окремих екземплярів алгоритмів (базових алгоритмів) з періодичним обміном даними між ними [4]. При використанні острівної модифікації для мурашиного алгоритму створюється декілька незалежних колоній мурах (островів), кожна з яких незалежно від інших розв'язує поставлену задачу комбінаторної оптимізації. Кожна колонія може використовувати як однакові, так і різні модифікації алгоритму мурашиної оптимізації. Керуючий метаалгоритм періодично створює агреговану матрицю феромон, що базується на матрицях кожного острова. Вага кожної окремої колонії в агрегованій матриці залежить від якості поточного найкращаого розв'язку, знайденого даним островом. Після створення агрегованої моделі метаалгоритм замінює матриці феромон кожного острова на агреговану.

Експериментальні дані засвідчують те, що острівна модифікація алгоритму дозволяє отримати меншу середню відносну похибку в порівнянні з незалежним використанням базових алгоритмів. Кооперативний алгоритм дозволяє використати усі переваги різноманітних модифікацій мурашиного алгоритму і диверсифікувати ризики передчасної збіжності.

Література. **1.** И.В. Сергиенко, Л.Ф. Гуляницкий, С.И. Сиренко Классификация прикладных методов комбинаторной оптимизации // Кибернетика и системный анализ. – 2009. – № 5. – С. 71–83. **2.** С.Д. Штовба Муравьиные алгоритмы // Exponenta Pro. Математика в приложениях. – 2003. – №4. – с. 70-75. **3.** M. Dorigo, C. Blumb Ant colony optimization theory: A survey // Theoretical Computer Science. – 2005. – P. 243-278. **4.** Л.Ф. Гуляницький, С.И. Сиренко Кооперативные моделе-ориентированные метаэвристики для задач комбинаторной оптимизации. // Intelligent Support of Decision Making. – Sofia: ITHEA. – 2009. – P. 165–172.

Шульган І.В., Боровська Т.М.
Вінницький національний технічний університет, факультет комп'ютерних систем та автоматики, Вінниця, Україна

Узагальнення моделі оптимального агрегування систем «виробництво-розвиток» з урахуванням невизначеностей зовнішнього середовища виробництва

Сьогодні фундаментом усталеного розвитку та ефективності регіональних структур є комплексна ефективність аграрного виробництва. Необхідна умова цього – оптимальність виробничих циклів агропідприємств різного рівня – від рослинництва і до тваринництва, до переробки в кінцеві продукти. Конкретні приклади – спеціалізовані господарства по вирощуванню картоплі, кукурудзи, буряка, свиноферми, птахофабрики, масложирком-бінати, пивзаводи. В таких виробничих одиницях постійно виникають питання: чи варто придбати певну техніку, чи варто будувати біореактор, а якщо варто – то якої потужності. Відповіді повинні бути розрахованими і комплексно обґрунтованими. Для цього потрібна відкрита персональна система, зручна для користувача. Основа системи – ефективні моделі (прості і адекватні) та зручні інтерфейси [1].

Невирішена задача – оптимальне агрегування неоднорідної структури «виробництво розвиток». Виконано розробку бінарного оператора оптимального агрегування систем «виробництво-розвиток», аналіз якого представлено на рисунку 1. Робота не має близьких аналогів. Всі розроблені моделі програмно реалізовані і досліджені [2].

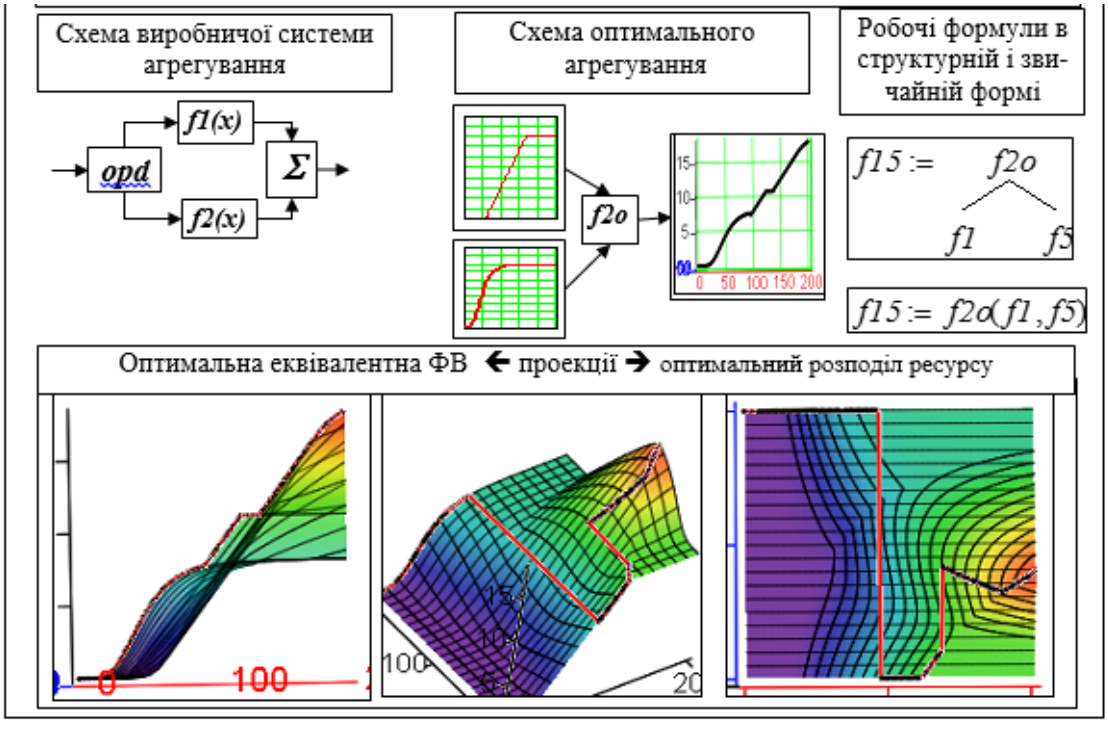

Рис. 1. Аналіз оператора оптимального агрегування

Методологія оптимального агрегування інтегрує в собі: агрегування моделі виробничої системи в еквівалентний елемент, субоптімізації розподілу ресурсу в підсистемах, і алгебру виробничих функцій, подібну алгебрі передавальних функцій для лінійних динамічниxx систем [3]. На рисунку 2 подано приклад порівняльного аналізу агрегування структури «виробництво, розвиток» та базової паралельної структури. Маємо єдину методологію і

суттєво різні методи. Три проекції 3Д графіка дозво-ляють побачити: оптимальну еквівалентну функцію виробництва і оптимальний розподіл ресурсу системи.

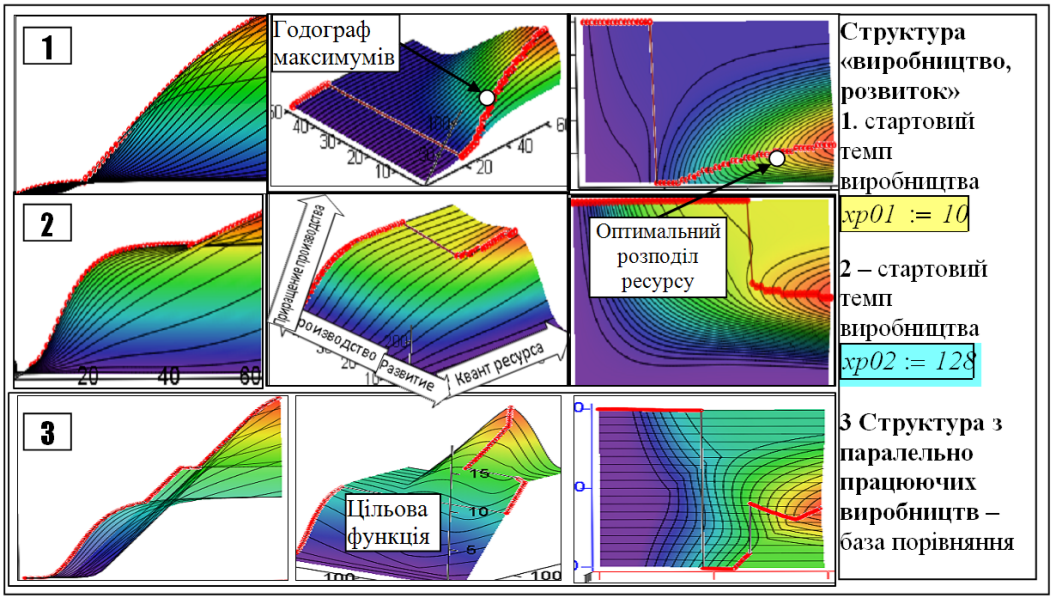

Рис. 2. Приклад. Порівняльний аналіз агрегування структури «виробництво-розвиток»

Висновки. На підставі аналізу процесів функціонування та розвитку сучасних високотехнологічних виробництв та існуючих методів інтеграції функціональних підсистем сформульована задача оптимального управління інтегрованими виробничими системами.

Виконано дослідження неоднорідних структур виробничих систем класу «виробництво, розвиток» з параметричними зв'язками, розроблено і програмно реалізовано оператор оптимального агрегування з урахуванням невизначеностей функцій виробництва і функцій розвитку і урахуванням збурень від зовнішнього середовища.

Розроблена модель є ефективним засобом для проведення наукових досліджень, придатна для створення підсистеми управління процесами виробництва та розвитку підприємства. Розроблена модель - параметризована і модульна, тому вона є ефективною основою для створення нових моделей інтегрованих виробництв.

Література. **1.** Месарович М. Д. Теория иерархических многоуровневых систем / М. Д. Месарович, З. Мако, М. Такахара – М.: Мир, 1973. – 310 с. **2.** Боровська Т. М. Метод оптимального агрегування в оптимізаційних задачах: монографія / Т. М. Боровська, І. С. Колесник, В. А. Северілов. – Вінниця: УНІВЕРСУМ–Вінниця, 2009. – 229 с. – ISBN 978–966–641–285–3. **3.** Боровська Т. М. Моделювання та оптимізація у менеджменті: навч. посіб. для студ. ВНЗ / Т. М. Боровська, В. А. Северілов, С. П. Бадьора, І. С. Колесник. – Вінниця: УНІВЕРСУМ – Вінниця, 2009. – 145 с. – ISBN 978–966–641–287–7.

Plenary talks

System analysis of complex systems of various nature

Intelligent systems for decision-making

High Performance Computing and Microsystems Engineering

Progressive information technologies

3 High Performance Computing and Microsystems Engineering

Section 3

High Performance Computing and Microsystems Engineering

1. High performance computing: Supercomputing, distributed, cloud, Grid, and quantum computing, high performance systems and networks..
2. Microsystems engineering and design.

Секция 3

Высокопроизводительные вычисления и разработка микросистем

1. Высокопроизводительные вычисления: Суперкомпьютерные, распределенные, облачные, Грид- и квантовые вычисления, высокопроизводительные системы и сети..
2. Разработка и проектирование микросистем.

Секція 3

Високопродуктивні обчислення та розробка мікросистем

1. Високопродуктивні обчислення: Суперкомпьютерні, хмарні, Грід- та квантові обчислення, високопродуктивні системи та мережі..
2. Розробка та проектування мікросистем.

Бабенко О.А., Гіоргізова-Гай В.Ш.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Балансування навантаження "хмарних" додатків

Говорячи про «хмарні» обчислення не можна не відзначити таку технологію, як балансування навантаження. Для забезпечення узгодженої роботи вузлів обчислювальної мережі на стороні хмарного провайдера використовується спеціалізоване проміжне програмне забезпечення, що забезпечує моніторинг стану обладнання і програм, балансування навантаження, забезпечення ресурсів для вирішення завдання.

Одним з основних рішень для згладжування нерівномірності навантаження на послуги є розміщення шару серверної віртуалізації між шаром програмних послуг та апаратним забезпеченням. В умовах віртуалізації балансування навантаження може здійснюватися за допомогою програмного розподілу віртуальних серверів по реальним.

Більшість хмарних додатків сьогодні створюється з використанням триланкового розподіленого підходу [1]. При доступі до ресурсів веб-додатку запит користувача спочатку обробляється DNS-сервером, який видає користувачеві IP-адресу одного з інтерфейсних серверів переднього плану (front-end). Користувач взаємодіє з одним з таких серверів, що забезпечують обробку статичного змісту і формування візуального результату виконання запиту, а також створення захищеного з'єднання. При обробці динамічних запитів, що вимагають обчислювальних дій чи запиту до динамічно сформованих даних, виконується переадресація запиту на сервер додатків середнього плану (middle-end). Сервер додатків, в свою чергу, взаємодіє з серверами заднього плану, що відповідають за зберігання даних (back-end).

З точки зору архітектури можна виділити наступні класи рішень, використовувані для балансування: балансування з пропуском трафіку через один пристрій балансування; балансування без пропуску трафіку через один пристрій балансування.

Процедура балансування здійснюється за допомогою цілого комплексу алгоритмів і методів, відповідних наступним рівням моделі OSI [2]:

- мережевому;
- транспортному;
- прикладному.

Балансування на мережевому рівні може здійснюватися за допомогою різноманітних способів: DNS-балансування, побудова NLB-кластеру, балансування за IP з використанням додаткового маршрутизатора, балансування за територіальною ознакою.

Балансування на транспортному рівні полягає у наступному: клієнт звертається до балансувальника, той перенаправляє запит одному з серверів, який і буде його обробляти.

При балансуванні на прикладному рівні балансувальник аналізує клієнтські запити і перенаправляє їх на різні сервери залежно від характеру запитуваної контенту.

Існує декілька критеріїв, в залежності від яких можна обрати необхідний метод балансування. До цих критеріїв відносяться:

- особливості архітектури "хмарної" системи та додатка;
- масштаб та територіальна розподіленість "хмарної" інфраструктури;
- необхідність підтримувати сесію користувача з одним і тим же вузлом;
- тривалість підключень;
- кількість підключень в даний момент часу.

На базі обладнання кафедри СП ННК "ІПСА" було встановлено систему дистанційного навчання в "хмарне" середовище на основі платформи CloudStack і досліджено ефективність використання різних методів "хмарного" балансувальника платформи.

Література. **1.** Балансировка масштабируемых приложений: http://www.osp.ru/os/2012/08/13019244 **2.** Балансировка нагрузки: основные алгоритмы и методы: http://habrahabr.ru/company/selectel/blog/250201

Булах Б.В., Крамар О.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Семантична інтеграція грід та хмарних ресурсів до складу комплексів інженерних обчислень

Сучасні комплекси інженерних досліджень потребують залучення високопродуктивних ресурсів, які б уможливили рішення складних обчислювальних задач при адекватних витратах часу. Як правило, доступ користувачів до таких ресурсів організований або через грід-середовище, або через хмарну інфраструктуру. Як правило, на етапі розробки програмної архітектури закладається орієнтація лише на один зі способів забезпечення ресурсами, хоча підтримка обох надала б програмним системам додаткової гнучкості.

Одним з варіантів побудови автоматизованих програмних комплексів вирішення інженерних задач є концепція "екосистеми сервісів". В рамках цієї моделі функціонал комплексу базується на функціоналі окремих веб-сервісів, які входять до складу комплексу, що надає ряд переваг: функціонал може легко бути розширений за рахунок реєстрації нових сервісів, комплекс може надавати нові можливості шляхом комбінування ("оркестрування") функіоналу доступних сервісів, окремі складові комплексу (сервіси) можуть динамічно заміщуватись, розроблятися й підтримуватись незалежними групами розробників.

Крім того, можна пропонувати таку модель для вирішення проблеми поєднання в складі комплексу інженерних обчислень можливостей грід та хмарних обчислень. В цьому випадку робота з проміжним шаром програмного забезпечення (грід / IaaS / PaaS) має бути прихована за відкритими інтерфейсами веб-сервісів, що б уніфікувало в рамках програмного комплексу запуск обчислювальних задач на різнорідних ресурсах. В рамках відкритої екосистеми сервісів таке рішення потребує впровадження бази знань, що містила б інформацію про наявні службові сервіси взаємодії з ресурсами, їх специфіку, формати даних та ін., описану за допомогою онтологій, а також — інструментарію здійснення логічних висновків над онтологіями. На рис.1 зображено схему, що пропонується для інтеграції грід та хмарних ресурсів до складу сервісно-орієнтованого обчислювального комплексу. Ключовим елементом є сервіс-місток виконання задач WS-E, що приховує деталі взаємодії з конкретним програмним забезпеченням доступу до ресурсів грід (через сервіси WS-G), або хмари (через сервіси WS-C), причому і самі грід ресурси можуть бути розгорнуті у хмарі [1]. Сервіс-місток взаємодіє з базою знань — семантичним реєстром сервісів (через веб-сервіси WS-R) для організації роботи з конкретним типом ресурсів. На заваді реалізації тут може стати той факт, що інтерфейс сервісів WS-C постійно модифікують постачальники хмарних ресурсів, що вимагатиме оперативного актуального оновлення інформації в семантичному реєстрі.

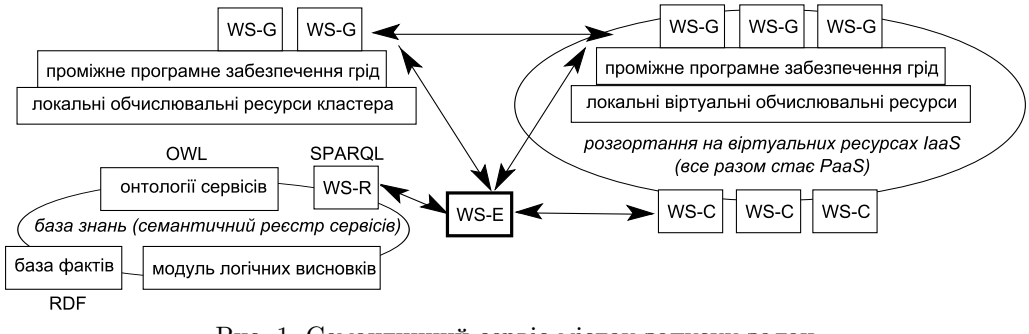

Рис. 1. Семантичний сервіс-місток запуску задач

Література. **1.** Rings T., Grabowski J. Pragmatic Integration of Cloud and Grid Computing Infrastructures / proc. int. conf. IEEE CLOUD, 2012. - p. 710-717.

Вергун О.С.
Національний технічний університет України "КПІ", факультет інформатики та очислювальної техніки, Київ, Україна

Розробка інтерфейсу для комп'ютерних кластерних систем

На теперешній час використання комп'ютерних кластерних систем стає все більш популярне. Вони використовуються якості хмарних обчислень, хмарних сховищ та систем для вирішення комплексних задач. Програмне забезпечення, що використовується у даних систем, у більшості, командний рядок операційної системи. Враховуючи велику кількість недоліків описаного програмного забезпечення, його необхідно модернізувати.

Одним із варіантів покращення цього програмного забезпечення є побудова інтерфейсу для користувачів. Це збільшить продуктивність користування, зменшить кількість помилок допущених при вводі запитів, збільшить зручність використання. Для побудови такої системи для початку потрібно найти спосіб взаємодії з командним рядком операційної системи, вибрати належну мову/фреймворк, які дозволять в повній мірі забезпечити необхідні інструменти для виконання даної задачі. Показана діаграма, яка ілюструє загальну структуру взаємодії користувача з комп'ютерною кластерною системою:

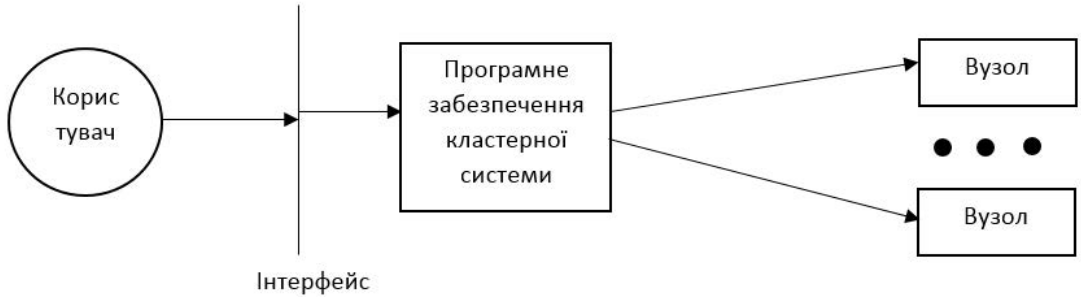

Рис. 1. Діаграма взаємодії користувача з кластерною системою

Основна складність розробки цього інтерфейсу це взаємодія з різними комп'ютерними кластерними системами. Ця задача вимагає аналізу всіх сучасних програмних забезпечень кластерних систем, їхнього синтаксису, специфіки і т.д. для створення уніфікованої структури або інтерфейсу для кожної кластерної системи. Також існують команди, які є унікальні для певної системи серед інших систем, тому виникне необхідність ще створити окремий список унікальних команд для кожної системи. Остання задача, яку необхідно вирішити, це взаємодія інтерфейсу з командним рядком та наоборот.

Для вирішення цих задач використовується фреймворк Ruby on Rails для написання серверної частини програми та фреймворк Javascript - AngularJS - для написання клієнтської частини. Описані програмні засоби повною мірою зможуть виконати поставлені задачі. Ця програма має вигляд веб-сторінки. Робочий процес цієї програми має наступний характер:

1. Користувач вибирає потрібну кластерну систему і в ній вибирає необхідну команду. У цій команді користувач заповнює поля даними і відправляє запит на сервер(нажимаючи кнопку "Process").
2. Ця команда оброблюється на сервері: приходить команда з клієнтської частини, відповідний фреймворк записує цю команду у командний рядок ОС і ОС виконує команду. Результат виконання ОС записується в командний рядок, а з командного рядка фреймворк передає результати назад користувачу.
3. Користувач отримує відповідь на веб-сторінку.

Описана програма стане засобом залучення не тільки висококваліфікованих спеціалістів у технічній сфері, а й спеціалістів, які не мають необхідних технічних знань, до використання кластерних систем для власних потреб, зокрема наукової діяльності.

Ігнатченко Л.О.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Оптимізація взаємодії веб-сервісів у складі системи GridALLTED

Працездатність системи під час високих навантажень є однією з найважливіших характеристик роботи будь-якого web-додатку. Це вимагає від розробників приділення особливої уваги до побудови архітектури програмного забезпечення (ПЗ). Так звані «вузькі місця» можуть спричиняти збій роботи всієї системи під час піку активності користувачів. Таким чином, важливим для підтримки працездатності ПЗ є виявлення проблемних компонентів програми, що через свою архітектуру не завжди можуть забезпечувати стабільність роботи системи.

Опис взаємодії архітектурних складових GridALLTED. Пакет GridALLTED є CAD-системою, що призначена для проектування, моделювання, аналізу різноманітних пристроїв, таких як системи управління та динамічні системи, що можуть складатися з компонентів різної фізичної природи – електронних, гідравлічних, пневматичних, механічних та інших [1]. На даний момент система реалізована таким чином:

1. Сформований користувачем маршрут проектування [2] надсилається з веб-сторінки на веб-сервер порталу.
2. З веб-серверу порталу після транслювання маршрут надсилається на сервер системи виконання завдання (СВЗ).
3. Після трансляції у сценарій виконання [3] BPEL-процес надсилає дані серверу грід-сервісу (СГС), що взаємодіє з самим грід-ресурсом.
4. Веб-сторінка періодично оновлює інформацію про статус виконання маршруту з бази даних порталу, а паралельно з цим сервер системи виконання завдання періодично запитує статус виконання у сервера грід-сервісу GridALLTED.

«Вузьким місцем» даного підходу є те, що під час виконання періодичних запитів взаємодія між СВЗ та СГС відбувається з використанням SOAP сервісів через HTTP з утворенням великої кількості пакетів, що завантажують мережу та під час пікових навантажень системи можуть приводити до критичних ситуацій. Також йде періодичне звертання веб-сторінки до бази даних, що також не сприяє покращенню стабільності роботи системи. У разі ж можливого подальшого горизонтального маштабування системи при теперішній її реалізації, наприклад, в результаті перенесення незалежних підсистем на окремі машини, ця проблема може стати ще більш критичною.

Спосіб усунення проблеми. Було запропоновано перепроектувати СВЗ та СГС таким чином, щоб перемістити цикл перевірки статусу виконання з СВЗ на СГС та додати елементи подієвої моделі взаємодії. Також необхідно додати аналогічні елементи з використанням технології AJAX для реалізації оновлення статусу виконання на стороні користувача. Тоді веб-сторінка після надсилання маршруту буде очікувати на сигнал, який надсилається, коли СГС отримує результати виконання цього маршруту. Результат виконання запланованого перепроектування системи полягає в тому, що буде прибрано потенційний ризик нестабільної роботи додатку під час пікових навантажень та підвищено ефективність роботи системи в цілому за рахунок зменшення часу очікування користувачем результатів виконання завдання.

Література. **1.** Петренко А.І. Дослідження архітектури комплексу схемотехнічного моделювання GridALLTED // Петренко А.І., Ладогубець В.В., Фіногенов О.Д., Булах Б.В. // Вісник університету «Україна» : Інформатика, обчислювальна техніка та кібернетика. – К.:Університет «Україна». – 2011. – № 2. – С. 65–70. **2.** Чкалов О. Особливості формування маршрутів проектування в системі GridAllted / Чкалов О., Булах Б., Безносик О., Мацулевич О. // Вісник Національного університету «Львівська політехніка», серія «Комп'ютерні науки та інформаційні технології». - №771. – Львів, 2014. – с. 77-84. **3.** Петренко А.І. BPEL-орієнтована система управління інженерними та науковими обчислювальними сценаріями / Петренко А.І., Булах Б.В. // Вісник університету «Україна»: Інформатика, обчислювальна техніка та кібернетика. – К.:Університет «Україна». – 2011. – № 2. – С.90-100.

Капшук О.О.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Системи розпізнавання людини за зображенням обличчя в IT-технологиях

Анотація: В роботі розглядаються методи і засоби біометричної ідентифікації і аутентифікації користувачів комп'ютерних систем з використанням технологій розпізнавання людини за зображенням обличчя, проблеми забезпечення надійності, хмарні технології і сервіси мобільної ідентифікації.

1.Вступ. Упродовж багатьох років в IT -технологиях використовуються засоби аутентифікації і ідентифікації користувачів (АІІК), які вважаються основою програмно-технічних засобів безпеки. Нині традиційні способи АІІК, в основі яких використовуються різні ідентифікаційні карти, ключі або унікальні дані, не є надійними в тій мірі, яка потрібна, і починають широко використовуватися технології ідентифікації і аутентифікації за допомогою біометричних даних самого користувача, які не треба тримати в пам'яті, неможливо втратити і імітація яких украй складна. Такий принцип дозволяє ідентифікувати безпосередньо людину, а не предмети, які він пред'являє, або інформацію, яку він повідомляє. Нині в біометричних системах контролю і управління доступом (СКУД), в основному, використовується багатомодальна та багатофакторна АІІК, яка вимагає застосування спеціалізованих біометричних сканерів і розробки відповідного програмного забезпечення, які гарантують необхідну надійність АІІК при високій вартості таких систем. У доповіді розглядаються питання реалізації систем АІІК з використанням технології ідентифікації людини за рисами обличчя (Face Recognition Technology, FRT). Впровадження цієї технології в системи АІІК - один з напрямів у біометричній індустрії, який динамічно розвивається і дозволяє досить просто реалізувати мобільну ідентифікацію (МІ) з використанням мобільних пристроїв користувачів (телефони, смартфони і планшети). Такі пристрої зазвичай оснащуються вбудованими відеокамерами (однією або двома), що дозволяє використовувати їх для реалізації FRT і спільно з обліковими даними цих пристроїв виконувати АІІК, що дає можливість діставати доступ до широкого спектру функцій і послуг з будь-якої точки світу і в будь-який час.

2.Проблеми надійності систем мобільної ідентифікації. Як відомо, будь-яка біометрична технологія не може гарантувати відсутність помилок АІІК, а її ефективність можна оцінити за допомогою коефіцієнтів помилкового пропуску FAR (False Acceptance Rate) і помилкової відмови в доступі FRR (False Rejection Rate), тобто в першому випадку система надає доступ незареєстрованому користувачеві, а в другому - забороняє доступ зареєстрованій в системі людині. Згідно з даними Національного Інституту Стандартів і Технології США(the National Institute of Standards and Technology (NIST)), коефіцієнт помилкового пропуску в системах розпізнавання осіб зменшувався в два рази кожні два роки, починаючи з 1993года і до кінця 2011 досяг значення 0,003%. Нині з метою підвищення надійності АІІК в стаціонарних системах контролю і управління доступом (СКУД) застосовують декілька відеокамер, інфрачервоне підсвічування і різні правила доступу : тільки особа, код і особа, карта і особа, тільки карта. Усе це дозволяє зменшити значення FAR і FRR. Так, наприклад, для біометричного сканера обличчя FaceId F710 з двома відеокамерами гарантуються значення FAR$<$ 0,0001% і FRR $<$ 1% [1].

Інший біометричний параметр людини, який може використовуватися для підвищення надійності АІІК, - це голос. Реалізація за допомогою мобільних пристроїв алгоритмів перевірки голосу не вимагає додаткових сенсорів, оскільки відповідні мікрофони є в кожному мобільному пристрої, а надійність АІІК при використанні FRT в якості основної може бути суттєво збільшена за рахунок введення додаткової перевірки.

3.Реалізація систем мобільної ідентифікації. Нині розробка систем МІ істотно спрощується за рахунок використання різних сервісів, які надаються деякими компаніями. В доповіді розглядаються можливості та застосування таких сервісів. Так компанія McAfee пропонує

кроссплатформовий сервіс McAfee LiveSafe, що захищає дані користувачів і усіх їх пристроїв, підтримує технології розпізнавання рис обличчя і голосу [2]. Сервіс надає користувачеві доступ до хмарного сховища Personal Locker. Ця функція дозволяє зберігати найбільш важливу інформацію і копії документів в Інтернеті, оперативно використовувати дані при будь-якому доступному підключенні до мережі, незалежно від місця, де знаходиться користувач і пристрою, який в даний момент у нього в руках. Сервіс виконує аутентифікацію не лише за рахунок комбінації методів розпізнавання голосу і рис обличчя, але також виконує ідентифікацію на основі PIN -коду, одноразового паролю, а також реалізованої на апаратному рівні технології Intel Identity Protection Technology (IPT). Компанія Face++ створила хмарну технологію розпізнавання обличчя людини, яка допомагає розробникам і компаніям впроваджувати більш якісні алгоритми виявлення та розпізнавання обличчя людей а також широкий перелік пошукових технологій у свої додатки і веб-сайти. Сервіс надає API, офлайн-SDK і кастомізований хмарний сервіс, як розробникам, так і бізнесу. За словами представників Face++, сервіс вже інтегрований більш ніж в 14000 застосувань, що охоплюють 40 млн. мобільних пристроїв за місяць[3].

Компанія Fujitsu працює над створенням принципово нової системи розпізнавання осіб навіть у великому потоці, яка зможе використовувати як джерело інформації відеозаписи навіть дуже низької якості [4].

Відомі пошукові компаній світу і соціальні мережі на протязі останніх 5 років (Google, “Яндекс”, Facebook, “Однокласники”) впроваджують технологію распознавания обличчя у своїх сервісах. Найбільша російська інтернет-компанія “Яндекс” впровадила технологію розпізнавання осіб Face.com. Функція розпізнавання осіб реалізована в проекті пошукової системи “Яндекс.Фотки”. Система самостійно знаходить на зображеннях людей і пропонує їх відмітити. Після того, як користувач кілька разів відмітить одну і ту ж людину на зображенні, система зможе самостійно її упізнати. Користувачі зможуть відмічати людей на своїх і чужих знімках.Аналогічна функція з червня 2011 року доступна усім користувачам найбільшої у світі соціальної мережі Facebook.

4. Висновки:

• Нині спостерігається значне зростання рівня використання біометричних технологій в мобільних і портативних пристроях. За даними звіту за 2014 рік, представленого Biometrics Research Group [5], до 2019 року кількість користувачів таких пристроїв досягне 5,5 млрд. у всьому світі.

• Використання таких пристроїв з интегрованою в них технологією розпізнавания осіб відкриває широкі можливості для мобільної ідентифікації і отримання зручного доступу до сервісів, які пропонують багато компаній.

• Використання технології розпізнавання осіб в пошукових системах і соціальних мережах розширює можливості спілкування людей, проте може погрожувати недоторканості приватного життя за рахунок несанкціонованого збору персональних даних цих користувачів і зловживання ними.

Література. **1.** Биометрический сканер лица FaceId F710 [Електронний ресурс]. – Режим доступу: http://faceid.ru/ : - Дата доступу 15.04.2015. **2.** McAfee выпустила сервис для защиты с распознаванием голоса и лица [Електронний ресурс]. – Режим доступу: http://www.km.ru/science-tech/2013/05/16/antivirusy-i-obespechenie-bezopasnosti-kompyutera/710816-mcafee-vypustila-se/: - Дата доступу 15.04.2015. **3.** Технология распознавания лиц Face++: новый капитал и коммерческое применение [Електронний ресурс]. – Режим доступу: http://apps4all.ru/post/11-14-14-tehnologiya-raspoznavaniya-lits-face-novyj-kapital-i-kommercheskoe-primenenie|/: - Дата доступу 15.04.2015. **4.** Special report mobile biometric authentication F710 [Електронний ресурс]. – Режим доступу: http://www.biometricupdate.com/201408/special-report-mobile-biometric-authentication/: - Дата доступу 15.04.2015. **5.** Special report mobile biometric authentication F710 [Електронний ресурс]. – Режим доступу: http://www.biometricupdate.com/201408/special-report-mobile-biometric-authentication/: - Дата доступу 15.04.2015.

Капшук О.О., Лугова О.Ю.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Аналіз надійності технологій розпізнавання облич в системах аутентифікації користувачів комп'ютерних систем

Біометричні системи знаходять широке використання в системах інформаційної безпеки, електронної комерції, при розкритті та запобіганні злочинів, судовій експертизі, митному контролі та т. і. Але вони вразливі до атак на різних стадіях обробки інформації. Ці атаки можливі на рівні сенсора, де сприймаєься зображення чи сигнал від індивідуума, атаки повтору (replay) на лініях комунікацій, атаки на базу даних, де зберігаються біометричні шаблони, атаки на модулі порівняння та прийняття рішень. Основну потенційну загрозу на рівні сенсора представляють атаки спуфінгу (spoofing). Спуфінг – це обман біометричних систем шляхом представлення біометричному сенсору копій, муляжів, фотографій, муляжів відбитків пальців, заздалегідь записаних звуків і т. і. Ціль атаки спуфінгу при верифікації – представлення незаконного користувача у системі як законного, а при ідентифікації – добитися невиявлення індивідуума, що знаходиться у базі даних (БД). Протидії атакам спуфінгу більш важкі, так як зловмисник безпосередньо має контакт з сенсором і неможливо використати криптографічні чи інші методи захисту.

В системах розпізнавання облич для спуфінг-атаки можна застосовувати фотографію обличчя, записане відео, 3D моделі обличчя з рухом губ, 3D моделі з різноманітними виразами обличчя і т. д. Значно більш стійкою до спуфінгу являється нова 3D-технологія 3D Hybrid Face Recognition, що враховує не тільки структуру людського обличчя, а і його форму.

В роботі проведено аналіз того, від чого залежить надійність роботи системи розпізнавання облич, і виявлено, що є сильна залежність від наступних факторів:

- Актуальності фотографії, занесеної в базу даних: це стосується зміни зовнішності - зміни кольору шкіри, наявності чи відсутності бороди та вусів, зміни поверхності шкіри після чи під час хвороби, зміни міміки;
- Величини бази даних: для підвищення ефективності розпізнавання доцільно використовувати багато зображень однієї людини (наприклад, 10);
- Якості зображення: примітно знижується вірогідність безпомилкової роботи системи, якщо людина, яку ми намагаємося розпізнати, дивиться не прямо в камеру або сфотографована при поганому освітленні. На практиці при використанні систем розпізнавання облич у складі стандартних електронних охоронних систем передбачається, що людина, яку необхідно розпізнати, дивиться прямо в камеру. Таким чином, система працює з відносно простим двомірним зображенням, що значно спрощує алгоритми і знижує інтенсивність обчислень. Але навіть в цьому випадку задача розпізнавання все ж таки не є тривіальною, оскільки алгоритми повинні приймати до уваги можливість зміни рівня освітлення, зміну виразу обличчя, наявність чи відстутність макіяжу та окулярів.
- Апаратури, що використовується (камера): технології розпізнавання облич добре працюють із стандартними відеокамерами, що передають дані та керуються персональним комп'ютером, і потребують роздільчої здатності 320x240 пікселів на дюйм при швидкості відеопотоку, принаймні, 3 – 5 кадрів за секунду. Для порівняння – прийнятна якість для відеоконференції потребує швидкості відеопотоку вже 15 кадрів за секунду. Більш висока швидкість відеопотоку при більш високій роздільчій здатності веде до покращення якості ідентифікації. При розпізнаванні облич з большої відстані існує сильна залежність між якістю відеокамери та результатом ідентифікації.

Враховуючи вищенаведені фактори, що впливають на розпізнавання облич, можна виділити наступні більш надійні методи захисту від спуфінгу, що використовуються на даний момент, та приклади їх використання:

1. Можна використовувати методи на основі виявлення емоцій, моргання та руху очей, адже вони дають змогу захиститись від фото-спуфінгу, хоч вони і не ефективні проти

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

відео-спуфінгу з використанням відео файлів, які містять в собі варіації емоцій, а також рух очей. Методи для визначення емоцій мають значно підвищену складність алгоритмів розпізнавання.
Компанія nViso, наприклад, використовує розпізнавання емоцій разом з розпізнаванням обличчя, при цьому впроваджує цей сервіс у хмарному середовищі.

2. Актуальним є метод на основі мультимодальної аутентифікації (наприклад, відбитки пальців та сканування обличчя водночас), що забезпечує більш ефективний захист, ніж методи для визначення емоцій та моргання. Водночас метод не є складним в реалізації. Цей метод використовує компанія Wired Technology Systems, яка впроваджує засоби біометричної ідентифікації, та яка є лідером в області рішень по комплексній біометрії. Wired розробила інфраструктуру управління аутентифікацією під назвою WICAM (Wired Centralized Authentication Management Suite), що передбачає повномасштабне впровадження засобів поведінкової та фізіологічної біометрії разом [1].
3. Мультифакторна аутентифікація, що може не тільки сканувати обличчя людини, а й запросити слово-пароль, наприклад. В такому випадку зловмисникам доведеться не тільки мати зображення для спуфінгу, а й знати паролі, токени та т. і.
Цю технологію впроваджено у системі Rohos Face Logon 2.9. компанії Rohos. Мультифакторна аутентифікація в цій системі передбачає захист доступу до комп'ютера, використовуючи розпізнавання обличчя та USB flash накоплювач як ключ до комп'ютера.
4. Одним з найефективніших методів для визначення «живучості» є метод на основі термограм обличчя (використовуються особливості розміщення кровоносних судин, що розташовані безпосередньо під шкірою обличчя). Метод забезпечує надзвичайну точність у визначенні живучості [2] і здатний розрізнити навіть близнюків.

В цілому, на сьогоднішній день розроблено ряд комерційних продуктів, призначених для розпізнавання облич. Лідерами на даний момент є системи Visionic, Viisage і Miros. Багато відомих виробників комп'ютерної техніки теж використовують системи розпізнавання облич. Наприклад, програма Lenovo VeriFace дозволяє у разі спроби доступу до комп'ютера особи, зображення обличчя (ЗО) якої не зареєстроване в системі контролю та управління доступом (СКУД), зробити знімок і записати час спроби несанкціонованого доступу в журналі VeriFace. Аналогічно працює і програма ASUS SmartLogon.

Висновки. Безсумнівно, системам розпізнавання облич ще є куди рухатись і є в чому вдосконалюватись. Наприклад, існує загроза спуфінгу, а також біометричні дані людини змінюються з часом, тому база шаблонів потребує постійного супроводження, що створює певні проблеми і для користувачів, і для адміністраторів. Звичайно, також потрібно захищати саму базу даних шаблонів від втручання зловмисників.

Але вже сьогодні можемо спостерігати надзвичайно високу точність розпізнавання. Наприклад, дослідники Китайської академії наук розробили систему розпізнавання облич, що здатна у натовпі виявити необхідну людину з точністю до 99,8%. Система пізнає людину з 91 різного ракурсу. Попередній рекорд точності у розпізнаванні облич знаходився на рівні 97,6% і належав американській системі [3].

Все це доводить, що майбутнє за біометричними технологіями є, і при поєднанні з іншими методами аутентифікації можна досягти надзвичайної точності розпізнавання.

Література. **1.** Multimodal biometrics beef up security, www.crn.com/news/security/18830496/multimodal-biometrics-beef-up-security.htm **2.** Thermal Imaging as a Biometrics Approach to Facial Signature Authentication, www.ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=6236169 **3.** У Китайській академії наук розробили надзвичайно точну систему розпізнавання облич, www.habrahabr.ru/company/web_payment_ru/blog/237979/.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Кирюша Б.А.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Анализ модели идеальной линии на ABCD параметрах

С ростом популярности программных продуктов с открытым исходным кодом, можно наблюдать как формирование новых подходов к решению классических задач, так и переосмысление старых. В большинстве современных коммерческих САПР моделирования СВЧ устройств на схемотехническом уровне используются либо модели в S-параметрах, либо эквивалентные схемы, для моделирования в Y-Z параметрах [1]. В пакете смешанного проектирования QUCS реализована возможность одновременного использования традиционных моделей и моделей в ABCD-параметрах, которые раньше использовались, преимущественно, для ручных расчетов [2]. Поскольку анализ моделей на Y и S параметрах уже был проведен [1], для полноты охвата типов моделей, используем возможность оценить характеристики программной модели на ABCD параметрах. В идеальном случае, модель должна иметь постоянный коэффициент передачи на всем пространстве частот и линейную зависимость фазы от частоты. Параметры частотного анализа в пакете QUCS заданы следующим образом:

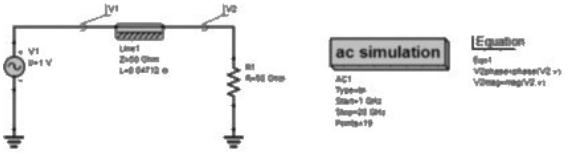

Рис. 1. Постановка задачи на моделирование

Результаты моделирования в S-параметрах (пакет Microwave studio 10.5) и ABCD-параметрах (пакет QUCS) изображены ниже.

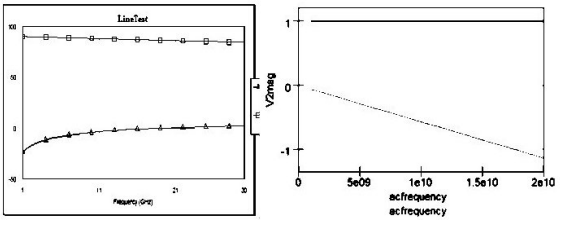

Рис. 2. Результаты моделирования по S-параметрам и ABCD/Y-параметрам

График моделирования в У-параметрах не приведен для экономии места, поскольку он совпадает с графиком в ABCD- параметрах, что подтверждено в работе [1]. Учитывая, что модель в ABCD-параметрах не требует внесения отрицательных слагаемых в элементы главной диагонали матрицы моделирования и демонстрирует полное совпадение с аналитическим представлением идеальной линий, предложенный в пакете QOCS алгоритм смешанного моделирования обеспечивает существенное превосходство над другими современными САПР по скорости и накоплению погрешности моделирования линий передач, сохраняя, при этом, возможность использовать существующие модели других типов. В качестве продолжения исследования планируется провести анализ погрешностей ABCD-моделей микрополосковых неоднородностей, по отношению к конечно-элементным моделям в трехмерном пространстве.

Литература. **1.** А.И. Петренко, Б.А. Кирюша, Проектирование СВЧ устройств с помощью ALLTED, Электроника и связь. – 2001. – № 13. – С. 3–6. **2.** Quite universal circuit simulator Technical papers, http://qucs.sourceforge.net/docs/technical/technical.pdf.

***Киселев Г.Д.*[1], *Киселева А.Г.*[2]**
[1]*Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина;* [2]*Национальный технический университет Украины "КПИ", Киев, Украина*

Управление сложными техническими системами с помощью контекстно-зависимых сервисов управления

Введение. Широкое применение сложных технических систем (СТС) состоящих из гетерогенных компонентов и включающих интегрированные в них управляющие воздействия человека требует развития методов автоматизации принятия решений. Для повышения качества управления сложными системами необходимо использовать параметры контекста таких систем, описывающего изменение во времени как состояние компонентов СТС, так и ее внешней среды. Оперативное извлечение параметров цифрового контекста, а так же анализ контекста качественной природы, слабо поддающийся аналитической обработке сокращает время принятия решений в проблемных ситуациях (ПС), а именно в аварийных и предаварийных ситуациях работы сложных технических систем.

Модели представления знаний. Формализация знаний о ПС в СТС позволяет использовать логическую дедукцию для принятия управляющих решений. Обработка знаний при таком подходе базируется на использовании интерпретаторов семантических сетей, основанных на инференциальных (дедуктивных) возможностях семантических сетей и реализующих механизмы логического вывода, механизмы поиска и анализа. Использование семантической сети для принятия решений позволяет выделять: логические отношения, каузальные отношения, отношения синонимии, зависимости и реализации. Эти отношения задают совместимость отдельных событий, порождающих возникновение аварийной и предаварийной ситуации в СТС, они позволяют строить причинно-следственные связи, процедуры проверки целостности и непротиворечивости знаний. Интеграция приведенных выше моделей представления знания осуществляется на основе онтологии. Онтология Onto – модель, определяющая и объединяющая релевантные информационно-логические и функциональные аспекты, описывающие гетерогенные источники данных контекста. Использование онтологии позволяет помещать знания на уровень, на котором они становятся понятными проектировщикам контекстно-зависимых сервисов управления (КЗСУ) для выхода из проблемных ситуаций [1].

Управление сложными техническими системами. Для реализации объектно-когнитивного подхода к управлению СТС модифицирован «треугольник предвиденья» [2]. Модифицированный треугольник предвиденья имеет следующий вид: одна из вершин треугольника предвиденья представляет дескриптивные данные субъективного представления о работе компонентов СТС. Другая вершина представляет объективные данные, которые получены при обработке контекстных параметров. Согласование или достижение релевантности субъективных и объективных данных происходит путем их преобразования в Onto. Несоответствие объективных и субъективных данных в Onto заставляет КЗСУ выполнить первую итерацию принятия решения, на которой осуществляется изменение субъективных и оценивание возможных объективных параметров контекста в Onto. Изменение объективного состояния СТС происходит после выполнения соответствующих сервисов управления. Последовательность итераций, которая выполняется на n-ом шаге преобразования онтологии Onto приводит к интерактивному выравниванию, наблюдаемому в третей вершине треугольника. Принятие правильных и своевременных управляющих воздействий в системах управления СТС поддерживается наличием необходимых параметров контекста, а также знаниями в области управления СТС, накопленными экспертами.

Литература. **1.** Kyselova A. Context data analysis for microgrid control system / Kyselova A. // Visnyk NTUU "KPI".Informatics, operation and computer science. — 2013. — №58. — P.46-50. **2.** Згуровский М. З. Системный анализ: проблемы, методология, приложения/М. З. Згуровский, Н.Д. Панкратова; НАН Украины; НТУУ "КПИ" ; ННК «ИПСА». - К.: Наукова думка, 2005. - 743 с.

Ковтун О.О.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Організація хмарних інженерних сервісно-орієнтованих обчислень (SOC) на базі пакету ALLTED

Постановка задачі. Комплекс інженерного проектування ALLTED [1] у вигляді локально встановлюваного програмного забезпечення має недолік: відсутність функціональної сумісності (interoperability), тобто користувачу необхідно обробляти вхідні дані та результати роботи ALLTED для забезпечення можливості їх обробки за допомогою інших програмних засобів. Цей недолік загалом може бути усунений шляхом впровадження програмного забезпечення як хмарних сервісів та організації на основі них сервісно-орієнтованих обчислень (SOC), у яких обмін даними між сервісами здійснюється відповідно до стандартів передачі даних [2].

ALLTED може використовуватися для організації віддалених обчислень з використанням протоколу SOAP на базі XML [3]. SOC можуть бути побудовані на основі альтернативного архітектурного принципу порівняно з SOAP – RESTful [4]. Перевага RESTful у тому, що основною концепцією є ресурси, доступ до яких здійснюється шляхом стандартних HTTP запитів, причому клієнти-споживачі сервісів можуть бути реалізовані довільними засобами, і таким чином досягається функціональна сумісність сервісів. ***Мета роботи*** – дослідити можливість організації RESTful SOC з використанням сервісів ALLTED.

Основний матеріал. Сервер та ALLTED розміщуються на обчислювальних ресурсах хмарної інфрастуктури (IaaS), яка автоматично масштабує ресурси і балансує навантаження під час виконання обчислювальних задач за допомогою сервісу. Сервер надає клієнтам RESTful програмний інтерфейс додатку (API) та складається з наступних програмних компонентів:

1. Веб-сервер, що надає RESTful API веб-клієнтам та взаємодіє з іншими компонентами відповідно до REST-запитів.
2. Адаптер вхідних даних, що перетворює дані, надіслані веб-клієнтом для обробки, з формату XML або JSON, визначеного у RESTful API, у формат вхідних даних ALLTED.
3. ALLTED – виконуваний файл, зчитує файл вхідних даних та створює файл результатів.
4. Адаптер результатів, що перетворює дані, отримані після сеансу роботи ALLTED, у формат, визначений у RESTful API веб-сервера.

Функції базового RESTful API для ALLTED надають можливість виконання асинхронних SOC шляхом обробки веб-сервером HTTP-запитів:

1. Функція передачі у форматах XML/JSON на виконання завдання ALLTED у PUT-запиті.
2. Функція перевірки стану виконання завдання у GET-запиті.
3. Функція завантаження результатів роботи ALLTED у форматах XML/JSON (повертає відповідний код стану HTTP у разі відсутності результатів) у GET-запиті.

Таким чином, ALLTED виступає у ролі набору ресурсів, на кожний з яких можна посилатися глобальним ідентифікатором (URI у HTTP) з боку веб-клієнтів або інших RESTful сервісів.

Висновки. Показано можливість створення програмних засобів інженерного проектування у вигляді хмарної SOC-системи, що використовує для обміну даними архітектурний принцип RESTful, на основі пакету ALLTED.

Література. **1.** Petrenko A. ALLTED – a computer-aided engineering system for electronic circuit design / Petrenko A., Ladogubets V., Tchkalov V., Pudlowski Z. // Melbourne: UICEE, 1997. – 205p. **2.** Tsai W. T. Introduction to Service-Oriented Computing / Tsai W. T., Chen Y., Bitter G., Miron D. // Arizona State University, 2006. **3.** Петренко А.І., Дослідження архітектури комплексу схемотехнічного моделювання GridALLTED / Петренко А.І., Ладогубець В.В., Фіногенов О.Д., Булах Б.В. // Вісник Університету "Україна". – 2011. – №2. – С. 65-70. **4.** Fielding R. T. Principled design of the modern Web architecture / Fielding R. T., Taylor R. N. // ACM Trans. Internet Techn. – 2002. – №2(2). – P. 115–150.

***Лавренюк А.М.*[1], *Лавренюк С.І.*[2]**
[1] *Фізико-технічний інститут НТУУ "КПІ", Київ, Україна;* [2] *Інститут кібернетики ім. В.М.Глушкова НАН України, Київ, Україна*

Аспекти використання мови прогамування C++ AMP для обробки великих обє'мів сейсмічних даних високопродуктивними кластерами з GPGPU

Розглянуто аспекти застосування мови програмування C++ AMP при оптимізації обчислень за допомогою GPU при обробці великих обє'мів сейсмічних даних, таких, як наприклад, 3D міграція дуплексних хвиль. Такі обє'ми даних оброблюються на високопродуктивних кластерах, які можуть містити вузли з GPU

Графічні процесори (GPGPU) вже зарекомендували себе, як обчислювальні ресурси, у різних областях, серед них і аналіз сейсмічних даних. Графічні процесори ефективні при виконанні швидких розрахунків завдячуючи їх будові, що дозволяє виконувати більше 320 обчислень за одиницю часу на одному GPU. На основі GPU вже побудовано кластери, які також доступні в грід-мережах, наприклад, CKIT-4 [1]. Так, для використання GPU компанією Nvidea розроблена мова програмування CUDA. Важливими недоліками якої є закритість стандарту - розвивається тільки однією компанією і працює тільки з обладнанням Nvidea. Окрім Nvidea є і інші виробники GPU такі як Intel, Amd у яких є своя велика частка ринку, а їх GPU мають свої переваги. Для роботи в гетерогенному середовищі, з обладананням різних виробників спільнотою розробників була створена мова програмування OpenCL (Open Computing Language) - відкритий, не потребує ліцензійних відрахувань стандарт для універсального паралельного програмування різних типів процесорів. Стандарт надає програмістам переносний та ефективний доступ до всієї мощі гетерогенних обчислювальних платформ. Проблема в використанні OpenCL в тому, що він базується на мові C, що унеможливлює використання певних конструкцій мови C++ і потребує значних часових затрат на програмування і стиковку двух частин програми - тієї, що запускається на хості (CPU) і взаємодіє з GPU і програм ядер, які виконуються уже на самому GPU. Компанія Microsoft розробила відкритий стандарт мови C++ AMP (Accelerated Massive Parallelism – ускорений масивний паралелізм) [2], що являє собою бібліотеку і невелике розширення мови, яке в сукупності дозволяє проводити гетерогенні обчислення в єдиній програмі на C++. В Visual Studio включені нові інструменти та механізми для підтримки відлагодження та профілювання програм C++ AMP, в тому числі відлагодження коду, що виконується на GPU, та візуалізація GPU-паралелізму. Завдячуючи C++ AMP розробники масових програм на C++ отримали в своє розпоряження знайомі інструменти для створення переносимих і розрахованих на майбутнє програм, що спроможні досягати вражаючого прискорення при виконанні алгоритмів, що допускають розпаралелювання по даним. Можна сказати, що C++ AMP вводить GPGPU в ужиток. Серед особливостей обробки великих обє'мів сейсмічних даних високопродуктивними кластерами з GPGPU аналогічним 3D міграції дуплексних хвиль [3] є те, що програма працює багато днів, а то і місяців і багато кластерів працює не на операційній системі Microsoft. Як CUDA так і OpenCL стабільно працюють на різних програмних платформах. Так, як C++ AMP запропонована та розроблена Microsoft, то і операційні системи Microsoft були першими на яких працювали програми C++ AMP. Але, так як стандарт C++ AMP відкритий, то вже є реалізації C++ компіляторів для гетерогенних систем на xUnix. Про аспекти використання таких компіляторів та обробки великих даних детально буде розглянуто в доповіді.

Література. **1.** CKIT-4, http://icybcluster.org.ua/index.php?lang_id=1&menu_id=5. **2.** C++ AMP, https://msdn.microsoft.com/ru-ru/library/hh265137.aspx. **3.** Мармалевський Н.Я. Міграція дуплексних хвиль - новий засіб формування сейсмічних зображень субвертикальних границь / Н.Я.Мармалевський, Ю.В.Роганов, З.В.Горняк та ін. // Зб. наук. праць. УкрДГРІ.-2007.-№2.

Петренко А.І.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Сервісно-орієнтоване дослідницьке середовище з складних обчислень

Під тиском ринку швидко зростає складність проблемно-орієнтованого програмного забезпечення, яке розробляється. Сучасні застосування вже не є незмінними, цілісними утвореннями, як це було в минулому. Це не монолітні ядра, що працюють на крупній комп'ютерній платформі, а скоріше набір динамічно змінюваних модулів, представлених бажано у вигляді веб-сервісів [1–3]. На сьогоднішньому етапі розвитку грід-технологій відбувається процес об'єднання веб-сервісів і грід- сервісів, коли веб-сервіси забезпечують ключі до грід-обчислень, а грід–середовище розглядається як розширюваний набір уніфікованих веб-сервісів.

Наші дослідження базується на використанні новітнього напрямку сервісно-орієнтованих обчислень (SOC) для відображення у вигляді веб-сервісів окремих проектних процедур (функціональних модулів), а не всього прикладного ПЗ в цілому , що забезпечує можливості модифікації і адаптації такого ПЗ, підтримки репозитарію таких веб-сервісів різними розробниками на різних мовах програмування [4]. Ця оригінальна концепція немає аналогів в світі. Для її реалізації пропонується наступне:

- Відтворити обчислювальні процедури, що використовуються при комп'ютерному моделюванні в будь-якій предметній галузі, у вигляді репозитарію уніфікованих веб-сервісів. При цьому, скоріш за все, вони не будуть співпадати з функціональними модулями існуючих монолітних прикладних програм, а потребують додаткового дослідження і програмного переоформлення. Але й такі веб-сервіси, написані на будь-якій мові програмування, мають взаємодіяти між собою на будь-якій платформі.
- Узагальнити поняття веб-сервісу, об'єднавши процедурно-орієнтовані, а також ресурсно-орієнтовані сервіси, що дозволить використовувати окремі програмні компоненти з первинними несумісними інтерфейсами.
- Перейти від функцій застосування до концепції завдання, що призведе до аспектно-орієнтованого планування потоків робіт. Разом з тим це дозволить вийти на більш високий рівень абстракції ПЗ, при якому користувач сам в інтерактивному режимі обирає маршрут математичного експерименту, тобто послідовність потрібних процедур моделювання чи обчислень, для чого створюється відповідний інтерфейс з мовою планування маршруту обчислень (рис.1).
- Використати слабку зв'язність між побудованими веб-сервісами для модифікації і адаптації проблемно-орієнтованого застосування, оскільки взаємодія між веб-сервісами в вибраній композиції не порушується кожного разу, коли змінюється дизайн або реалізація якого-небудь сервісу.
- Надати можливість створення депозитарію веб-серсівів декільком командам розробників за допомогою різних мов програмування, із застосуванням безлічі даних, які можуть надходити у режимі реального часу ("он-лайн") з декількох, як правило, географічно розподілених джерел.

На базі SOC [5]можна побудувати мережу світового масштабу слабко пов'язаних сервісів, які можуть бути без особливих зусиль скомпоновані користувачами за своїми сценаріями у гнучкі прикладні програми з динамічною структурою, що виконуються в розподіленій обчислювальній інфраструктурі. Основна наукова ідея проекту полягає в побудові інтегрованих обчислювально-інформаційних сервісів та новітнього інструментарію їх реалізації на базі об'єднання веб, грід і хмарних технологій, яка дозволяє реалізовувати відкриту динамічну архітектуру з використанням різноманітних функціональних модулів.

Взаємодіючи у мережі в певній послідовності, інтегровані сервіси ефективно реалізують той або інший обчислювальний процес. Динамічна архітектура проблемно-орієнтованого програмного забезпечення носить міждисциплінарний характер, що позитивно вплине на розвиток вітчизняної е-науки в цілому, зокрема, її грід-інфраструктури і впровадження грід-технологій

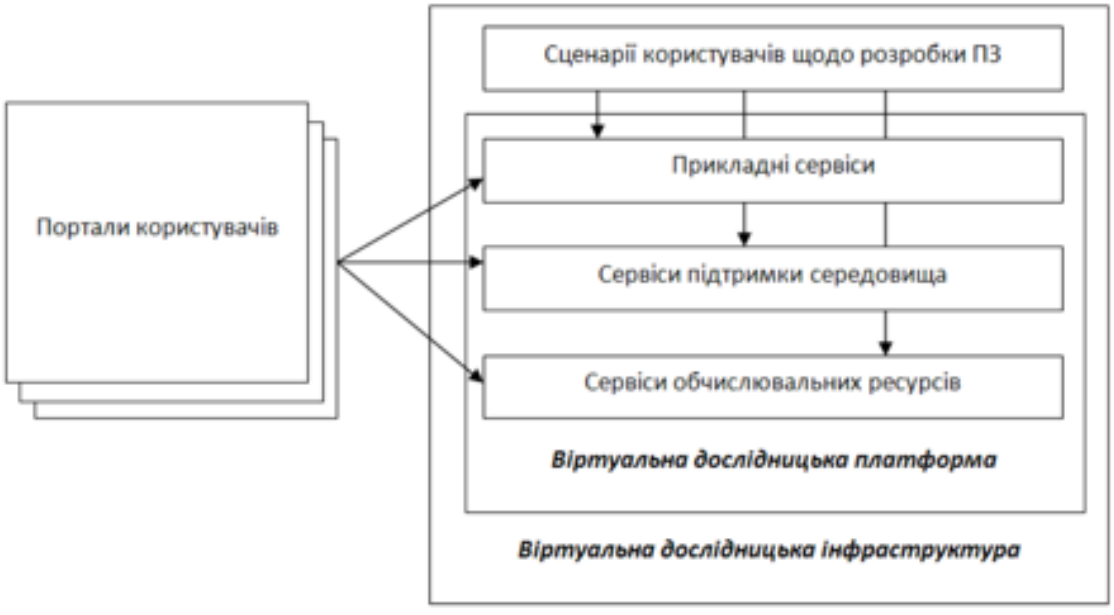

Рис. 1. Загальна архітектура SOC системи

для проектування нових технічних об'єктів і виконання математичного моделювання та обчислень при здійсненні наукових досліджень. Окрім всього, це буде перше в Україні використання об'єднаних грід та хмарних ресурсів для вирішенні практичних задач. Інноваційність підходу визнана і підтримана Європейськими країнами у вигляді спільного проекту для програми «Горизонт-2020.

Запропонований програмний комплекс призначений, в першу чергу, для задоволення потреб невеликих науково-дослідних лабораторій при виконанні складних обчислювальних експериментів під час дослідження, що проводяться ними, а також для малих і середніх підприємств в сучасному інструментарії проектування складних технічних об'єктів і технологічних процесів.

Література. **1.** Petrenko A.I. Service-Oriented Computing in a Cloud Computing Environment.-//Computer Science and Applications (USA), Volume 1, Number 6, 2014, pp. 349-358. **2.** Petrenko A.I. Virtual Research Environment "Optimal Engineering Design as a Service".-// Proc. EGI Community Forum 2014 – Helsinki, May 2014. **3.** Petrenko A.I. Service-oriented computing (SOC) in Engineering Practice.-// Proc. of- the Thirteenth Cracow Grid Workshop (CGW'13), Krakow, Poland, November 4 - 6, 2013. **4.** Petrenko A.I. Service-oriented computing (SOC) in Engineering Design.-// Proc. of Third International Conference on High Performance Computing (HPC-UA-2013), Kiev, October 2013 (http://hpc-ua.org/hpc-ua-13/files/proceedings/58.pdf) **5.** Petrenko A.I. Service-oriented computing (SOC) in a cloud computing environment.-// Proc. of East-West Design and Test Conference (EWDT-13), Rostov-on-Don, 15-18 Sept.2013.

Петренко О.О.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Наука про сервіси, управління та інжиніринг

У сучасній світовій економіці, особливо в розвинених країнах, постійно збільшується частка послуг в загальному обсязі ЗПС і в міжнародній торгівлі. Управлінці та фахівці приходять до розуміння того факту, що технології управління виробництвом і просуванням продуктів значно мірою відрізняються від управлінських технологій створення та надання послуг.

Всесвітня ініціатива компанії «IBM» з дослідження відмінних особливостей управління послугами (сервісами) отримала назву «Service Science, Management and Engineering (SSME) [1–3]. Ця наука про сервіси покликана дослідити основні принципи функціонування складних систем сервісів, шляхи створення, масштабування і вдосконалення таких систем. Індустрія послуг потребує створення своєї наукової бази, розроблення методик і інструментарію для розроблення систем сервісів, розгортання підготовки відповідних кадрів, спроможних забезпечити подальше поширення і зміцнення індустрії послуг.

Може бути проведена певна аналогія між наукою про сервіси (SSME) і комп'ютерними науками (CS). Успіх комп'ютерних наук полягає не в дослідженні фундаментальної законів природи (як у фізиці чи хімії, наприклад), а в здатності об'єднати різні дисципліни (такі як математика, електроніка, системотехніка, інформатика) для вирішення комплексних проблем, що вимагають участі всіх цих дисциплін та їх взаємодії. Наука про сервіси може бути такою ж і стати міждисциплінарним парасолькою, що дозволяє економістам, соціологам, математикам, програмістам і законодавцям співпрацювати для досягнення більш важливої мети: аналізу, побудови, управління та розвитку складних систем, які коли-небудь люди спробували побудувати.

У даній роботі розглядаються питання широкого розповсюдження систем сервісів, їх проникнення в технічні системи (зокрема, в сервісно-орієнтовані структури програмного забезпечення). Результати досліджень, отримані в рамках SSME, знаходять застосування як міждисциплінарні курси в провідних ВНЗ розвинутих країн на факультетах, де знання про технології управління послугами є актуальними (медицина, фінанси, освіта, транспорт, зв'язок, інженерія і т.д.) [4, 5].

Література. **1.** Priorities: Succeeding through Service Innovation - A Framework for Progress (http://www.ifm.eng.cam.ac.uk/ssme/) **2.** Source: Workshop and Global Survey of Service Research Leaders (IBM & IBM 2008), http://wpcarey.asu.edu/csl/knowledge/Research-Priorities.cfm **3.** Maglio P., "Service systems, service scientists, SSME, and innovation", Communications of ACM, Volume 49, Issue 7, Pages: 81 – 85, 2006. **4.** Stauss B., Engelmann K. Kremer A., Luhn A. (Eds.), "Services Science: Fundamentals, Challenges and Future Developments", Springer, Berlin 2007. **5.** Succeeding through service innovation: A service perspective for education, research, business and government, ISBN: 978-1-902546-65-0.-//University of Cambridge Institute for Manufacturing (IfM) and International Business Machines Corporation (IBM), April 2008.- 30 p.

Пикалов А.С.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Использование методов вычислительного интеллекта для повышения точности распознования изображения

Работа посвящена повышению точности распознавания движения глаз человека при съемке с веб-камеры в режиме реального времени.

На данный момент существует большое количество алгоритмов распознавания образов. Так же было разработано множество программных продуктов. которые реализуют эти алгоритмы. Примером может послужить набор библиотек OpenCV, в которую, в часности, входят библиотеки, реализующие алгоритмы распознавания, а так же средства для их обучения. Поэтому, на сегодняшний день создание простой системы распознавания какого либо объекта является вопросом выбора инструментария. Однако, при использовании существующих решений в более узкой предметной области, оказывается необходимым вносить свои модификации в существующие алгоритмы, так как качество стандартных средств является неудовлетворительным.

Недостатки существующих подходов. Использование существующих алгоритмом распознавания изображения в реальном времени позволяют с низкой вероятностью ложного распознавания находить область, в которой находится интересующий нас объект.

Целью процесса распознавания является нахождение некоторой области на изображении, внутри которой и находится искомый предмет. Зададим эту область следующими параметрами

$$C = (xc, yc)$$
$$R = (w, h)$$

Вследствии физических ограничений качества камеры, с которой мы получаем изображение, при распознавании объекта на изображении, параметры области на каждом следующем кадре будут отличаться на некоторое значение погрешности (обозначим это как dC и dR).

Основным недостатком классических алгоритмов распознавания является то, что при уменьшении линейных размеров интересующего нас объекта на изображении относительная ошибка области распознавания сильно увеличивается.

При отслеживании движения глаз, нам необходимо находить координаты зрачка в глазном яблоке человека. Проблема заключается в том, что область изображения, в которой находится зрачок, сопоставима по размерам с величиной ошибки распознавания положения зрачка. Так, при применении метода Виолы-Джонса значение относительной ошибки (отношение абсолютной ошибки распознавания зрачка к размеру области глаза на изображении) находится на уровне 10% - 30%, что является существенным для поставленной задачи.

Предлагаемая математическая модель. Схема работы системы распознавания представленна на рис. 1. Нас интересует этап фильтрации изображения. Для уменьшения значения ошибки предлагается использовать следующий способ обработки данных: Пусть

$c(t) = (x(t), y(t))$ – положение зрачка в момент времени t.

Автором предлагается значение зрачка в момент времени $t + 1$ рассчитывать следующим образом:

$$c(t + 1) = 0,5 * c[real](t + 1) + 0.5 * f(c(t), c(t - 1), \ldots, c(t - k)),$$

где

$c[real](t + 1)$ – координата зрачка, полученная в результате распознавания изображения в $t + 1$ момент времени.

$c(t), c(t - 1), \ldots, c(t - k)$ – k последних измерений.

$f(c(t), c(t - 1), \ldots, c(t - k))$ – некая функция прогнозирования $c(t + 1)$ по k прошлым значениям.

В качестве возможных вариантов предлагается рассмотреть функции следующих видов:

1) Экспоненциальное сглаживание;
2) Авторегрессия со скользящим средним АРКС(n,k).

$$f(t) = \sum_{i=1}^{n} a(i) * f(t-i) + \sum_{i=1}^{k} b(i) * avg(t-i),$$

где $avg(t)$ - функция скользящего среднего.
3) Нейронная сеть BackPropagation

Рис. 1. Схема работы системы распознавания

Выводы. Сравнительный анализ, проведенный в работе, выбрать найлучший способ для фильтрации шумов для данной задачи, что повышает качество системы распознавания

Содержание доклада. В предлагаемом докладе будут представлены различные варианты функции $f(c(t), c(t-1), \ldots, c(t-k))$, а так же сравнительный анализ качества нахождения зрачка на изображении для каждого варианта функций.

Литература. **1.** Зайченко Ю.П. Нечёткие модели и методы в интеллектуальных системах. Учебное пособие для студентов высших учебных заведений. - К.: «Издательский Дом «Слово»», 2008. - 344 с.
2. Згуровский М.З., Зайченко Ю.П. Основы вычислительного интеллекта. - К.: НВП «Видавництво «Наукова думка» НАН України», 2013. - 408 с.

Ревуцька І.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Задача розміщення елементів при проектуванні електронних схем в системах автоматизованого проектування та методи її вирішення

Вступ. Однією з найважливіших задач, яку виконують САПР при конструюванні електронних схем, є задача оптимального розміщення елементів. Критерії оптимальності можуть бути різними: мінімум сумарної довжини з'єднань; мінімум кількості з'єднань, довжина яких перевищує задану; мінімум перетинів провідників; максимальна кількість з'єднань між сусідніми елементами. У будь-якому разі, основна ціль розміщення – створення найкращих умов для подальшого трасування з'єднань. Досягти цього можна різними методами, виходячи з метричних та топологічних параметрів схеми.

Постановка задачі. В загальному вигляді задача розміщення конструктивних елементів[1,2] на комутаційній платі формулюється наступним чином. Існує множина конструктивних елементів $R = \{r_1, r_2, \ldots, r_n\}$; і множина зв'язків між ними $V = \{v_1, v_2, \ldots, v_p\}$, а також множина установочних позицій на комутаційній платі $T = \{t_1, t_2, \ldots, t_k\}$. Необхідно знайти таке відображення множини R на множину T, яке забезпечить екстремум цільової функції. В більшості випадків мінімізується загальна довжина з'єднань між елементами :

$$L = \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} c_{ij} \tag{1}$$

де d_{ij} - відстань між позиціями встановлення елементів, c_{ij} - число зв'язків між елементами x_i та x_j.

Це пов'язано з тим, що зменшення довжини з'єднань призводить до покращення електричних характеристик пристрою та полекшує процесс трасування друкованої плати. При розміщенні, елементи представляються точками для спрощення обчислень, а найбільш зв'язані елементи об'єднуються в групи, що в свою чергу виступають як один елемент.

Порівняльний аналіз алгоритмів. Всі алгоритми,які використовує САПР, є наближеними, адже при розрахунках враховується лише величина d_{ij}- відстань між центрами елементів, а не справжня довжина з'єднань, яка буде визначена лише на наступному етапі – етапі трасування з'єднань.

На сьогодні розроблено достатню кількість алгоритмів для вирішення задачі розміщення[1,2],які умовно поділяють на:

- Силові алгоритми розміщення. Елементи представляються у вигляді точок між якими діють сили тяжіння та відштовхування, інтерпретуючи зв'язки між елементами. Сили тяжіння між двома точками пропорційні кількості зв'язків між ними, сили відштовхування збільшуються зі зменшенням відстані між точками. Також вводяться сили відштовхування від країв комутаційної плати, та сили опору середовища, щоб позбавитись від незатухаючих коливань у системі. Таким чином, задача оптимального розміщення зводиться до задачі визначення статичного стану моделі механічної системи матеріальних точок.

  Перевагою такого методу є можливість отримання глобального екстремуму цільової функції, а також зведення пошуку до процедур, які вирішуються розробленими чисельними методами.

  Недоліком є важкість реалізації на ЕОМ(підбір коефіцієнтів для силових зв'язків).
- Алгоритми послідовного розміщення. В основі цих алгоритмів лежить припущення, що для отримання оптимального результату необхідно елементи, максимально з'єднані один з одним, розташовувати на сусідніх позиціях. На першому етапі алгоритму роблять початкове розміщення елементів, які штучно зміщують до країв плати. На кожному

наступному етапі для встановлення на комутаційну плату обирають елемент з числа ще не розміщених, що має максимальну ступінь зв'язаності з закріпленими елементами.
Алгоритми, що використовують послідовне закріплення елементів в позиціях, є найбільш швидкодіючими на сьогоднішній день.
Проте якістю отриманого рішення послідовні алгоритми поступаються ітераційним.

- Ітераційні алгоритми. Для покращення розв'язку в цих алгоритмах вводять ітераційний процес перестановки місцями пар елементів. Наприклад, в ітераційному методі парних замін кожен елемент міняється з кожним іншим і визначається, чи зменшилася сумарна довжина з'єднань. Кількість проб на кожній ітерації складає $n(n-1)/2$, де n- кількість елементів. Обирається заміна, яка приводить до найбільшого скорочення сумарної довжини з'єднань.
Даний алгоритм показує непогані результати розташування, проте час обрахунку пропорційний n^2, що робить застосування алгоритму на великих розмірностях недоцільним.
- Алгоритми, що використовують принцип випадкового розміщення. Цей клас алгоритмів передбачає вирішення багатокритеріальної та багатоекстремальної задачі про призначення. Прикладом такого алгоритму може бути модель квадратичного призначення.[3]
Припустимо, що деякі елементи закріплені в позиціях,тоді задача розміщення за критерієм мінімізації сумарної зваженої довжини з'єднань зводиться до мінімізації функції:

$$F(p) = 1/2 \sum_{i=1}^{n} \sum_{j=1}^{n} d_{p(i)p(j)} r_{ij} + \sum_{i=1}^{n} a_{ip(i)} \tag{2}$$

де $F(p)$- сумарна зважена довжина з'єднань елементів при довільному розміщенні; p- деяка перестановка $p = p(1), ..., p(n)$, що представляє собою довільне розміщення елементів в позиціях; $d_{p(i)p(j)} r_{ij}$ - довжина з'єднань між елементами x_i та x_j; r_{ij}- ваговий коефіцієнт, визначаючий ступінь зв'язаності елементів; $a_{ip(i)}$- сумарна зважена довжина з'єднань між елементом x_i та елементами з множини фіксованих елементів. Величину $a_{ip(i)}$ можна знайти за формулою

$$a_{ip(i)} = \sum_{s=1} r_{is} d_{p(i)s} \tag{3}$$

де $d_{p(i)s}$ - відстань між елементом, розташованим в позиції $p(i)$, та елементом x_s; x_s- елемент з множини фіксованих елементів.
Сильною стороною алгоритму є точність обчислень, а слабкою - велика кількість машинного часу, оскільки розглядається повний перебір закріплення елементів на посадочні місця.

Висновок. Для отримання найбільш оптимального результату при автоматизованому проектуванні електронних схем доцільно комбінувати запропоновані алгоритми. Наприклад, зробити початкове розміщення одним з алгоритмів послідовного розміщення та корегувати отриманий результат одним з ітераційних алгоритмів або алгоритмів, що використовують принцип випадкового розміщення, для досягнення найкращого результату.

Література. **1.** Общая постановка задачи, http://www.intuit.ru/studies/courses/650/506/lecture/11525?page=1 **2.** Классификация алгоритмов размещения, http://bigor.bmstu.ru/?cnt/?doc=010_EDA/eda081.mod/?cou=Default/020_ECAD.cou **3.** Модель квадратичного назначения, http://studopedia.ru/4_5327_model-kvadratichnogo-naznacheniya.html

***Романов В.В.*[1], *Ткаченко В.А.*[2]**
[1] *Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина;* [2] *ООО НПО "Кристалл", Киев, Украина*

О некоторых аспектах шумовых параметров полупроводниковых интегральных микросхем

Рассмотрены шумовые параметры интегральных микросхем, некоторые конструктивно-технологические способы уменьшения этих шумов для увеличения процента выхода ИМС.

Для описания шумов и их источников для различной элементной базы ИМС применяются различные модели с их временной и пространственной структурой. Для количественной оценки пользуются усредненными параметрами, определяемые на основе статистических законов. Классификацию электрических шумов проводят по различным признакам: по физической природе и механизму возникновения, по поведению спектральной плотности и плотности вероятности, с точки зрения случайных процессов, по внешнему проявлению[1].

Применительно к ИМС шумы обычно подразделяют на стационарные и нестационарные. Стационарные шумы характеризуются постоянством средних (типовых, номинальных) значений, на основе которых с определенной степенью вероятности можно установить наиболее вероятные предельные значения того или иного шума. Это позволяет характеризовать шум его спектральной плотностью по мощности, квадратом тока или квадратом напряжения. Физическим смыслом спектральной плотности S(f) является мощность, которая приходится на единицу полосы частот и выделяется на сопротивлении 1 Ом. Единица измерения спектральной плотности – B^2/Гц или A^2/Гц.

Исследование шумов, источников его происхождения преследуют ряд определенных целей: изыскание источников шумов и способов их уменьшения в радиоэлектронной аппаратуре (в т.ч. и в ИМС), способов и средств оптимального обнаружения и измерения параметров слабого сигнала на уровне (и ниже) шума; повышение точности измерений в аналоговых и цифровых устройствах обработки информации; использование шума в качестве источника информации.

Относительно методов измерения шумов, то их большое количество. Наиболее точный метод измерения - это метод сравнения с эталоном, а при расчетах – использование физических констант. При точных измерениях параметров используется эталонный генератор шума. Для усилителей низкой частоты и операционных усилителей представляет интерес типовые источники шума и предельные значения шума в диапазоне 0,1Гц – 100 кГц. В диапазоне инфранизких частот измеряется интегральный шум в микровольтах, в диапазоне низких частот измеряется спектральная плотность напряжения шума в нВ/$\sqrt{\text{Гц}}$. Обычно приводятся в справочных данных спектральная плотность на частотах 10, 100, 1000, 10000Гц, что не всегда достаточно на практике. Шумовой ток в диапазоне инфранизких частот измеряется в пА (10^{-12} А), в диапазоне 10Гц – 100000Гц измеряется спектральная плотность тока шума в фА/$\sqrt{\text{Гц}}$.

Шум твердотельных ИМС обычно принято делить на три стационарных и один нестационарный: тепловой, дробовый, фликкерный (шум 1/f), взрывной (импульсный) шум.

Тепловой шум (шум Джонсона). Тепловые колебания атомов и ионов можно рассматривать как тепловой шум – совокупность продольных и поперечных плоских волн различных частот, распространяющихся по всем направлениям. Колебания в кристалле можно рассматривать как распространение квазичастиц – фононов, т.е. фононо-термическое возбуждение. Структурные частицы колеблются около положений равновесия – узлов кристаллической решетки. Амплитуда их колебаний тем выше, чем выше температура окружающей среды T_A, При достижении некоторого критического значения решетка разрушается. При понижении T_A амплитуда колебаний уменьшается, однако по законам квантовой механики полного прекращения колебаний не происходит.

Моделирование теплового шума и его зависимость от температуры T_A, полосы частот возможно с помощью свободно подвешенного проволочного или углеродистого резистора.

Резистор генерирует на своих выводах некоторое напряжение, которое и принимается за напряжение “белого” (“чистого”) теплового шума.

Тепловая энергия может быть преобразована в другой вид энергии, в частности, в электромагнитную энергию фононого излучения. Используя закон Джоуля-Ленца можно получить формулу Найквиста: $U^2_{\text{ш.т}} = U^2_{\text{ш.г}} = U_{\text{ш.т.эфф.}} = U_{\text{ш.г.эфф.}} = \sqrt{4K * T * R_{\text{ш}} * B}$
где К- постоянная Больцмана, Т- абсолютная температура в Кельвинах, В –полоса частот в герцах.Исходя из формулы Найквиста, для уменьшения уровня шумов необходимо понижать температуру, при которой работает полупроводниковый прибор, снижать входное сопротивление приемного устройства и понижать полосу пропускания.

Дробовый шум. Электрический ток через резистор представляет собой движение дискретных зарядов. Конечность заряда ведет к статистическим флуктуациям тока. Если считать, что каждый заряд действует независимо (закон Кулона), то флуктуирующий ток определяется как: $I_{\text{ш.эфф.}} = I_{\text{ш.г}} = \sqrt{2q * I_{\text{уст}} * B}$

где q – заряд электрона, Iуст – установившееся значение тока, А, В-полоса частот измерения, Гц. Приведенное соотношение выведено из условия того, что носители зарядов действуют независимо друг от друга, но практически это не соответствует действительности, I ш.эфф. на практике меньше теоретического. Величина тока дробового шума через обратно -смещенный p-n переход определяется как $I_{\text{ш.эфф.p-n}} = \sqrt{2q * I_{\text{уст}} * B * (e^{qU/KT} + 1)}$

Флинкер-шум. или шум типа 1/f отличается от теплового и дробового шумов специфическим спектральным распределением, пропорциональным 1/f и играющим огромную роль на низких и особенно на инфранизких частотах, что имеет большое значение в интеграторах и усилителях постоянного тока. Природа этого шума до конца не изучена. Однако следует отметить, что, например, шум тока базы биполярного транзистора является шумом такого типа. Исследования проведенные в КНИИМП при создании низкошумящих операционных усилителей показали, что на шум типа 1/f огромное значение оказывают поверхностные эффекты, т.е. шумы носят релаксационный характер. Поскольку состояние поверхности ИМС можно характеризовать плотностью поверхностных состояний, то возникновение шумов 1/f определяется временем релаксации τ_i уровней энергии поверхностного состояния. Уменьшение шума 1/f должно достигаться путем уменьшения времени релаксации. При разработке в КНИИМП малошумящих операционных усилителей за счет применения специальных конструктивно-технологических мер был достигнут уровень спектральной плотности напряжения шума на частоте f = 10Гц (полоса $\Delta f = B = 1$Гц) и R_{Γ} =10 Ом, $U_{\text{ш.эфф.}} = 1 - 1,5\text{нВ}/\sqrt{\text{Гц}}$ –(типовое значение). Время релаксации $\tau_i = 1/f_i$ определяется свойствами макроскопической системы, зависит от её состояния и внешних условий и могут достигать минимальных $\tau_{i.} = 10^{-14} - 10^{-12}$ сек и до сотен секунд применительно к ИМС [2].

Взрывной (импульсный) шум. Этот вид шума не относится к стационарным видам шума. Обычно в справочных данных на ИМС он производителями микросхем не приводится, хотя фактически он существует. Образование взрывного шума можно объяснить диэлектронной рекомбинацией на межузельных атомах примеси и дислокациях. В случае диэлектронной релаксации происходит первоначально захват электрона ловушкой. После захвата электрон переходит на более низкий уровень энергии с испусканием кванта энергии, что фиксируется как импульс мощности, напряжения или тока.

Вывод. Глубокое изучение шумовых параметров ИМС способствует разработке новых конструктивно-технологических способов уменьшения этих шумов, что безусловно увеличивает процент выхода годных ИМС, разработке неразрушающих методов диагностики надежности ИМС и тем самым позволяют косвенным образом обнаружить потенциально ненадежные изделия.

Литература. **1.** П.Хоровиц, У. Хилл Искусство схемотехники, Издание 7-е,-М.: БИНОМ, 2014.-704 с. **2.** Нучно-технический отчет по НИОКР “ Крейсер 1”. КНИИМП. Киев

Харченко К.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Метод визначення відстані до об'єктів за допомогою характеристичних рис в стереозображеннях

Пропонується метод визначення параметру глибини за допомогою визначення характеристичних рис у зображеннях лівої та правої стереопари.

Постановка задачі. У відомих бібліотеках комп'ютерного зору [1] описані та реалізовані різноманітні підходи до визначення параметрів глибини сцени за допомогою стереозображень. Однак, відомі методи для розв'язання задач у реальному часі з великим розміром зображень у стереопарах потребують застосування багатопроцесорних систем, високопотужних обчислень або виконання частини розрахунків на відеокартах GPU. Для багатьох застосувань в практичних проектах для робототехніки та автомобільних систем визначення глибини зображення не завжди є найважливішим завданням тому, що головною задачею є визначення відстані до найближчого об'єкту. Зменшення вхідної інформації на етапі передобробки може суттєво вплинути на кількість обчислень, необхідних для визначення відстані до найближчого об'єкта. Оскільки відстані до об'єктів, які розташовані на більшій відстані не відіграють великої ролі у аналізі зовнішньої ситуації, то їх можна виключити з розгляду. Таким чином, розв'язання задачі пошуку найближчого об'єкта стає доступним на малопотужних вбудованих системах і т.і.

Визначення диспаритету для характеристичних рис у стереопарі. За допомогою алгоритмів визначення характеристичних рис зображення (feauture detection) є можливість за короткий час встановити точки зображення на лівій та правій стереопарі. Для цього можна застосувати добре відомі та якісно реалізовані методи SIRF, SURF, FAST або інші [2]. Надалі можна порівняти дескриптори характеристичних рис з обох зображень. Точки з однаковими дескрипторами будуть вказувати на один і той самий об'єкт у сцені. Вимірявши горизонтальну відстань від краю зображення до лівої та правої характеристичних точок можна розрахувати диспаритет відповідної точки об'єкта. Надалі диспаритет можна перерахувати у значення відстані до об'єкту, яке буде залежати від розміру стереобази, зведення оптичних осей та фокусної відстані камер. Даний підхід дозволяє визначити пару відповідних точок на стереозображенні з найменьшим диспаритетом за лінійний час від кількості характеристичних точок. Оскільки сучасні алгоритми мають налаштування параметрів визначення характеристичних рис, то метод можна гнучко налаштовувати на певні типи сцен, з якими працює система визначення відстані з застосуванням стереопари. Також метод дозволяє виділяти важливі зони зображень, обробляти зображення без втрати деталізації, виявляти відстані до невеликих об'єктів.

Оцінка швидості зближення. Для оцінки швидкості зближення з найближчим об'єктом послідовність кадрів у стереопарі можна обробляти для визначення характеристичних рис і використовувати сортування точок за величиною диспаритета. Це дозволить більш ефективно виявляти характеристичні риси у стереопарі для наступних кадрів у послідовності, за допомогою відшукання прямокутників з найімовірнішим розташуванням найближчих об'єктів у лівому та правому зображенні стереопари.

Недоліки методу. Недоліком запропонованого методу може бути наявність ефекту "сліпих" зон в сценах, де присутні однотонні об'єкти великого розміру, тому що визначення характеристичних рис в такому випадку не буде давати суттєвих значень для генерації точок, що порівнюються.

Література. **1.** Sung-Il Joo, Sun-Hee Weon, Ji-Man Hong, and Hyung-Il Choi, "Hand Detection in Depth Images Using Features of Depth Difference", http://worldcomp-proceedings.com/proc/p2013/IPC7121.pdf. **2.** OpenCV, http://opencv.org/.

Plenary talks

System analysis of complex systems of various nature

Intelligent systems for decision-making

High Performance Computing and Microsystems Engineering

Progressive information technologies

4 Progressive information technologies

Section 4

Progressive information technologies

1. Support (mathematical, algorithmic, linguistic, informational-organizational, technical, software) of the control systems, information processing, and development technologies.
2. E-commerce.
3. Information security and guarding.
4. Highly productive operating systems and networks, telecommunication technologies.
5. Data and knowledge bases as the environment of information support for control and design.

Секция 4

Прогрессивные информационные технологии

1. Обеспечение систем управления (математическое, алгоритмическое, лингвистическое, информационно-организационное, техническое, программное), обработка информации и технологии их создания.
2. Электронная коммерция.
3. Информационная безопасность и защита информации.
4. Высокопродуктивные ОС и сети, телекоммуникационные технологии.
5. Базы данных и знаний как среда информационной поддержки управления и проектирования.

Секція 4

Прогресивні інформаційні технології

1. Забезпечення систем управління (математичне, алгоритмічне, лінгвістичне, інформаційно-організаційне, технічне, програмне), обробка інформації та технології їх створення.
2. Електронна комерція.
3. Інформаційна безпека та захист інформації.
4. Високопродуктивні ОС і мережі, телекомунікаційні технології.
5. Бази даних і знань як середовище інформаційної підтримки управління та проектування.

Belevets M.S., Bidyuk P.I.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Systemic approach to estimation of operational risks

*T*he problems of financial and other risks estimation and management are highly urgent as for today due to substantial losses that emerge as a result of underestimation of the risks or inappropriate countermeasures [1]. It is especially important task to develop new mathematical approaches and techniques that will enable correct estimation of risk situations, prediction of possible losses and making quality managerial decisions.

. There exist various approaches to solving the problem dependent on specific application area. For example, the problem could be considered separately for information technologies that are hired in financial institutions, for clients of these institutions, for the methods and techniques relevant to financial operations, for influence of market etc. The systemic approach to modeling and estimation of operational risks is based on system analysis ideas that refer to hierarchical data processing schemas, modern techniques for analyzing model structure and parameters, identification and taking into consideration possible uncertainties related to data and estimation algorithms. The approach also includes the possibilities for control of relevant computational processes by appropriate set of quality criteria so that to provide for the high quality of final result. It supposes performing of analysis of internal and external influence factors to various sides of financial company activities including stochastic disturbances of different nature and types, application of statistical simulation techniques in the frames of decision support systems (DSS) constructed on the purpose. The idea of the systemic approach to risk analysis and management illustrates Fig. 1.

. The most general approach is based on general system theory and more specific techniques could be found in operations research, modern system analysis theory, cybernetics science, intellectual data analysis and data mining, and many others. Among specific analysis tools are modern theories of estimation, model building, forecasting and optimal control. Correct application of the techniques mentioned provides for development of specialized DSS that are characterized by necessary functional orthogonality and completeness as well as inherent intellectual features [2].

. The DSS proposed has features necessary for identification and taking into consideration possible uncertainties of structural, statistical and parametric type. The uncertainties are encountered in the process of preliminary data processing, model constructing, and computing forecasts. It is possible to substantially reduce their influence on the quality of final result using such modern techniques as data normalization, imputation, and filtering together with application of appropriate model structure and parameter estimation techniques. *P*rocessing statistical uncertainty. The most often met statistical uncertainties related to model development and estimation of forecasts are provoked by the following factors: measurement errors (noise) that is available practically in all cases of data collection independently on the data origin (including economy, finances and industrial control systems); stochastic external disturbances that usually negatively influence the process under study and shift the processes from desired mode (say, offshore capital transfer from some country, low quality of higher administration, unstable often changed laws, substantial corruption, local hybrid wars); missed measurements (observations) and outliers (there is necessary special literature on both subjects); multicollinearity, that requires special data processing techniques to reduce the degree of mutual correlation between separate time series.

. The most often means used to fight the measurement noise and external stochastic disturbances are digital and optimal filters (among them Kalman filter is a widely used instrument). Digital filters (DF) help to select for subsequent processing the frequency band of interest by processing the time series data with linear structures that could be represented, for example, by autoregression (AR) or autoregression with moving average (ARMA) equations. Note, that functioning of DF doesn't require the model of a process under investigation [2].

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

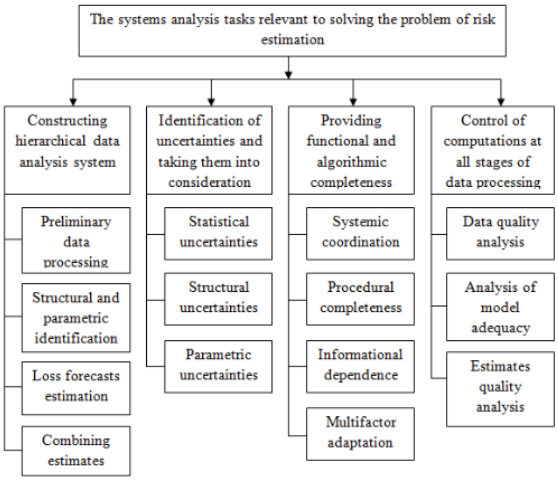

Figure 1. Tasks arising from application of the systems analysis approach to risk estimation and forecasting

. Today there is a number of Kalman filtering techniques designed for special cases. Appropriately designed adaptive Kalman filter provides a possibility for covariances estimation for stochastic disturbances and measurement noise as well as estimation of quality short-term forecasts. Optimal filter design requires model of the process (system) under study in the state space form.

. Among IAD techniques hired for risk analysis, estimation and management are static and dynamic Bayesian networks, decision trees, nonlinear Bayesian regression, special statistical data processing techniques, multivariate distributions, and fuzzy logic. Also very promising is combination of various techniques in the frames of scenario based approach.

. A key feature of DSS is in its possibilities for appropriate control of intermediate and final results. First, the quality of data should be tested with the criteria disclosing degree of information contained in data. Here variance and entropy based criteria are suitable as well as the number of derivatives that could be computed with the polynomial approximating the time series under study. As far as some observations can be lost or skipped during their collection it is also advisable to apply appropriate data imputation techniques to solve the problem of incomplete data sets. Another set of criteria should be hired for detecting nonlinearities, disclose integrated and co-integrated time series, identify heteroskedasticity etc. All these measures are directed towards improvement of a process model structure. After constructing candidate models it is necessary to test the model quality (adequacy) with appropriate set of statistics (say, determination coefficient, Durbin-Watson and Bayes-Schwartz statistics, and others dependently on a model type). Quality of forecasts are also to be determined with a set of statistical parameters among which are mean squared error, mean absolute percentage error, Theil coefficient and many others. Thus, all computational procedures in the frames of DSS are controlled with appropriate statistics what enhances substantially quality of the final result. The forecasts quality could also be improved with application of various combining techniques, such as simple averaging, weighted averaging, and optimization procedures based combining.

References. **1.** McNeil A.J., Frey R., Embrechts P. Quantitative risk management. – Princeton: Princeton University Press, 2005. – 554 p. **2.** Bidyuk P.I., Romanenko V.D., Timoshchuk O.L. Time series analysis. – Kyiv: Politechnika, NTUU KPI, 2013. – 600 p.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Boldak A.O., Yefremov K.V.
National Technical University of Ukraine "Kyiv Polytechnic Institute", Kyiv, Ukraine; World Data Center for Geoinformatics and Sustainable Development of NTUU "KPI", Kyiv, Ukraine

WDC-Ukraine's distributed platform for supporting research data life cycle

Introduction. In order to implement interdisciplinary scientific researches effectively it is necessary to create appropriate conditions for the information exchange in the process of solving scientific problems. This task is assigned to the International Council for Science World Data System (ICSU WDS) [1].

The existing WDS structures provide users with an access to heterogeneous data, accrued over a long period of time, and with the various tools and instruments for their processing and analysis [2]. However, users have to overcome some difficulties when trying to obtain the necessary data if the latter are distributed between several sources and require treatment using several tools and instruments for their processing and analysis [3].

One of the major tasks delegated to the WDC-Ukraine is to organize and support interdisciplinary researches. The work of the Centre is organized on the "Network of Networks" model. According to this model, a group of scientific institutions of the National Academy of Sciences of Ukraine, each of which conducts one or more scientific fields, has organized cooperation within the scope of WDC-Ukraine [2]. Institutions collect and provide scientific data for each scientific area to WDC-Ukraine.

Proposals and Concepts. The organization of interdisciplinary research requires the implementation of a number of data maintenance processes which basically form a continuous life cycle of scientific data [4-6] (Fig. 1).

Therefore, the task of the Center lies in technological support of the full life cycle of scientific data for a wide range of users.

To solve the mentioned problem it is necessary to develop a unified approach to implementing the data management model. This approach includes agreements on the basis of international standards for data formats, data communication protocols, and interaction of data processing services.

The introduction of this unified approach can be accomplished by:

- confining oneself to the declaration of standards / agreements and formal requirements to partners for their compliance;
- providing tools (platform) to implement the data management model by the uniform rules.

In practice, when implementing joint projects, the first way proved to be not effective enough even to provide a raw material for researches – data from partners in a compliant form.

Therefore, there was a need for developing information technology platform which would provide the ability to organize easily and quickly the support of full life cycle of scientific data (hereinafter - Platform).

To provide a minimum set of tools for organizing an interdisciplinary research, the created Platform should meet the following requirements: to provide publication and access to data; tools for partners' registration in the system as subjects – data sources; tools for forming and managing own domain-specific section (for data owners) within the common data ontology; tools for transforming data from one format to another; provision of data binding to specific research projects; data publishing tools supporting the assignment of access privileges that is controlled by a data owner; to provide Re-using stage; ontology-based search engine subsystem; tools for data selecting and merging based on search results.

Moreover, this Platform should enable users to deploy and implement data analysis methods. Data analysis methods can be divided into general and special. Special methods change to general after being approved and piloted. Ranking, bug-tracking, history of use in a variety of tasks and others are performed for each method.

This predetermines the formation of a common methodology of data analysis within interdisciplinary researches. Also an essential component of the Platform lies in the ability to design and

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

perform complex data processing sequences – scenarios (Workflow Management Subsystem).

Therefore, there is a necessity to provide information on data processing services also in the form of ontology. The history of use, applicability to a particular area, to the different types of data should be considered.

The Platform implements a cloud model within which data, services, subsystems, performing certain functionality and etc., are regarded as various-typed resources which have to be properly described and registered/published in the system. Their physical location does not matter.

Implementation. Nowadays, software tools, which implement particular elements of the required functionality, have already been developed, and Platform implementation presupposes the integration of existing systems. The scheme of Platform components' interaction is shown in Fig. 2.

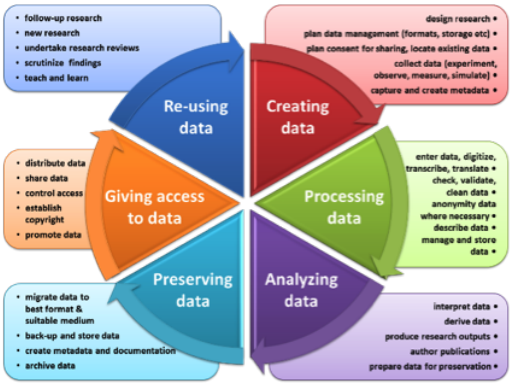

Figure 1. Scientific data lifecycle

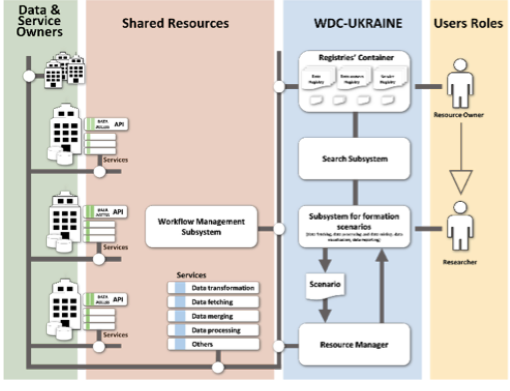

Figure 2. The scheme of Platform components' interaction

The pilot version of the Platform has been implemented by WDC-Ukraine and has been successfully applied in a number of projects [7]. The purpose of the pilot is to specify the concept of interaction of separate subsystems.

Certain data processing and data analysis services, services that are related to data visualization have been developed. A modified CKAN [8] (without the support of ontologies yet, the links have been established through metadata) is used as a registers container and search subsystem. Taverna Workflow Management System [9] has been deployed and connected. The work on automating the creation of data processing scenarios using intelligent agents is being performed.

The implementation of the Platform, using suggested approaches, allows evolving from stand-alone data owners and tools for their processing into the environment of interdisciplinary researches' organization.

References. **1.** Constitution of ICSU WDS, http://www.icsu-wds.org/organization/constitution/at_download/file1/WDS_Constitution_06.11.13.pdf [accessed on: 06.2014] **2.** Zgurovsky, M., Yefremov, K. et al., Integration of the Ukrainian Science into the WDS. Cybernetics and Systems Analysis, Vol. 46, No 2, pp. 211-219, 2010 **3.** Zgurovsky, M., Gvishiani, A., Yefremov, K., et al., Development of WDS Russian-Ukrainian segment. CODATA Data Science Journal, Vol. 12, pp. 17-26, 2013 **4.** Create and manage data, http://www.data-archive.ac.uk/create-manage/life-cycle/ [accessed on: 10.2014] **5.** Shearer, C., The CRISP-DM model: the new blueprint for data mining. J Data Warehousing, Vol. 5, pp. 13-22, 2000 **6.** Rohanizadeh, S. S., Moghadam, M. B., A Proposed Data Mining Methodology and its Application to Industrial Procedures. Journal of Industrial Engineering, Vol. 4, pp. 37-50, 2009 **7.** Zgurovsky, M., Boldak, A., Yefremov, K., Intelligent analysis and the systemic adjustment of scientific data in interdisciplinary research. Cybernetics and Systems Analysis, Vol. 49, No 4, pp. 541-552, 2013 **8.** CKAN data management system, http://ckan.org/ [accessed on: 10.2014] **9.** Taverna Workflow Management System, http://www.taverna.org.uk/ [accessed on: 10.2014]

Kachko N.A.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Prime table generation

There are an infinite set of numbers, each of which meets the next criteria: *an odd natural number k such that* $k2^n + 1$ *is composite, for all natural numbers n* [1]. Every number with this property is called "*Sierpinski number*".

The major challenge defining number Sierpinski is to test some particular number on being prime. Considering the fact those numbers might be infinitely large, time consumed on this test should be negligible small. Hereby the problem arises: how to test a number on being prime using as less time as possible?

The integer factorization problem is the computational problem of determining the prime factorization of a given integer. Phrased as a decision problem, it is the problem of deciding whether the input has a factor less than k [3]. The very same deciding method referring in the article "Primality Test Problem" [2] could be used in this particular case. It should be modified in order to use capabilities of computer computation distribution, thus decreasing time used for prime look-up table generation. Task distribution is a generic method to achieve better performance over limited computing power of single unit.

To ensure that each node of cluster takes approximately equal amount of data to compute, the next algorithm of distribution is proposed:

1. Each node receives index "k" and upper number bound to test from root node
2. The "k"-th node uses formulae $i * N + k$ to pick the next number to test (where i - current local iterator, N – number of all nodes)
3. The "k"-th node make a local table of primes using probabilistic method
4. The root node receives and merges tables from all nodes into the general one
5. The root checks the last number's index corresponding with credible source

The example of distribution to three nodes is shown on the Figure 1:

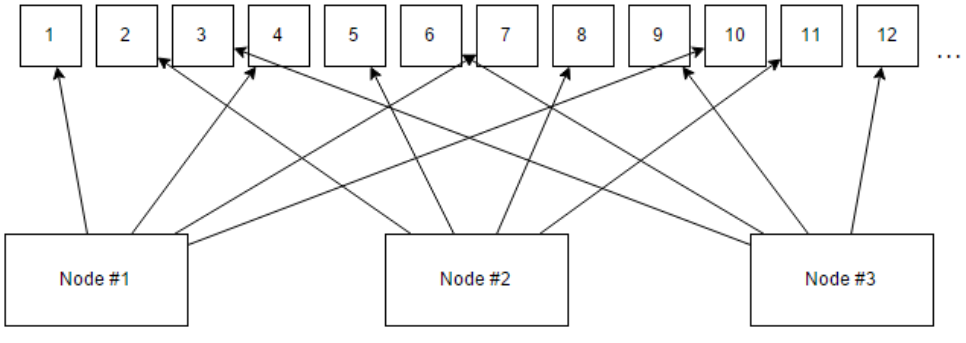

Figure 1. Example of nodes distribution

In the proposed way the system will work as a set computers combined together. With proper linkage and solid amount of numbers set the node-to-node transportation costs will be negligible small. The outcome of proposed method is a potentially infinite look-up table fulfilled with prime numbers that can be used in various tasks, Sierpinski number test in particular.

References. **1.** Wikipedia, the free encyclopedia (2015, Feb 17). *Sierpinski number* [Online]. Available: http://en.wikipedia.org/wiki/Sierpinski_number **2.** N. Kachko and V. Yaremenko, "Primality Test Problem," unpublished. **3.** Wikipedia, the free encyclopedia (2015, Apr 14). *Computational complexity theory* [Online]. Available: http://en.wikipedia.org/wiki/Computational_complexity_theory

Kyrysha B.A., Kazachenko O.D.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Authorization and Accounting for Distributed Systems

The authentication problem in computer networks is really significant nowadays. Authentication is often only a step to control that the individual making a request is authorized to do so and provide correct access to operations. There are few models of authentication for distributed systems.

Proxy-based model. Distributed authorization can be easily supported by providing restricted proxies support .A proxy is a token that allows one to operate with the rights and privileges of the principal that it granted. Naturally, a principal with the credentials needed to authenticate itself can often grant a proxy to another principal simply by passing on those credentials. A restricted proxy is a proxy that has had conditions placed on its use. The end-server can verify that the restrictions have not been falsified. Due to a specified restrictions proxy server can be used only in allowed principal or with restriction of operations that may be performed. Restricted proxies provide the vehicle for implementing a wide range of authorization mechanisms in distributed systems. An authorization server based on using restricted proxies does not directly determine that a particular principal is authorized to access a particular object. Instead, when requested by an authorized client, server grants a restricted proxy allowing him to act as the authorization server to have rights for access.

Security agent model. The security agent based authorization model has the following elements: Security Management Authority (SMA), Security Management Component (SMC), Object Managers (OM) and Security Agents (SA). All resources that need a protection for its objects has a SMC and each domain has an SMA. Both SMCs and SMA are secure entities and they are involved in the creation and distribution of the SAs. OMs are in charge of the management of the objects that located in the host. OMs use the SMC to generate and ratify the SAs. The SMC also maintains security information such as public and private keys as well as a list of some of the commonly used certificates and name servers. The OM defines the privileges of an object at the time of objects creation and these are passed to the SMC which then encyst them in a SA and shared with clients. A client request passes along with the SA with additional agent methods. The OM at the target interacts with the SMC to control access. All unauthorized repeated use of the SA are defined in duplication list. SA should be unforgeable and make only those decisions for which it has been allowed to do so.

Logic-based model. There was invented a logic for authentication and access control in distributed systems. Later also was designed and implemented a security system based on this logic. The basic concepts of SRC logic are principals and statements. Principals make statements. Principals include: simple principals, channels, groups, conjunction of principals, principals quoting principals, principals in roles, principals acting on behalf of others. However, as a language for authentication and access control, SRC logic is rather limited in several aspects. First, primitive statements cannot state properties of principals for use in the logic. Second, SRC logic does not have variables. It leads to expression a delegation to any principal that has a certain property, without explicitly listing all the principals. Third, every delegation can be freely re-delegated. Even with the above limitations, validity in SRC logic is undecidable in general except two classes. One class is worst-case exponential-time solvable. The other class is the result of even further simplification and is computationally tractable.

References. **1.** W.Farmer, J.Guttman and V.Swarup, "Security for Mobile Agents: Authentication and State Appraisal", Proc. of the European Symposium on Research in Computer Security, 1996. **2.** V. Varadharajan, P. Allen, S. Black. "An Analysis of Proxy Problem in Distributed Systems". Proceedings of the Symposium on Security and Privacy, pp 255-275. IEEE, 1991. **3.** IEEE, www.mit.edu/~bgrosof/paps/delegation-logic-tissec.pdf. **4.** B. Cliford Neuman. Proxy-based authorization and accounting for distributed systems. Technical Report 91-02-01, Department of Computer Sci-ence and Engineering, UniversityofWashington, March 1991.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Litvinov V.A., Maistrenko S.Ya., Khurtsilava K.V.
Institute of Problems of Mathematical Machines and Systems of NAS of Ukraine, Kyiv, Ukraine

Regarding quality of the basic vocabulary of the subject-oriented spellchecking system and criteria of its evaluation

The basic vocabulary (BV) of the admissible ("correct") words is the basis of the automated system of the user error identification in the natural language words (generally in the irregular alphabetical and digital codes). BV quality is primarily determined by two factors:

- ability to detect the maximum amount of the most probable (typical) mistakes;
- capability to generate the minimum amount of the false error notifications.

The first factor can be estimated by the relative quantity ρ of the BV words, which have been distorted by certain, typical mistakes (k - distortions), and which have coincided with real admissible words.

The second factor can be assessed by the sum probability π of the absence of the required word in BV.

As described in [1,2], diminution of the vocabulary volume, with other equal conditions, leads to lesser ρ due to increase of the relative redundancy of the word presentation, and accordingly due to diminution of the probability of the accidental coincidences of the mistaken words with the admissible ones. On the other hand, exclusion from the vocabulary of the words with non-zero request probability increases the value of π.

Simulation results for a number of the Russian and Ukrainian dictionaries, as described in [1,2], have been obtained by the use of the simple criterion of the "brute force" – i.e. deletion of the word A_j with minimum value of the request probability p_j . Through simplicity of this criterion, these results more likely illustrate a possible existence of the task of BV improvement by eliminating words of the lesser value, rather than resolution of this task.

While identifying the "pinpoint" criterion, which assesses the concrete contribution of the potentially deletable words to the values of the quality factors, one should take into account the following considerations:

1) Elimination of a neutral word, k - distortion of which does not cause coincidences with real words, - does not diminish the value of ρ , but increases the value of π.

2) For each word A_l, direct k - distortions of which cause coincidences with words A_s, there are inverse distortions of words A_s, coinciding with a word A_l.

3) Exclusion of a word A_l changes distribution of the consequences of the possible k -distortions in the remaining part of BV (in particular, that can turn into neutral words the related words A_s).

The above said qualitative considerations sum up the following criterion, which can serve as a foundation for the step-by-step task-solving algorithm:

$$A_l = A\left(\min_j \frac{p_l}{p_l \sum_k P_k \cdot \frac{{v_{kl}}^s}{V_{kl}} + \sum_s p^s \sum_k P_k \cdot \frac{1}{V_{ks}}}\right) \quad (1)$$

where P_k - is a relative quantity of k - distortions of BV words,

${v_{kl}}^s$ - is quantity of coincidences with BV words of the k - distortions of the word A_l.

V_{kl}, V_{ks} - is total quantity of all possible k- distortions of the word A_l, words A_s.

After exclusion of the word A_l the values of ${v_{kl}}^s$, p^s, V_{ks} are re-determined.

References. **1.** В. А. Литвинов, С. Я. Майстренко, К. В. Хурцилава, К оценке качества базового словаря в системе автоматического обнаружения ошибок пользователя и возможностей его улучшения // Материалы XIV международной научной конференции им. Т. А. Таран «Интеллектуальный анализ информации».-Киев.-2014.-14-16 мая.- С. 139-143. **2.** В. А. Литвинов, С. Я. Майстренко, К. В. Хурцилава, Оценка контролирующих свойств базового словаря допустимых слов в системе автоматического обнаружения ошибок пользователя // Математичні машини і системи.-2014.-№2.- С. 65-70.

Piddubna Yu.I., Bidyuk P.I.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Systemic approach to estimation of operational risks

A systemic approach is proposed for modeling and estimation of operational risks using statistical data.

Operational risks include many different types of risk, from the simple 'operations' risks of transactions processing, unauthorized activities, and system risks to other types of risk that are not included in market or credit risk: human risk, legal risk, information risk, and reputational risk.

The major problem with any model for operational risk is that these data are inadequate. For example:

- Internal loss event data for low frequency high impact risks such as fraud may be too incomplete to estimate an extreme value distribution for measuring the tail loss. But augmenting the database with external data may not be appropriate;
- Operating costs have a tenuous a relationship with operational loss. So the proportional charges that regulators are considering for operational risk, that are based on a fixed percentage of operating costs, may be very inaccurate;
- Internal risk ratings are based on assessments of the size and frequency of operational losses from the different activities in a business unit. These data are likely to be inaccurate because they lack objectivity.
- Regression models of operational risk that are based on the CAPM or APT framework produce betas that are based on many subjective choices for the data.[2]

The idea of the systemic approach to risk analysis and management illustrates Fig. 1

The most general approach is based on general system theory and more specific techniques could be found in operations research, modern system analysis theory, cybernetics science, intellectual data analysis and data mining, and many others. Among specific analysis tools are modern theories of estimation, model building, forecasting and optimal control. Correct application of the techniques mentioned provides for development of specialized DSS that are characterized by necessary functional orthogonality and completeness as well as inherent intellectual features.

The DSS proposed for risk estimation has features necessary for identification and taking into consideration possible uncertainties of structural, statistical and parametric type. The uncertainties are encountered in the process of preliminary data processing, model constructing, and computing forecasts. It is possible to substantially reduce their influence on the quality of final result using such modern techniques as data normalization, imputation, and filtering together with application of appropriate model structure and parameter estimation techniques. [1]

Processing statistical uncertainty. The most often met statistical uncertainties related to model development and estimation of forecasts are provoked by the following factors: measurement errors (noise) that is available practically in all cases of data collection independently on the data origin (including economy, finances and industrial control systems); stochastic external disturbances that usually negatively influence the process under study and shift the processes from desired mode (say, offshore capital transfer from some country, low quality of higher administration, unstable often changed laws, substantial corruption, local hybrid wars); missed measurements (observations) and outliers (there is necessary special literature on both subjects); multicollinearity, that requires special data processing techniques to reduce the degree of mutual correlation between separate time series.

The most often means used to fight the measurement noise and external stochastic disturbances are digital and optimal filters (among them Kalman filter is a widely used instrument). Digital filters (DF) help to select for subsequent processing the frequency band of interest by processing the time series data with linear structures that could be represented, for example, by autoregression (AR) or autoregression with moving average (ARMA) equations. Note, that functioning of DF doesn't require the model of a process under investigation.

During the last few years there has been increasing interest in the use of causal networks to model operational risks, and in particular to use Bayesian belief networks (BBNs). Reasons for this surge of interest include:

- A BBN describes the factors that are thought to influence operational risk, thus providing explicit incentives for behavioural modifications;
- BBNs have the ability to perform scenario analysis to measure maximum operational loss, and to integrate operational risk with market and credit risk;
- BBNs have applications to a wide variety of operational risks. Over and above the categories where available data enable the modelling of operational risk with standard statistical models, BBNs have applications to areas where data are more difficult to quantify, such as human risks;
- Augmenting a BBN with decision nodes and utilities improves transparency for management decisions. [2]

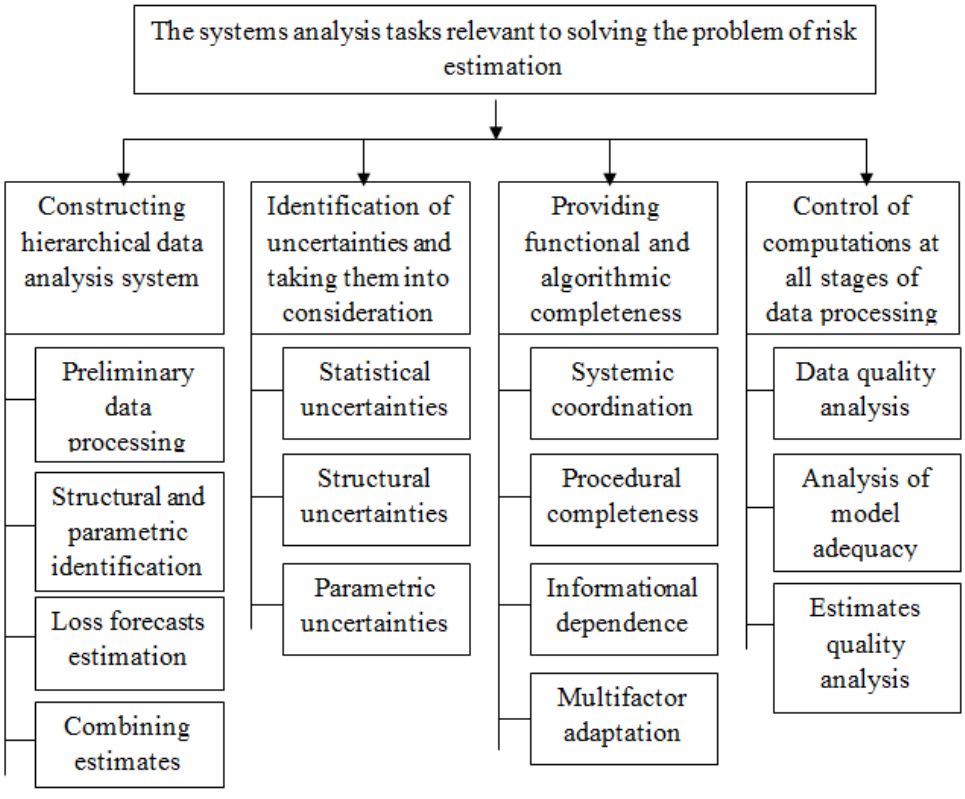

Figure 1. Tasks arising from application of the systems analysis approach to risk estimation and forecasting

Among IAD techniques hired for risk analysis, estimation and management are static and dynamic Bayesian networks, decision trees, nonlinear Bayesian regression, special statistical data processing techniques, multivariate distributions, and fuzzy logic. Also very promising is combination of various techniques in the frames of scenario based approach.

References. **1.** McNeol A.J., Frey R., Embrechts P. “Quaytitative risk mvnagement.” – Princetin: Princeton Uniaersitn Press, 2005. – 554 p. **2.** CaroC Alexander. “Bayesian Methods fop oeasuring Operatronal Risks” — ISMA lentre. The Univeisity Mf Reading, 2000. — 22p.

Yaremenko V.S.
Institute for Applied System Analysis of NTUU "KPI", Kyiv, Ukraine

Modelling of semantic Web applications based on annotated REST Web services

Semantic web is a collective movement towards adding semantics or meaning to the data available on the internet thus making them machine readable. Data with semantics will be key in the future web where human interaction can be reduced for exploring and using this information [1]. So, it is very actual problem to improve usage of semantic Web applications.

The basic idea is that the semantic orchestration engine with help of ontology decides in what order and in what format the data must pass through a variety of REST services to get the desired result. The proposed model of semantic Web application consists of 4 main parts: annotated REST services, ontology, task manager and semantic orchestration engine (figure 1).

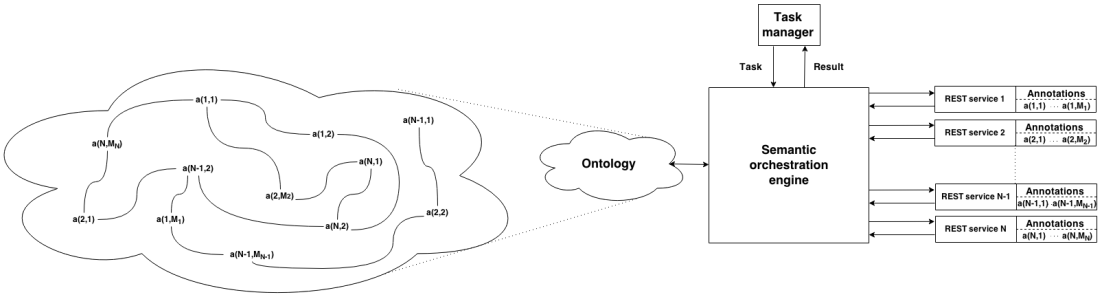

Figure 1. Semantic Web application connections and example of ontology based on annotated REST Web services

Each REST service can solve its own type of problem and returns a result in a specified format. Also, each service has annotations that describe its work, input data type, output data type, etc. E.g. REST service 1 has M1 annotations: a(1,1), a(1,2), ..., a(1,M1); REST service 2 has M2 annotations and so on.

Ontology is based on annotations from all REST services and semantically connections between them (figure 1). E.g. a(1,1) has connections with a(1,2), a(2,M2) and a(N,MN) annotations. Task manager (human, other service or application) prepares a problem for solving using the semantic WEB application and transfer it to this application. A task should be processed by in such a way as to comply with application's entry. The main part of the proposed model is semantic orchestration engine. After receiving a task this application starts the analyzing process. It analyzes annotations of inputted task, than analyzes an ontology and decides how the problem should be solved. It determines the route of data (which services will solve this problem and in what order) to get the desired result. The application processes inputted data and transfer a result to one of the REST services. After processing is finished this service transfers the data to semantic Web application. It determines whether to continue the solving process or the solution is being found. If the problem is solved, the process finishes. Otherwise the application processes the received data and prepares it for transferring to the next REST service and so on.

When used correctly this model, the semantic Web application can solve various kinds of problems. Also, when the ontology is changed by another one and when other types of REST services are being connected, this application is able to switch easily between the solutions of the problems of various kinds (e.g., engineering, medicine, weather forecast, etc.).

References. 1. Davis John, Rajasree M. S., "RESTDoc: Describe, Discover and Compose RESTful Semantic Web Services using Annotated Documentations" in *International Journal of Web & Semantic Technology (IJWesT)*, Vol.4, No.1, January 2013.

Акінфієва А.О.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Створення бази знань підсистеми контролю успішності студентів вузу

База знань - частина експертної системи, що містить подання знань, яке стосується визначеної предметної області.

Система, заснована на знаннях - це система програмного забезпечення, основними структурними елементами якої є бази знань, яка спеціалізується для накопичення, збереження, пошуку та демонстрації знань,і механізм (машина) логічних виведень. [4]

У системі, заснованій на знаннях, знання подаються у конкретній формі в базі знань, що дозволяє їх легко визначати, модифікувати і поповнювати. Функції вирішення задач реалізуються самостійним механізмом логічних висновків, що працюють на основі знань, які зберігаються у базі.[2]

Інтелектуальна система контролю успішності є комплексом програмних, лінгвістичних та логіко-математичних засобів для реалізації підтримки діалогу з системою дистанційного навчання. Такий режим спілкування з інформаційною системою є режимом набуття знань.[1] Тобто, створення та правильна побудова спеціалізованої бази знань значно вплине на всю структуру інформаційної системи.

Таким чином, досі актуальною є проблема адекватної реалізації бази знань та її роботи. Основною задачею, що має бути вирішена в роботі над нею, це: дослідження існуючих рішень структурування інформації у базах знань та їх ефективності з точки зору викладача і студента, розробка сервіс-орієнтованого прототипу для систем дистанційного навчання, який дозволить виконувати швидкий пошук по курсам, матеріалам, а також відстежувати успішність студентів.

База знань включатиме в себе оцінки та будь-які коментарі до кожної з них. Тоді ось що отримуємо. По-перше, у нас можуть бути додатки, які в зручному вигляді дозволятимуть змінювати та додавати бали за курс. По-друге, і що найбільш важливо, з'явиться можливість створювати програми, що будуть: 1) аналізувати інформацію про студента (оцінки та коментарі) і будувати персональний план для студентів; 2) аналізувати інформацію про предмет та давати поради по його поліпшенню. Це і є основним призначенням навчальних експертних систем - аналіз знань студентів за визначеними предметами, відшукування пробілів у знаннях і запропонування засобів їх ліквідації.Також вона виконуватиме функціїї системи контролю - забечпечуватиме адаптивне керування поведінкою складних людинно-машинних систем, прогнозуючи появу можливих збоїв і плануючи дії, необхідні для їх попередження. [3]

Така система підконтролю успішності буде динамічною (тобто, допускатиме зміни даних і знань у процесі рішення) та досить простою (мала витратінсть ресурсів).

Висновок. У підсумку отримуємо, що побудова інформаційної підсистеми на основі баз знань дозволяє створювати нові методи, способи та технології організації гнучкого і різнобічного контролю успішності студентів вузу. Перевагами даної реалізації є забезпечення постійного доступу до процесу навчання та актуальність набутої студентами інформації, стійкість та відтворюваність результатів.

Література. **1.** Н. Б. Шаховська,В. В. Пасічник, "Сховища та простори даних", 2009. **2.** П.І. Федорук, Н.М. Дяків, "Технологія оцінки ефективності роботи баз знань», 2004. **3.** Т.А. Гаврилова, "Бази знань інтелектуальних систем", 2000. **4.** Субботін С. О., "Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень: Навчальний посібник.",2008

Артюхов В.Г., Бритов О.А.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Розрахунок цифрових еліптичних фільтрів в Micro-Cap

Наведені формули, рекомендації по організації розрахунку цифрових еліптичних фільтрів в середовищі популярної програми схемотехнічного моделювання Micro-Cap.

Донедавна для розрахунку аналогових та цифрових фільтрів використовувовались різні програми. Сучасна потреба в проектуванні змішаних аналого-цифрових систем вимагає об'єднання цих розрахунків в одному середовищі. Поширені програми схемотехнічного моделювання включають елементи, які дозволяють моделювати як аналогові, так і цифрові фільтри [1].

Крім того, програма Micro-Cap [2] включає модуль проектування аналогових фільтрів, який дозволяє отримувати передаточну функцію нормованого аналогового фільтру. Як показано в нашій попередній роботі [3], цей фільтр може бути використаний як прототип для розрахунку цифрового фільтру з заданими параметрами АЧХ.

Формули розрахунку коефіцієнтів передавальної функції цифрового фільтру були виведені для фільтрів Батерворта та Чебишева. В даній роботі наводяться формули для еліптичних фільтрів (фільтрів Золотарева-Кауера). Вбудований в Micro-Cap модуль проектування аналогових фільтрів видає передавальну функцію нормованого прототипу у вигляді добутку секцій першого та другого порядку, які мають вигляд:

$$H_j(U) = k_{0j}\frac{1}{U + b_{0j}}; H_j(U) = k_{0j}\frac{U^2 + a_{0j}}{U^2 + b_{1j}U + b_{0j}}$$

Причому форма передавальної функції не залежить від виду АЧХ. Завдяки цьому формули для розрахунку коефіцієнтів передавальної функції ЦФ будуть однакові для всіх чотирьох типів АЧХ (низьких частот, високих частот, смугових та режекторних).

Формули для розрахунку коефіцієнтів передавальної функції цифрового фільтру отримуються шляхом використання білінійного перетворення з компенсацією спотворення частот [4]. Обчислення програмуються у вигляді директив .define, розміщених на листі схеми. Формули для розрахунку секцій еліптичних ЦФ другого порядку наведені на рис.1 в складі макровизначення секції ЦФ.

Для практичного використання зручно всі обчислення виконати у вигляді макровизначення секції цифрового еліптичного фільтру. Таке макровизначення включає схему релізації секції ЦФ. Реалізація може виконуватись з використанням керованих джерел з заданою передвальною функцією [3], з використанням примітивів, включених в бібліотеку Micro-Cap, або з використання примітивів користувача. При реалізації можна врахувати точність представлення коефіцієнтів, як показано в [1].

Приклад макровизначення секції еліптичного ЦФ з формулами для розрахунку коефіцієнтів наведений на рис. 1. Це макровизначення виконане на основі бібліотечних примітивів суматора, множника та блока затримки і ралізує канонічну структуру секції.

Механізм макровизначень дозволяє задавати параметри, такі як частота дискретизації, характеристична частота фільтра, параметри передавальної функції прототипу. Чисельні значення цих параметрів задаються під час інстанціювання макровизначення в схемі у вікні параметрів компоненту, приклад якого наведено на рис.2. Параметри нормованогоо аналогового фільтру-проторипу розраховуються модулем Design симулятора Micro-Cap і виводяться цим модулем у вигляді директиви .define. Наприклад, для полосового фільтру ця директива має вигляд:

.define BP (1.713)*(U*U+0.93395)/(U*U+0.0149*U+0.96575)*
(1.602)*(U*U+1.0707)/(U*U+0.015429*U+1.0355)*(0.11269)

Частотний аналіз дозволяє отримати частотні характеристики ЦФ, а аналіз перехідних процесів забезпечує отримання імпульсної та перехідної характеристик, а також реакції ЦФ на довільні вхідні сигнали.

```
Recursive low pass digital elliptic filter second order section
canonical implementation based on
Components-Digital Primitives-Digital Filter Macros.
parameters (FS=1, fc=0.1, ka0, aa0=1, ba0=1, ba1=1)
help FS "Sampling frequency."
help fc  "Cutoff frequency. The same units as FS."
help ka0 "Analog prototype scaling coefficient"
help aa0 "Analog prototype numerator coefficient"
help ba0 "Analog prototype denominator coefficient"
help ba0 "Analog prototype denominator coefficient"
```

```
.define yc cos(2*pi*fc/Fs)
.define ys sin(2*pi*fc/Fs)
.define k1 (1+ba0)+yc*(1-ba0)
.define k2 (1-ba0)+yc*(1+ba0)
.define B2 (1-ba1*ys/k1)/(1+ba1*ys/k1)
.define B1 -(k2/k1)*(1+B2)
.define A0 (ka0*(1-yc)*(1+B2)/2/k1)*(1+yc+aa0*(1-yc))
.define A1 2*((1-yc)*aa0-(1+yc))/(1+yc+(1-yc)*aa0)
```

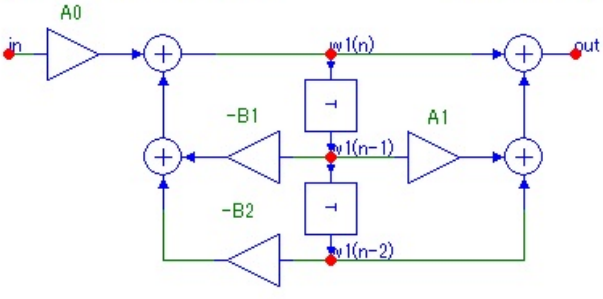

Рис. 1. Макровизначення секції другого порядку.

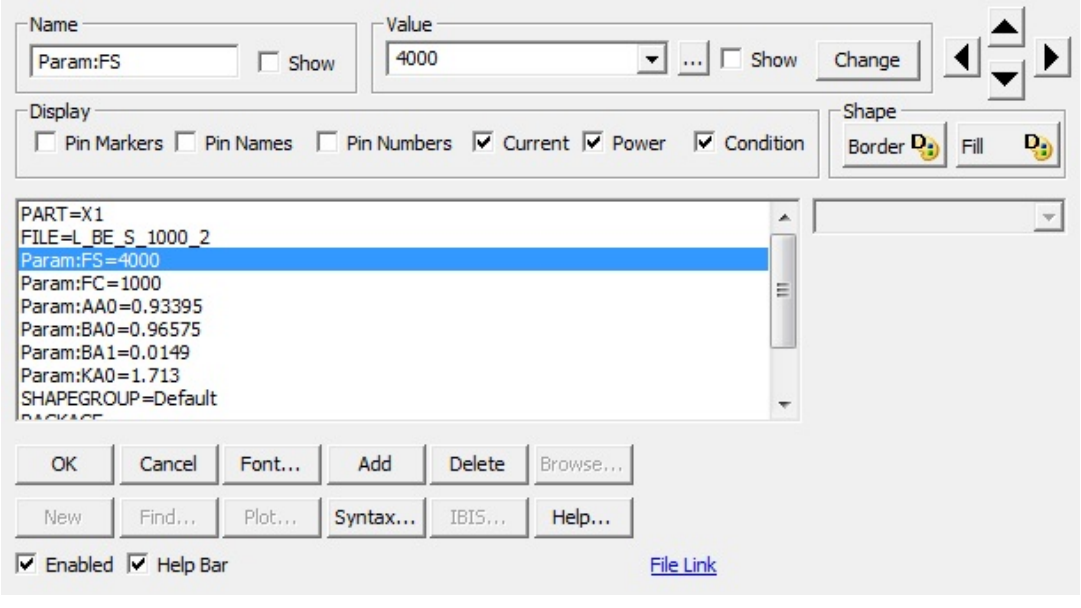

Рис. 2. Приклад задачі параметрів секції другого порядку.

Література. **1.** Dalibor Biolek, Inas Faisel Abuetwirat, "Analysis of digital filters via SPICE-family programs", http://www.elektrorevue.cz/clanky/06026/english.htm, дата доступу 16.03.2015. **2.** М.А.Амелина, С.А.Амелин, Программа схемотехнического моделирования Micro-Cap. Версии 9, 10.-Смоленск, Смоленский филиал НИУ МЭИ, 2012.-617 с., ил. **3.** В.Г.Артюхов, А.А.Бритов, Цифрові фільтри в Micro-Cap/ Системний аналіз та інформаційні технології: Матеріали 16-ї Міжнародної науково-технічної конференції SAIT 2014, Kyiv, 26-30 травня/ ННК "ІПСА" НТУУ "КПІ", 2014 – с.333-334. **4.** А.Пелед, Б.Лиу, Цифровая обработка сигналов. Теория, проектирование, реализация. Пер. с англ. – Киев: Вища Школа. Головное издательство, 1979. – 264 с., ил.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Архангельский К.Е.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Алгоритмы компьютерного зрения в системах дополненной реальности

Алгоритмы компьютерного зрения в системах дополненной реальности

В данный момент устаревают традиционные графические интерфейсы, управляемые клавиатурой и мышью.

Стремительное развитие интерактивных мультимедийных технологий требует появления новейших интерфейсов человеко-машинного взаимодействия.

Данные интерфейсы не используют привычное графические меню, формы или панели инструментов, они опираются на методы взаимодействия, присущие человеку, т.е. вместо традиционных средств управления используются жесты [1].

Одним из примеров новейших интерфейсов взаимодействия человека и компьютера является технология, получившая название «дополненная реальность» (от англ. augmented reality).

В парадигме привязки технологии к интерфейсу взаимодействия человека и компьютера, дополненную реальность в большинстве случаев представляют как процесс совмещения объектов реального мира и объектов, сгенерированных компьютером. Большинство исследований в области дополненной реальности сконцентрировано на использовании живого или интерактивного видео, подвергнутого цифровой обработке, «дополненного» компьютерной графикой [2].

Исследования включают отслеживание движения реальных объектов, распознавание координатных меток при помощи машинного зрения и конструирование управляемого окружения.

Дополненная реальность отличается от виртуальной. Их различие в том, что дополненная реальность вносит отдельные искусственные элементы в восприятие реального наблюдаемого мира, а виртуальная реальность конструирует новый (абсолютно мнимый) искусственный мир.

Технологии дополненной реальности позволяют демонстрировать совершенно новые свойства объектов и получать новые ощущения от привычных реальных вещей, используя стандартный персональный компьютер и стандартные периферийные устройства (рис. 1). Поэтому данная технология может эффективно применяться в широких практических задачах [3].

Результатом работы над данным проектом есть приложение для мобильной платформы Android с возможностью демонстрации дополнительного графического контента при наведении камеры устройства на метку.

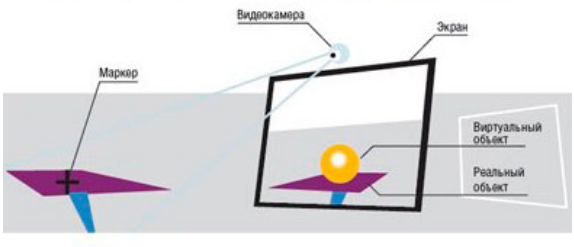

Рис. 1. Как работает дополненная реальность

Литература. **1.** Хабрахабр. Многофункциональный новостной сайт. Режим доступа: http://habrahabr.ru/ **2.** Geektimes. Новостной сайт. Режим доступа: http://geektimes.ru/post/118123/ **3.** Дополненная реальность: настоящее и будущее (конференция от Microsoft). Режим доступа: http://www.optimization.com.ua/reports/augmented-reality-conference-2012.html

***Арчвадзе Н.Н.*[1]*, Пховелишвили М.Г.*[2]*, Шецирули Л.Д.*[3]**

[1] *Тбилисский государственный университет им. И.Джавахишвили, Тбилиси, Грузия;* [2] *Институт вычислительной математики им. Н.Мусхелишвили Грузинского технического университета, Тбилиси, Грузия;* [3] *Батумский государственный университет им. Ш.Руставели, Батуми, Грузия*

Шаблоны для Haskell функций с бесконечными рекурсивными типами аргументов

В парадигме функционального программирования использование методик конструирование структур данных позволяет параллельно создавать и типовые шаблоны (каркасы) функций для обработки этих структур. Такие каркасы можно рассматривать как шаблоны для наполнения необходимой функциональностью. Общий вид таких шаблонов остается неизменным, меняется только содержимое, связанное с требованиями к функциям, определяемыми целями разработчика.

Шаблоны-каркасы функций можно использовать в задачах синтеза программ. Задачу синтеза можно представить как разработку метода автоматической генерации программы компьютером для задачи, которая ранее им не решалась и для которой у него отсутствует алгоритм решения [1-3].

Задача синтеза программ представляет интерес для исследований искусственного интеллекта и разработки программного обеспечения. Построенные программы синтезатором не требуют процесса верификации, чем увеличивается надежность программного обеспечения.

Традиционно функциональное программирование занималось изучением и обработкой такой структуры обработкой данных, как список. Первый функциональный язык Lisp называется так «list processing»-«обработка списка». В работе [4,5] рассмотрены шаблоны Лиспа, которые назывались «обобщенными формами представления» и были использованны для рекурсивных функций Лиспа.

Для представления и обработки таких структур данных, как деревья различной природы (бинарные и любой арности, сбалансированные и прочие), векторы и матрицы, различные сложные отображения, применяется метод синтаксически ориентированного конструирования (был предложен Ч.Хоаром) [6]. Данный метод заключается в конструировании типов данных из других типов (в том числе и рекурсивно из самих себя) при помощи применения двух простых операций — декартова произведения (*) и размеченного объединения (+).

В работе [7,8,9] рассмотрены шаблоны для функций с хвостовой рекурсии и рекурсии на голову списка для языка Haskell. Они определены для функций, обрабатывающее списки элементов некоторого типа A. Шаблоны, для рекурсивных функций, получающую на вход один аргумент-список, имеют следующий вид:

Хвостовая рекурсия:

f un[ ] = g1 [ ]

f un(x : xs) = g2 (g3 x) (g4 (fun (g5 xs)))

Рекурсия приходит на голову списка:

fun [ ] = g1 [ ]

fun (x : xs) = g2 (f un(g3 x)) (g4 (g5 xs))

Функции g1, g2, g3, g4 и g5 зависят от целей разработчиков.

g1–функция для обработки пустого списка; g2–функция для сцепки результатов обработки головы и остатка не пустого списка; g3–функция для обработки головы не пустого списка; g4–функция для обработки результата рекурсивного вызова для остатка не пустого списка; g5–функция для предварительной обработки остатка не пустого списка перед рекурсивных вызовом.

В работе [10] рассмотрены шаблоны для функций с дополнительным аргументом (аккумулятором):

fun n =fun' n a - - Вызов функций, параметр n и a имеют конкретные значения

fun' n a= g1 a

fun' (x : xs) = g2 (g3 x)(g4 (fun' (g5 xs) g6 a))

где: g1 – это функция, которая принимает то значение, которое возвращается начальной функцией; g2– это функция, которая объединяет результаты обработки головы и хвоста списка; g3– это функция, которая обрабатывает голову списка; g4– это функция, которая

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

для непустого списка обрабатывает рекурсивный вызов для хвоста списка; g5 – это функция, которая предварительно обрабатывает для хвоста списка рекурсивный вызов; g6 – это функция, которая обрабатывает аккумулятор.

Возможно определить основные принципы построения определений с накапливающим параметром [10], но невозможно любую функцию преобразовать для вычисления с аккумулятором. Построение функций с накапливающим параметром — приём не универсальный, и он не гарантирует получения хвостовой рекурсии. С другой стороны, построение определений с накапливающим параметром является делом творческим.

В данной работе на примере Haskell рассматриваются шаблоны для бесконечных структуры данных. Бесконечные структуры данных можно определять на основе бесконечных списков, можно также использовать механизм рекурсии. Рекурсия в этом случае используется как обращение к рекурсивным функциям. Третий способ создания бесконечных структур данных состоит в использовании бесконечных типов. Этот последний способ является для нас самым интересным. В языке Haskell для определения рекурсивных типов используется ключевое слово data, что делает структуры данных таких типов потенциально бесконечными.

Например, определение типа для представления двоичных деревьев в Haskell имеет вид:

data Tree a=Leaf a

| Branch (Tree a)(Tree a)

Branch : : Tree a->Tree a->Tree a

Leaf : : a-> Tree a

В этом примере показан способ определения бесконечного типа. Видно, что без рекурсии тут не обошлось.

Примерно таким же образом можно определить и свой собственный тип для представления списков:

data List a = Nil | Cons a (List a)

Как видно, подобное определение полностью соответствует теоретическому определению типа List (A).

При помощи ключевого слова data определяются так называемые алгебраические типы данных, которые характеризуются тем, что над значениями таких типов может быть определена алгебра. Также, ключевое слово data может быть использовано для определения новых типов, которыми могут быть простые типы, типы-контейнеры и перечисления.

Рассмотрим конкретный бесконечный рекурсивный тип:

data Name a = Value a [Name a] (1)

и для функций, которые будут работать с данными этого типа, построим шаблоны. Довольно многие задачи используют рекурсивный тип такого вида, например, дерево любой структуры имеет тип:

data Tree a = Node a [Tree a],

которое означает, что у каждога узла дерево (Node) определяется список поддеревьев. Для листа дерево не нужно специальное значение, так как его можно представить пустым списком - [].

Выделили несколько типов функций, которые могут работать с данными (1) типа:

hasElement :: Eq a => Name a -> a -> Bool (a)

isProp :: Ord a => Name a -> Bool (b)

levelList :: Name a -> Int -> [a] (c)

Функции типа a по смыслу объединяют функции, которые выясняют, имеет ли в определенном данным – Name типом аргументе значение Value типа. Естественно, что нужно ограничение типа на эквивалентность. Функции типа b объединяют функции, которые выясняют, имеет ли аргумент Name типа определенное свойство. Естественно, что нужно ограничение типа на упарядоченность. Третий тип с функции, создают список данного типа (Name) на определенном уровне (второй аргумент). Из-за того, что тип Name бесконечный и рекурсивный, такие задачи встречаются довольно часто.

Каждая функция для обработки значений типа Name должна содержать как минимум две клаузы. Первая обрабатывает Value NIL, вторая — Value nonNIL. Этим двум частям типа Name обычно соответствуют образцы Value [] и Value (x:xs). Обычно, два клауза можно объединить в один с использованием технологии охраны. В теле второго клауза (или второго

выражения охраны) обработка элемента Value xs (или Value tail) выполняется той же самой функцией.

Нужно обратить внимание на то, что, из-за рекурсивности типа, строится многократные рекурсивные обращение, как для просмотра в вертикальном направление, как и горизонтальном. Поэтому нужно организовать многократный рекурсивный процесс как вглубину, так и в ширину.

Шаблон для функций a типа имеет вид:

hasElement :: Eq a => Name a -> a -> Bool

hasElement (Node z []) s = g1

hasElement (Node z (x:xs)) s = if then vBool else g3

Рассмотрим конкретный пример описания рекурсивного типа. Дерево любой структуры имеет следующее описание типа данных:

data Tree a = Node a [Tree a]

Этот описание гласит, что каждый узел дерева имеет список поддеревьев.

Допустим нужно определить следующие функций для обработки дерева:

hEl :: Eq a => Tree a -> a -> Bool

hEl :: Eq a => Tree a -> a -> Bool

Определение этой функций имеет вид:

hEl (Node z []) s = if z==s then True else False

hEl (Node z (x:xs)) s = if z==s then True else (hEl x s)||(hEl (Node z xs) s)

Например, вызовы этой функции имеют вид:

>hEl (Node 5 [Node 6 [Node 7 [], Node 8 []], Node 9 []]) 8

True –значение после вызова

>hEl (Node 's' [Node 'd' [Node 'f' [], Node 'g' []], Node 'h' []]) 's'

True –значение после вызова

Если сравнить hEl функцию с шаблоном hasElement получим, что функция g1 является условным выражением вида: if z==s then True else False. Функция g2 равно предикату z==s. Функция g3 объединяет с помощью оператора ||(or) многократные рекурсивные вызовы как вглубину, так и в ширину и имеет вид: (hasElement x s)||(hasElement (Node z xs) s). Переменная vBool имеет логичесое значение True.

Шаблон для функции b типа имеет вид:

isProp :: Ord a => Name a -> Bool

isProp (Value z []) = vBool1

isProp (Value z ((Value s list):xs)) = if g1 then vBool2 else g2(isProp (Value s list)) (isProp (Value z xs))

Функции g1, g2 и значения организаций многократного рекурсивного процесса как вглубину, так и в ширину. vBool1, vBool2 зависят от целей разработчиков. g1-функция-предикат, который проверяет имеет ли каждый узел и его дочерный узел то свойство, что дано в задаче. g2 - логическая функция and или or, для vBool1, vBool2 –конкретые значения Bool типа.

Пример: для заданного типа данных data Tree a = Node a [Tree a] определена функция isHeap, которая проверяет, является ли дерево пирамидным, т.е. значение в каждом узле меньше чем значения в любых поддеревъях этого узла.

isHeap :: Ord a => Tree a -> Bool

Определение этой функций имеет вид:

isHeap (Node z []) = True

isHeap (Node z ((Node s list):xs)) = if z>s then False else (isHeap (Node s list))and(isHeap (Node z xs))

Например, вызовы этой функции имеют вид:

>isHeap (Node 5 [Node 6 [Node 7 [], Node 8 []], Node 9 []])

True –значение после вызова

>isHeap (Node 's' [Node 'd' [Node 'f' [], Node 'g' []], Node 'h' []])

False –значение после вызова

Если сравнить isHeap функцию с шаблоном isProp получим, что значение vBool1 равно True, vBool2 равно False. Функция g1 возвращает истину, если z>s и ложь, в противном случае. Функция g2 тоже самое, что логический оператор and.

Шаблон для функции с типа имеет вид:
levelList :: Name a -> Int -> [a]
levelList (Value z []) c = g1
levelList (Value z (x:xs)) c = if g2 then vBool else g3
Пример: Функция lList для заданного типа данных data Tree a = Node a [Tree a] создает список значений всех узлов, которые размешены на уровне, который задан вторым аргументом:
lList :: Tree a -> Int -> [a]
Определение этой функций имеет вид:
lList (Node z []) c = if c==1 then [z] else []
lList (Node z (x:xs)) c = if c==1 then [z] else (llList x (c-1))++(llList (Node z xs) c)
Например, вызовы этой функции имеют вид:
>lList (Node 5 [Node 6 [Node 7 [], Node 8 []], Node 9 []]) 3
[7,8] –значение после вызова
>lList (Node 's' [Node 'd' [Node 'f' [], Node 'g' []], Node 'h' []]) 2
['d','h'] –значение после вызова
Если сравнить lList функцию с шаблоном levelList, получится, что функция g1 тоже самое, что выражение if c==1 then [z] else []. g2 функция определяется как выражение c==1. Значение vBool равно списку [z]. Функция g3 объединяет с помощью оператора ++ (append) многократные рекурсивные процессы как в глубину, так и в ширину- (levelList x (c-1))++(levelList (Value z xs) c).

Обобщенные формы для рекурсивной функций применяются в решение задач синтеза и верификации функциональных программ для языков Lisp и Haskell.

Литература. **1.** N.Archvadze, M.Pkhovelishvili, L.Shetsiruli. The complexity of program synthesis from examples. Proceedings of the Eleventh International Conference Pattern Recognition and Informaton Processing (PRIP'2011). ISNB 978-985-448-772-7. pp. 275-279. http://lsi.bas-net.by/conferences/prip2011/ 2011. **2.** Archvadze N.N., Pkhovelishvili M.G., Shetsiruli L.D. Several issues of programs synthesis. Proceedings of the International Conference on System Analysis and Information Technologies. ISSN 2075-4086. pp. 403. http://sait.kpi.ua/books/sait2011.ebook.pdf/view 2011. **3.** N.Archvadze, M.Pkhovelishvili, L.Shetsiruli . Automatically building the "basic recursive" part of the data structures programs descriptions. Proceedings of the System Analysis and Information Technologies 14-th International Conference SAIT 2012. p.323. http://sait.kpi.ua/books/sait2012.ebook.pdf/view. **4.** N. Archvadze,M. Pkhovelishvili, L. Shetsiruli, M. Nizharadze. Program Recursive Forms and Programming Automatization for Functional Languages. WSEAS TRANSACTIONS on COMPUTERS. Volume 8, pp. 1256-1265, ISSN: 1109-2750,2009, http://www.wseas.us/e-library/transactions/computers/2009/29-531.pdf. **5.** N. Archvadze, M. Pkhovelishvili, L.Shetsiruli. Construction of the Generalized Recursive Forms for Functional Languages and their Application Verification of. Electronic Scientific Journal: "Computer Sciences and Telecommunications". No. 3(26), pp. 133-141. ISSN 1512-1232, 2010. http://gesj.internet-academy.org.ge. **6.** Душкин Р. В. Функциональное программирование на Haskell. 2007. **7.** N. Archvadze. M. Nizharadze. Typical Template Verification for List Editing In Haskell Language. Proceedings of the International Conference Management systems and modern information technologies. pp 170–172. ISSN 1512-3979. **8.** N.Archvadze, M.Pkhovelishvili, O.Ioseliani, L.Shetsiruli. THE AUTOMATIC SYNTHESIS OF HASKELL FUNCTIONS. Electronic Scientific Journal: "Computer Sciences and Telecommunications". file:///C:/Users/acer/Downloads/2346.pdf. pp. 20-26. http://gesj.internet-academy.org.ge/en/title_en.php?b_sec=&section_l=comp. **9.** N.Archvadze, M.Pkhovelishvili, O.Ioseliani,L.Shetsiruli. Function Templates for the Synthesis of Functional Programs. International Journal of Computer and Information Technology. ISSN: 2279 – 0764. Volume 03 – Issue 06, November 2014. http://www.ijcit.com/archives/volume3/issue6/Paper030610.pdf..pp.1241-1247. http://www.ijcit.com/Vol3Issue6.php. **10.** N.Archvadze, M.Pkhovelishvili, L.Shetsiruli. Function Definitions with Accumulators in Functional Languages. Proceedings of the System Analysis and Information Technologies. 16-th International Conference SAIT 2014. pp. 335-336. file:///C:/Users/acer/Downloads/sait2014.ebook.pdf.ISBN978-966-2748-33-8..

Білорус Д.С.
Національний технічний університет України "КПІ", факультет інформатики та очислювальної техніки, Київ, Україна

Проблематика задачі пошуку і композиції веб сервісів для підтримки бізнес-процессів

У загальному випадку бізнес-процес — це сукупність взаємопов'язаних заходів і завдань, спрямованих на створення певного продукту або послуги для споживачів. Бізнес-процеси існують всередині кожної організації, незалежно від того, формалізовані вони чи ні. Популярна сьогодні система умовних позначень (нотація) BPMN визначає кінцевий набір графічних елементів для опису бізнес-процесу [1]:

- Об'єкти потоку управління: події, дії та логічні оператори
- З'єднувальні об'єкти: потік управління, потік повідомлень та асоціації
- Ролі: пули і доріжки
- Артефакти: дані, групи і текстові анотації.

Найменшою елементарною дією в процесі є завдання (task). З точки зору автоматизації бізнес процесів за допомогою Web-сервісів, саме завдання є головним цільовим об'єктом. Інші елементи відносяться до задач оркестрування, хореографії, координування, тощо, які частково, або повністю, вирішуються за допомогою мов виконання бізнес-процесів (наприклад: BPEL).

Розглянемо Web-сервіси як автономні, модульні програми, призначені для реалізації бізнес-процесів. Web-сервіси спираються на ряд галузевих стандартів: WSDL (для опису), UDDI (для інформування та публікації) і SOAP (для обміну повідомленнями). Ці специфікації не залежить від платформи і мови, завдяки чому користувачі можуть пов'язувати різні компоненти з різних організаційних структур. Проте, жоден з цих стандартів не передбачає визначення бізнес-семантики Web-сервісів. Таким чином, Web-сервіси залишаються ізольованими. Для вирішення цієї проблеми необхідно з'єднати Web-сервіси та вказати, як їх спільно використовувати для реалізації більш складної функціональності — як правило, для реалізації бізнес-процесів.

Численні існуючі Web-сервіси являють собою розподілену базу обчислень, в яких інформація та послуги надаються за запитом в формі, що може бути оброблена машиною [2]. Однак, якщо взяти будь-яку довільну проблему, малоймовірно, що вона буде розв'язана повністю одним з доступних Web-сервісів; швидше за все, вирішення проблеми, ймовірно, буде потребувати інтеграції результатів, отриманих декількома сервісами. Композиція Web-сервісів та інтеграція отриманої інформації — камінь спотикання сервісно-орієнтованої архітектури.

Проблема композиції Web-сервісів може бути скорочена до трьох основних проблем: перша — зробити план, який описує як взаємодіють Web-сервіси і як функціональність, що вони пропонують, може бути інтегрована для забезпечення вирішення поставленої проблеми. Друга проблема полягає у виявленні Web-сервісів які можуть виконати необхідні завдання. Третя проблема полягає в управлінні взаємодією з цими Web-сервісами. Важливо відзначити, що планування, виявлення та взаємодія строго пов'язані: план визначає тип Web-сервісу, що треба виявити, але це також залежить від Web-сервісів, які наразі доступні. Крім того, процес взаємодії залежить від специфіки плану, але сам план залежить від потреб взаємодії.

Ці три підзадачі диктують набір завдань, які мають бути вирішені будь-якою Web-сервісною системою, що підтримує композицію. Для виявлення Web-сервісів, система повинна бути в змозі представити можливості, надані Web-сервісами, та повинна бути в змозі розпізнати схожість між можливостями, що надаються, і необхідною функціональністю. Друге завдання для Web-сервісної системи — це підтримка взаємодії між Web-сервісами. Така система повинна визначати специфікацію за якою Web-сервіси отримують інформацію та повертають результат, а також протоколи на механізми низького рівня, що необхідні для виклику Web-сервісів.

Література. **1.** Information technology — Object Management Group Business Process Model and Notation, ISO/IEC 19510, 2013. **2.** D. Both, M. Champion, C. Ferris, F. McCabe, E. Newcomer, and D. Orchard. Web services architecture. http://www.w3.ord/TR/2003/WD-ws-arch-20030514/, 14 May 2003. W3C Working Draft.

***Булах Б.В.*[1], *Лашко О.В.*[2]**
[1]*Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна;* [2]*Національний технічний університет України "КПІ", Київ, Україна*

Семантичні технології в неруйнівному контролі

Неруйнівний контроль (НК) якості та надійності виробів, їх стану в процесі експлуатації є високотехнологічною галуззю, і залучення інформаційних технологій (ІТ) нині є необхідністю при вдосконаленні методів контролю, розробці та впровадженні апаратно-програмних засобів контролю та обробці результатів НК. Серед актуальних нині напрямків розвитку ІТ, які можуть стати в нагоді при вирішенні всіх вищезгаданих задач, окреме місце посідають семантичні технології. Доцільність та конкретні переваги їх використання в НК досліджені відносно слабко.

Семантичні технології (СТ) передбачають вилучення та обробку знань про дані певної предметної області окремо від обробки самих даних. У вузькому сенсі це робота на рівні метаданих, що дозволяє машинам "розуміти" дані, здійснювати логічні висновки на їх основі, а людині — простіше вносити зміни до структур даних та програм, просто додаючи нові взаємовідношення на рівні метаданих (редагуючи онтології або додаючи нові факти в базу знань, а не змінюючи структуру баз даних або програмний код).

База знань засобів контролю. Одна з можливих областей застосування СТ в НК — програмне забезпечення, що допомагає спеціалістам НК проводити контроль складних промислових виробів в умовах, коли множина таких засобів контролю досить динамічна. Як правило, для здійснення періодичного контролю тих або інших виробів у відповідній довідковій документації наводиться список існуючих на той момент рекомендованих засобів, і навіть кваліфікований спеціаліст може не мати точного уявлення про наявні відповідники всіх таких засобів, особливо на великих промислових підприємствах з великою кількістю об'єктів контролю. Ситуація погіршується у випадку наявності застарілих інструкцій, де вказані засоби контролю, що вже не випускаються. Тому нині висуваються пропозиції використати СТ для створення бази знань засобів контролю з метою автоматизації визначення їх функціональної еквівалентності та з урахуванням їх характеристик [1].

Інтеграція програмних засобів та узгодження форматів даних. Інша можлива область застосування СТ в НК - програмне забезпечення, що моделює методи або засоби контролю. Часто такі програми мають справу з моделюванням об'єктів різної фізичної природи, а тому постає питання узгодження роботи різних програмних модулей, що мають свою специфіку конфігурування, функціонування, свої формати даних тощо. Створення онтології таких програмних компонентів дозволить автоматизувати створення та диспетчеризацію потоків виконання обчислень при моделюванні складних або комбінованих засобів НК [2].

Аналіз результатів контролю. Нарешті, СТ можуть стати в нагоді при обробці та аналізі вихідних даних НК: при класифікації отриманої інформації щодо знайдених дефектів, у допомозі в прийнятті адекватних та ефективних рішень за результатами контролю та ін.

Загалом слід зазначити, що темі застосування семантичних технологій у неруйнівному контролі досі присвячена досить невелика кількість опублікованих робіт, і ця перспективна область перетину НК та ІТ потребує подальших досліджень.

Література. **1.** Kamsu-Foguem B. Knowledge-based support in Non-Destructive Testing for health monitoring of aircraft structures / Advanced Engineering Informatics Journal. - Vol. 26 Issue 4. - October, 2012. - P. 859-869. **2.** Mckenna J.P., McLean N.I., Gachagan A. et al. The design of a multi-agent system for flaw-response modelling / 45th Annual British Conference on Non-Destructive Testing, Stratford-upon-Avon, United Kingdom. - 2006.

Галатенко Д.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Адаптивний фільтр для правильної локалізації підзображень

На сьогоднішній день, комп'ютерний зір (Computer Vision) набув досить широкрго застосування в різних галузях IT-технологій. Це пов'язано з розробкою та впровадженням алгоритмів, які можуть автоматично розпізнавати об'єкти, класифікувати їх та обробляти отриману інформацію. Однією із фундаментальних проблем комп'ютерного зору є розпізнавання об'єкту за допомогою заданого шаблону, тобто локалізація підзображення, що міститься в зображенні.

Загальна техніка вирішення цієї проблеми полягає у заповненні матриці шаблонного зображення (або підзображення) нулями або середніми значеннями одного із зображень. Ми можемо легко позбутися від областей, що оточують підзображення(A), встановлюючи значення за його межами рівними нулю. Таким чином змінюється діапазон підсумовування, під час проходу великого зображення(B). Значення за межами A стають нулем, незважаючи на значення в B. В якості інструменту для оперування зображеннями вибираємо алгоритм базований на FFT - Fast Fourier Transform [1], оскільки він забезпечує вищу швидкість виконання, ніж інший популярний алгоритм базований на кроскореляції [2]. Але практика показує, що заповнення нулями не дає очікуваного результату. Справа в тому, що одне із зображень може мати середнє значення більше, ніж середнє для вихідного зображення, що призведе до хибних розрахунків.

Шляхом вирішення цієї проблеми є використання нормалізації. Нормалізація не буде скасовувати попередні рішення (заповнення матриці нулями), а також працюватиме на локальному рівні у більшому зображенні. Загальний алгоритм нормалізації зображення має такий вигляд:

1. знайти середню величину відстані між пікселями використовуючи низькочастотний фільтр (ФНЧ) [3];
2. відняти середню величину від значень відстані між пікселями початкового зображення;
3. обчислення середньоквадратичного відхилення величини відстані між пікселями;
4. нормалізація середньоквадратичного відхилення;

В результаті отримуємо нормалізоване зображення, яке придатне для перетворення Фур'є.

Однак існує дві проблеми на границях зображення B, де ковзаюче вікно A виходить за межі B. Перша проблема - числова точність: обчислюючи дисперсію в кумулятивній формі, не варто користуватись типом даних double, бо це призводить до значної похибки. Друга проблема полягає в тому, що на границях B ми маємо найменшу концентрацію корисної інформації. Очевидно, що залежно від розміру вікна, ми будемо отримувати різні результати кроскореляції. Оскільки границі - найчастіші області помилок в обчисленні кроскореляції і розмір вікна пов'язаний з точністю результату, ми повинні затемнити ті пікселі, що входять в граничну область. Цей процес називається *адаптивним контрастним збільшенням* і саме він дає можливість вирівняти дисперсію та середнє значення зображення B і шаблону A, а також виконати корректну процедуру локалізації зображення.

Завершальним кроком вирішення задачі є SAD відповідність [4] (Sum of the Absolute Differences). SAD відповідність використовує значення кореляції, щоб поділити повне зображення на комірки розміром підзображення. У кожній комірці відображається позиція найбільшої кореляції. Для цієї позиції ми визначаємо суму абсолютних відмінностей (SAD) і серед усіх комірок вибираємо ту, яка має найменше значення SAD.

В результаті отримуємо готовий адаптивний фільтр, який здатний правильно локалізувати підзображення за прийнятний час.

Література. **1.** Rao, K. R., Kim, Do Nyeon, Hwang and Jae Jeong, Fast Fourier Transform - Algorithms and Applications 2010. http://www.springer.com/us/book/9781402066283 **2.** JULIUS O. SMITH III, MATHEMATICS OF THE DISCRETE FOURIER TRANSFORM. http://www.dsprelated.com/freebooks/mdft/Cross_Correlation.html **3.** Принципи Ущільнення та Перетворення Зображень, В. П. Кожем'яко, ВНТУ, 2011. pp. 31 - 47. http://publish.vntu.edu.ua/txt/Kojemiako_PrincipyUschilnZobrazh.htm **4.** SAD, http://en.wikipedia.org/wiki/Sum_of_absolute_differences.

Джунковський В.О., Джунковський Ю.О.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Задача вибору лідера в колективах однорідних агентів (роботів)

Роботи міцно увійшли в наше життя і стали основою економічної і військової могутності розвинених країн світу. На даний час у світі нараховується більше 1,3 мільйонів роботів. А за оцінками міжнародної федерації робототехніки (IFR), кожного року продається 160 тис. роботів різного призначення [1].

В наш час найбільш поширеними є роботи, які не можуть обійтися без втручання в їхні дії оператора, який допомагає їм виконувати те чи інше завдання. Однак, на сьогоднішні день все більше розробників працюють над створенням роботів, які будуть працювати і виконувати задачі за своїм призначенням самостійно. Змога роботів працювати самостійно, без допомоги оператора, за складних умов зовнішнього середовища, допомагає виконувати завдання значно вищі за складністю і робити це з набагато більшою якістю і швидкістю. Але змога роботів працювати самостійно це далеко не межа їх вдосконалення. Наступним етапом розвитку робототехніки є створення колективів роботів. Даний вектор розвитку в галузі робототехніки із застосуванням роботів що мають елементи штучного інтелекту значно розширює область завдань, які можуть виконувати роботи.

Відмітимо, що штучний інтелект і робототехніка завжди були тісно пов'язані один з одним. Одним з важливих напрямів штучного інтелекту до сих пір вважається цілеспрямована поведінка роботів (створення інтелектуальних роботів, котрі зможуть автономно виконувати операції для досягнення цілей, що задаватиме людина).

Активні дослідження в області створення систем взаємодії роботів ведуться вже майже чверть століття. Такі напрямки, як колективна, ройова, зграйна і ін. робототехніка, зайняли важливі позиції в сучасній робототехніці і теорії багатоагентних систем. Однак, до сих пір переважне число досліджень в цій області залишається на теоретичному, модельному рівні.

Переваги групового застосування роботів очевидні. По-перше, це більший радіус дії, по-друге, розширений набір виконуваних функцій і нарешті, більш висока ймовірність виконання завдання, що досягається за рахунок можливості перерозподілу цілей між роботами групи в разі виходу з ладу деяких з них. Тому такі складні завдання як, наприклад, масштабне дослідження і зондування поверхні інших планет, збірка складних конструкцій в космосі і під водою, участь у бойових діях і забезпечуючих операціях, розмінування територій тощо, можуть бути ефективно вирішені роботами тільки при їх груповій взаємодії.

Однією з принципових особливостей колективної робототехніки є локальний характер взаємодії роботів одне з одним, а також роботів з середовищем [2]. Така взаємодія називається неявною комунікацією(implicit communication) [3]. Мова йде про те, що кожен робот групи безпосередньо взаємодіє лише зі своїми сусідами, які знаходяться в деякій обмеженій зоні видимості. Звідси як правило слідує, що в такій системі роботи самостійно приймають рішення про подальші дії, спираючись на деякі прості правила локальної взаємодії. Однак, разом з тим, переважна кількість прикладів вирішення задач в області колективної робототехніки стосується узгодженого руху і узгоджених дій колективу роботів. Тому основними задачами даної області є створення методів управління даними колективами. На сьогодні одним з основних методів управління групою роботів є задача визначення лідера.

Визначення лідера в колективі роботів дає змогу використовувати основні переваги колективної робототехніки. Одну з найбільших переваг яку надає обрання лідера є живучість такої багатофункціональної системи. Живучість такої системи забезпечується тим, що у випадку виходу з робочого стану робота, який мав права лідера, система не стане неконтрольована, навпаки вона одразу відреагує на цю подію і почне проведення процедури голосування,в якому братимуть участь всі роботи даного колективу, за допомогою цієї процедури і визначить наступного лідера групи, який візьме на себе контроль над групою. Критерії за якими члени колективу будуть обирати лідера залежать від типу одиниць колективу та завдання що перед

ними стоїть.

Велику роль у визначенні лідера відіграє рівень когнітивних можливостей агента, адже він будучи головним центром має командувати діями колективу знаючи ситуацію навколо. Враховуючи те що роботи обмежені локальною взаємодією, це може вирішуватись з міркувань, що агент котрий має найбільшу кількість зв'язків з іншими агентами стає претендентом на місце лідера, в чиї функції буде входити прийняття рішень щодо подальших дій колективу. Ці рішення він прийматиме на основі аналізу інформації, збір якої будуть здійснювати роботи що знаходяться на периферії , а обробку роботи, що безпосередньо знаходяться в його оточенні. І саме кількість агентів що знаходяться безпосередньо біля агента-претендента може бути основним критерієм прийняття рішення колективом під час процедури голосування. Це можна побачити на прикладі рис. 1.

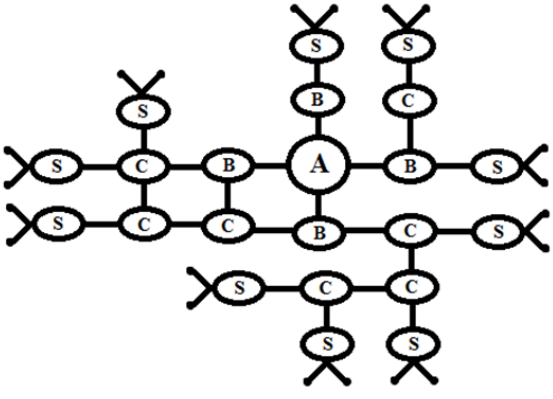

Рис. 1. Приклад багатоагентної системи

На рис.1 агент A стає керуючим центром, його найближчі сусіди B – аналізаторами, а агенти S що знаходяться на периферії будуть відповідальні за зовнішню сенсорику.

Отже, організація колективу роботів на чолі з лідером дозволяє, по-перше, значно розширити області застосування роботів, по-друге, вприту наблизитися до вирішення проблеми масового застосування мікроробототів у складі великих груп, які налічують тисячі і десятки тисяч мікроробототів. Застосування груп роботів, здатних виконувати складні роботи, наприклад, при ліквідації наслідків природних або техногенних катастроф, зростання числа яких спостерігається останнім часом, в свою чергу, дозволить знизити ризик для людей в умовах шкідливих або навіть небезпечних середовищ, скоротити витрати, пов'язані із забезпеченням безпеки, надасть змогу досліджувати важко досяжні місця космосу в такий спосіб, який є неможливим для групи вчених в силу людської фізіології, допоможе збільшити продуктивність підприємств і таке інше.

Література. **1.** Добрынин Д.А. Интеллектуальные роботы вчера, сегодня, завтра//X национальная конференция по искусственному интеллекту с международным участием КИИ-2006(25-28 сентября 2006 г., Обнинск):Труды конференции. В 3-х т. Т.2. М.: Физмамлит, 2006. **2.** Zhigou Shi, Jun Tu, Qiao Zhang, Lei Liu, Junming Wei, A Surwey of Swarm Robotics System//Proc. of the Third Intern. Conf. on Advances in Swarm Intelligence, Shenzhen, China, 2012. V.1, P. 564. **3.** Yogeswaran M. and Ponnambalam S. G. (2010). Swarm Robotics: An Extensive Research Review, Advanced Knowledge Application in Practice, InTech. P. 259.

Діденко Д.Г.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Нечітке відновлення довідника за кодом та назвою у базі даних

Застосовується алгоритм нечіткого порівняння рядків для видалення «схожих дублікатів».

Вступ. Щодня OLTP системи наповнюються новими даними за допомогою численої кількості операторів та кінцевих користувачів. Така ручна робота містить 5% помилок у вигляді опечаток, які згодом під час чергового імпорту-експорту потрапляють до OLTP систем [1]. Для виправлення помилок можна застосовувати різні методи: кластерізація, пошук за довідником, генетичні алгоритми та нейронні мережі [1, 2]. Але всі ці підходи дуже трудоємні.

Постановка задачі. Проблема якісного аналізу даних важлива, особливо для кластерів та хмарних баз даних, які мають дуже великий розмір. В даному випадку необхідно відтворити довідник T зі стовпчиками «ID» та «NAME», який не містить дублікатів «NAME» та «схожих дублікатів». Схожість дублікатів визначається за відстаню між кортежами, використовуючі «редакційну відстань», яка більш детально розглядається в [3].

Реалізація. Після отримання вхідного файлу імпорту у вигляді XML-файлу, необхідно виконати наступні дії:

1. Створення таблиці довідника T («ID», «NAME») з первинним ключем «ID», причому поки альтернативний ключ «AK_NAME» не створюється, тому що він буде заважати далі робити навмисне дублювання рядків.
2. Імпорт вхідного XML-файлу за допопомогою JPA [4].
3. Українізація довідника - замінити всі входження англійських літер на українські.
4. Відновлення регістру даних: якщо в стовпчику всі дані лише з великої літери, тоді замнюємо їх на відповідні за допомогою нейронної мережі.
5. Видалення повторів зі стовпчику «NAME» з автоматичним відновленням значень зовнішнього ключа в залежних таблицях. Це найшвидше робиться за допомогою курсору, всередині вкладеного курсору.
6. Знаходження відстані між рядками дозволяє знайти схожі значення в стовпчику «NAME», та вилучити їх так само як в пункті 5.
7. Створення унікального (альтернативного) ключа «AK_NAME» на поле «NAME».

Висновок. Наведений алгоритм використовується при роботі з великою кількістю даних для якісного покращення БД та видалення зайвих дублікатів. Цей алгоритм має достатню швидкодію для великих наборів даних. Подальшою розробкою є реалізація цього підходу в розподілених системах для MapReduce (наприклад для Hadoop) за допомогою мови програмування Java.

Література. **1.** Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP. - 2-е изд., перераб. и доп. - СПб.: БХВ-Петербург, 2007. - 384 с.: ил. ISBN: 5-94157-991-8. **2.** Пасічник В.В. та ін. Глобальні інформаційні системи та технології: моделі ефективного аналізу, опрацювання та захисту даних. Монографія / В.В. Пасічник, П. І. Жежнич, Р. Б. Кравець, А. М. Пелещишин, Д. О. Тарасов. Львів: Видавництво Львівської політехніки, 2006. 348 с. **3.** Райкова О. А. Нестрогое сопоставление записей реляционных баз данных с использованием редакционного расстояния между кортежами и ключевого набора атрибутов [Текст] / О. А. Райкова // Технические науки: проблемы и перспективы: материалы междунар. науч. конф. (г.Санкт-Петербург, март 2011 г.). — СПб.: Реноме, 2011. — С. 137-140. **4.** Шилдт Г. Java 8. Полное руководство. 9-е издание . - М.: Диалектика, 2015. - 1376 с.: ил.

Діденко Д.Г., Попова А.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Прискорення DataMining за допомогою Hadoop для пошуку схованих залежностей даних в БД

Вступ. У сучасному світі особливу цінність набувають ефективні способи обробки інформації. Бази даних, а також системи керування цими базами (СКБД) стали необхідними у будь-якій організації. Освітні заклади, банки, страхові, комерційні та інші компанії збирають та накопичують у своїх базах гігабайти інформації про співробітників, послуги, товари і т.д. Цінність подібних даних безперечна. Але добути з них нетривіальні, практично корисні і доступні для інтерпретації знання, необхідні для прийняття рішень, часом стає дуже складно. Недоліки статистичних методів спонукали до розвитку нових методів дослідження складних систем, постала нагальна потреба в такій технології, яка автоматично видобувала б із даних нові нетривіальні знання у формі моделей, залежностей, законів тощо, гарантуючи при цьому їхню статистичну значущість. Новітні підходи, спрямовані на розв'язання цих проблем, дістали назву інтелектуального аналізу даних (Data Mining) [1].

Постановка задачі. В даній роботі об'єктом дослідження було обрано базу даних зовнішньо-економічної діяльності України. Для пошуку залежностей у заданій БД як засіб DataMining був реалізований метод кластерного аналізу. Для оцінки адекватності отримані результати порівнюються з результатами існуючих програмних продуктів.

Так як обробка такої великої кількості матеріалу займає багато часу і вимагає багато ресурсів, її виконують за допомогою проекту Hadoop, що реалізує модель розподілених обчислень MapReduce. Ця модель створена для обчислення деяких наборів розподілених задач з використанням великої кількості комп'ютерів об'єднаних у кластер. Її робота складається з двох етапів: Map та Reduce. На Map-етапі здійснюється попередня обробка вхідних даних. Для цього один з комп'ютерів (головний вузол – master node) отримує вхідні задачі, розділяє на частини і передає іншим комп'ютерам (робочим вузлам – worker node) для попередньої обробки. На Reduce-етапі відбувається згортка попередньо оброблених даних. Головний вузол отримує відповіді від робочих вузлів і на їх основі формує результат – розв'язок задачі, що була сформульована на початку (рис.1). Операції попередньої обробки працюють незалежно одна від одної і можуть виконуватися паралельно, що також надає можливість відновлення після часткового порушення роботи серверів [2]. Тому не дивно, що зараз модель Map Reduce складає основну частину обчислювальної інфраструктури таких компаній як Yahoo, Facebook, LinkedIn та Twitter [3].

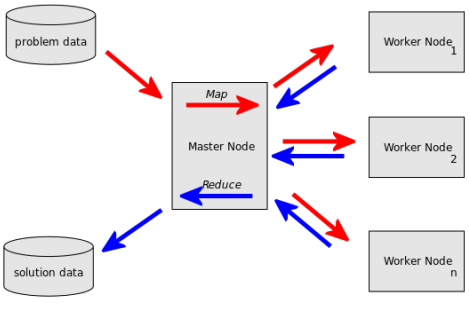

Рис. 1. Схематичне зображення моделі MapReduce

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Реалізація. Розглянемо докладніше формальну постановку задачі кластеризації. Нехай X - множина об'єктів, Y - множина кластерів. Задана функція відстані між об'єктами $\varphi(x, x')$. Існує кінцева навчальна вибірка об'єктів $X^m = \{ x_1, ..., x_m\} \in X$. Необхідно розбити вибірку на підмножини, що не перетинаються так, щоб кожна множина (кластер) складався з об'єктів близьких за метрикою φ, а об'єкти різних кластерів значно відрізнялись. При цьому кожному об'єкту $x_i \in X^m$ ставиться у відповідність номер кластера y_i.

На першому етапі відбувається виділення характеристик об'єктів. Обираються властивості, що характеризують наші об'єкти. Ними є кількісні або якісні характеристики. Так як в роботі ми працюємо з базою даних зовнішньо-економічної діяльності України, то об'єктами виступають здійснені платіжні операції, а в якості характеристик – загальна ціна операції, форма розрахунку, приналежність до категорії імпорт чи експорт, країна-партнер та інші. Далі зменшується, якщо це можливо, розмірність простору характеристичних векторів, тобто обираються найбільш важливі властивості об'єктів. Нормуються обрані характеристики.

На другому етапі визначається метрика. Обирається функція для обчислення відстані між об'єктами, вона являє собою один з основних управляючих факторів, що впливає на результати кластеризації.

Після цього проходить безпосереднє розбиття об'єктів на групи. Програмується алгоритм кластеризації – функція $a : X \to Y$, яка будь-якому об'єкту $x \in X$ ставить у відповідність номер кластера $y \in Y$. Множина Y в деяких випадках відома , однак в даній роботі ставиться задача встановити оптимальне число кластерів з точки зору критерія якості кластеризації.

Введення даних та представлення результатів відбувається за допомогою web-додатка, що додає зручності і зменшує вимоги до програмного забезпечення користувача.

Висновок. При роботі з великими обсягами даних методи кластеризації можуть стати дуже важливим етапом інтелектуального аналізу, адже шляхом виявлення кластерної структури, можливе:

- краще розуміння даних, їх природи (розбиття вибірки на групи схожих об'єктів дозволяє спростити подальшу обробку даних, застосовуючи до кожного кластера свій метод аналізу) ;
- стиснення даних (якщо початкова вибірка надлишково велика, то можливо її скоротити залишивши по одному найбільш типовому представнику від кожного кластера) ;
- виявлення новітніх знань (виділяються нетипові об'єкти, які не вдається приєднати до жодного з кластерів) [4].

Не менш важливою і актуальною задачею було використання моделі розподілених обчислень. Адже зараз, коли обсяги даних сягають терабайт, Hadoop та інші технології розподіленої обробки великих масивів даних швидко стають важливим вмінням для широкого кола програмістів, і ігнорувати такі потужні інструменти означає свідомо лишатись позаду.

Література. **1.** Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. — 3-е изд., перераб. и доп. — СПб.: БХВ-Петербург, 2009. — 512 с. **2.** Hadoop в действии / Лэм Чак - М.: ДМК Пресс, 2012. - 424 с. **3.** Сайт компанії The Apache Software Foundation - http://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html **4.** Scaling Clustering Algorithms to Large Databases / Bradley, P., Fayyad, U., Reina, C., Proc. 4th Int'l Conf. Knowledge Discovery and Data Mining, AAAI Press, Menlo Park, Calif., 2008.

Дідик С.А., Бідюк П.І.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Проблема відтоку абонентів телеком-операторів та можливі підходи до її вирішення

Виконується оцінювання проблеми відтоку клієнтів та методів боротьби з ним в телекомунікаційній індустрії.

На сьогоднішній день спостерігається відсутність динамічного зростання ринку телекомунікаційних послуг за рахунок чого прибуток від залучення нових абонентів щорічно зменшується. За такої ситуації особливо важливим та актуальним питанням індустрії є контроль та зменшення відтоку абонентів, тобто збереження існуючих доходів. У зв'язку з цим телеком-оператори звертають велику увагу на напрям дослідження CRM (Customer relationship management) для підвищення лояльності та утримання існуючих абонентів. В сукупності із залученням нових користувачів утримання існуючих абонентів забезпечує вищі доходи та динамічне збільшення клієнтської бази у довгостроковій перспективі.

Проникнення послуг мобільного зв'язку на ринку становить 120%, а це означає, що можливостей суттєвого збільшення прибутку виключно за рахунок залучення нових клієнтів майже немає. За такої ситуації на ринку важливим пріоритетом стає утримання існуючих користувачів послуг компанії, збереження лояльності і захист своїх позицій від «атак» конкурентів. Ринок перенасичений і напрям вектору розвитку компанії змістився у бік роботи з існуючою абонентською базою та її розвитком, що гарантує збереження існуючих доходів та можливе їх примноження у перспективі. [1].

Існує певний перелік причин відтоку абонентів. Серед них такі: вигідніша конкурента пропозиція, незадоволення якістю наданих сервісів тощо. Ці характеристики необхідно врахувати при побудові математичної моделі для прогнозування відтоку клієнтів оператора зв'язку [2].

Ймовірність відтоку абонента вираховується за допомогою аналізу трафіку та прихованих закономірностей поведінки, які вказують на бажання припинити користування послугами поточного оператора. Додатково оцінюється рівень лояльності до компанії (показник NPS – Net Promoter Score), який ілюструє загальну тенденцію задоволеності клієнтів. Значення NPS дає можливість додатково сегментувати абонентів та точніше встановити причини незадоволення послугами оператора зв'язку за рахунок опитування за рядом важливих параметрів.

Знаючи ймовірність відтоку та специфічні характеристики його поведінки, у рамках напряму використання CRM можна побудувати стратегію та загальний концепт утримання абонентів. Іншими словами: «Попереджений значить озброєний». Таким чином, наявність розгорнутих знань стосовно поведінки клієнта надає можливість підібрати індивідуальну пропозицію, яка максимально здатна утримати або вплинути на його поведінку [3].

Прогнозування відтоку полягає у визначенні ймовірності відтоку абонентів —схильності до відтоку для кожного абонента генеральної сукупності на певному майбутньому часовому інтервалі. Результати прогнозування відтоку (значення скорингу) — схильність до відтоку для кожного абонента з генеральної сукупності — використовуються для планування цільових кампаній по утриманню клієнтів [4].

Для прогнозування відтоку клієнтів використовують різні математичні моделі і методи, які на них грунтуються, серед них — логістична регресія, нейронні мережі, дерева рішень, байєсівські мережі (БМ), генетичні алгоритми та скорингові моделі.

Переваги використання методу дерева рішень полягають у швидкості процесу навчання, генеруванні правил у тих сферах, де знання складно формалізувати, а також відносно проста і зрозуміла класифікаційна модель, а також прийнятна точність оцінок прогнозів. Загалом мережі Байєса надають розробникові моделі можливість комбінувати експертні знання і статистичні дані з метою отримання оцінок структури і параметрів моделі. Дерево рішень і логістична регресія, працюють лише з обмеженою кількістю вхідних даних. Тому при побудові моделей на основі цих методів застосовують невелику кількість факторів (близько 10), які

вибираються або експертами, або програмою, але при цьому виникає загроза, що будуть не враховані і відкинуті суттєві фактори, які врешті-решт можуть спричинити серйозний вплив на остаточний результат. Безперечними перевагами БМ є такі: – можливість коректного встановлення існуючих причинно-наслідкових зв'язків між змінними; – об'єднання в одній моделі дискретних та неперервних змінних; – можлива висока розмірність моделі; – наявність множини методів оцінювання структури і параметрів моделі у формі спрямованого графа [5]. Крім того, БМ – це хороша альтернатива для регресійного аналізу, яка логічно доповнює результати, отримані за допомогою регресійної моделі. Однак, мережа Байєса має схильність до «перестраховки», тобто вона частіше помиляється, відкидаючи тих клієнтів, які б не змінили оператора. Саме тому необхідно запропонувати підхід, який унеможливить випадкове відкидання факторів і не буде занадто суворим для прийнятних клієнтів [6].

Логістична регресія, дерева рішень та байєсівські мережі надають можливість будувати наглядні і легко інтерпретовані моделі, якими можна досить просто скористатись для розв'язання практичних задач.

У подальших дослідженнях будуть проаналізовані та використанні сучасні підходи до прогнозування відтоку абонентів. Пропонується побудова нелінійної логістичної регресії на основі певного набору предикторів, вибраних у якості основних із трафіку абонента, а також періоду перебування в якості абонента у мережі конкретного оператора (наприклад, lifetime) та витрат на послуги (ARPU).

Для побудови та дослідження прогнозуючих моделей буде використано програмне забезпечення SPSS, яке дає можливість випробувати наявні альтернативні підходи до моделювання та якісно оцінити проміжні і остаточні результати оцінювання (прогнозування) стану абонентів. Очевидно, що для прискореного отримання високоякісного розв'язку цієї задачі доцільно побудувати систему підтримки прийняття рішень на принципах системного аналізу.

Література. **1.** [Електронний ресурс] Режим доступу: http://befirst.com.ua/interview/show/153-analitika-sposobna-prinosit-millioni/ **2.** Sas Institute Inc. Прогнозирование Оттока Клиентов. // Sas Institute Whate Paper. — 1999. **3.** Албитов А., Соломатин Е. CRM (Customer Relationship Management). [Електронний ресурс] Режим доступу: www.cfin.ru/itm/crm-review.shtml. **4.** Давнис В.В. Адаптивные модели: аналіз и прогноз в экономических система / В.В. Давнис, В.И. Тинякова. – Воронеж: Воронежский государственный университет, 2006. – 380 с. **5.** Бідюк П. І. Аналіз та математичне моделювання економічних процесів перехідного періоду/Бідюк П. І. Половцев О. В. - Київ : ПЛАБ-75, 1999. - Т. 209с. **6.** Бідюк П.І. Економетричний аналіз часових рядів / Конспект лекцій. –К., НТУУ «КПИ», 2007. - 250 с.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Драган А.В.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Построение модели фреймворка с пакетным разбиением кода для внедрения в приложение как зависимость

Введение. Фреймворк – это библиотека классов, которая облегчает написание кода определенного программного продукта. Моделирование фреймворка в первую очередь включает в себя применение архитектурных решений построения кода. С ростом функционала размеры такой библиотеки будут очень быстро увеличиваться, и зачастую не все компоненты нужны для использования. Фреймворк является зависимостью в основном программном продукте и должен быть легко доступен для загрузки и подключения к исходникам. В этих целях используется технология NuGet [1] приложений, которая позволяет создавать пакеты кода, загружать их в централизованное хранилище и легко добавлять как зависимость в код.

Постановка задачи. С ростом функционала библиотеки классов будет увеличиваться размер скомпилированного кода, а также количество самих классов и проектов. Полезной была бы возможность разбиения одного фреймворка на несколько пакетов с кодом и подключения в продукт только необходимых компонентов. Это позволит сделать код чище и сэкономит размер приложения и время загрузки кода с серверов. В докладе представлена модель фреймворка на примере использования технологии NuGet, которая учитывает данные характеристики.

Основной материал. В докладе была построена такая модель фреймворка:

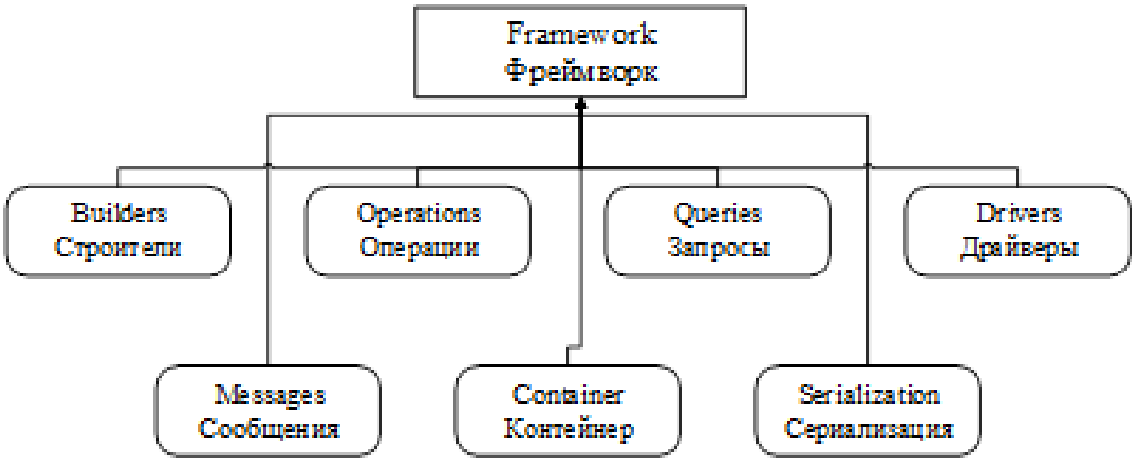

Рис. 1. Схема компонентов фреймворка

Рассмотрим предназначение всех компонентов:

1. Builders (рус. строители, далее - билдеры) – возвращают сущности. Они выполняют функцию «строителя», который собирает по элементам сущность для использования ее в коде.
2. Operations (рус. операции) – представляют все возможные операции, которые выполняет разрабатываемый продукт. Имеют реализацию для каждого из варианта использования.
3. Queries (рус. запросы) – это все запросы для предоставления информации из баз данных разрабатываемого приложения. Так же имеют реализацию всех вариантов использования.
4. Drivers (рус. драйверы) – набор драйверов, необходимых для интеграции фреймворка. Такими драйверами выступают: драйвер для связи с базой данных, драйвер для отправки сообщений в очередь MSMQ, драйверы для отправки API запросов [2].
5. Messages (рус. сообщения) – содержит в себе набор сообщений, т.е. классов, которые отправляются и обрабатываются в пределах разрабатываемого приложения.

6. Container (рус. контейнер) – собирает экземпляры классов, которые разбиты по вариантам использования в зависимости от конфигурационного файла фреймворка.
7. Serialization (рус. сериализация) – предоставляет возможность преобразовывать экземпляры классов в XML строку для отправки ее в запрос.

Каждая компонента – независимый проект кода. Все связи реализованы с применением архитектурного решения Dependency Injection [4]. Внедрение зависимости (англ. Dependency injection) — процесс предоставления внешней зависимости программному компоненту. Является специфичной формой «обращения контроля» (англ. Inversion of control), где изменение порядка связи является путём получения необходимой зависимости. Важной характеристикой фреймворка, как и любого другого программного продукта, является тестируемость [5], т.к. он изменяется вместе с тестируемым приложением, а это не должно повлиять на его функциональность и правильность выполнения тестов. Для этого не следует использовать статические классы, т.к. они не поддаются эффективному юнит-тестированию. Исходя из этого, при построении фреймворка используется внедрение зависимостей с использованием свойств. Каждое из таких свойств является типом необходимого базового класса. Решением проблемы создания региональных экземпляров классов являются контейнеры – они в зависимости от файла конфигурации приводят нужные типы к базовым, а при обращении к контейнеру через свойства и внедрения зависимостей, будет создан необходимый экземпляр базового класса. Такое решение позволяет разбить код на независимые пакеты. Они будут небольших размеров, что позволит их быстро загружать во время сборки приложения и позволит сократить время прохождения стадий релиза в системе непрерывной интеграции. Пакеты будут содержать определенные проекты кода, что позволит сделать код таргетированным. Таким образом, фреймворк станет сборкой небольших библиотек классов.

Выводы. Построение вспомогательных фреймворков является необходимым для больших проектов, так как позволяет сэкономить время разработки и тестирования в будущем. При создании модели фреймворка необходимо изначально определить его функционал и разбить его по содержанию по проектам.

Предложенная модель фреймворка учитывает следующее:

1. Все компоненты должны быть слабо связаны друг с другом, что реализуемо при помощи Dependency Injection.
2. Контейнер должен создавать экземпляры только нужных классов в зависимости от вариантов использования разрабатываемого продукта (например, регион работы продукта).
3. Разбитие фреймворка по компонентам позволяет сделать возможным создание пакетов с кодом. Они должны быть минимальных размеров для увеличения скорости сборки приложения и загрузки зависимостей.
4. Все пакеты хранятся в централизованном хранилище и маркируются версиями и описаниями.
5. Должна быть возможность поддержки стабильных версий для релизного кода.

Литература. **1.** Управление библиотеками проектов с помощью NuGet – Режим доступа: https://msdn.microsoft.com/ru-ru/magazine/hh547106.aspx **2.** Биберштейн Н. Компас в мире сервис-ориентированной архитектуры / Биберштейн Н. – Санкт-Петербург: КУДИЦ-Пресс, 2007. – 256 с. **3.** Osherove, R. The Art of Unit Testing [Text] / Osherove, R. – 2nd edition. – New York: Softbound, 2012. – 375 p. **4.** Seemann, M. Dependency Injection in .NET [Text] / Seemann, M. – USA: Munning Publications Co., 2012. – 584 p. **5.** Cwalina, K. Framework Design Guidelines: Conventions, Idioms, and Patterns for Reusable .NET Libraries [Text] / Cwalina, K. – USA: Addison-Wesley Professional, 2008. – 480 p.

***Дубовой В.М.*[1], *Дуда М.Е.*[1], *Байас М.М.*[2,3]**
[1] *Вінницький національний технічний університет, факультет комп'ютерних систем та автоматики, Вінниця, Україна;* [2] *Universidad Estatal Peninsula de Santa Elena (UPSE), Santa Elena, Ecuador;* [3] *Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación SENESCYT, Ecuador*

Інформаційна технологія координації багатолінійного технологічного процесу

Розглядається інформаційна технологія координації роботи багатолінійного технологічного процесу, яка ґрунтується на використанні генетичного алгоритму, цільова функція якого враховує як економічний ефект, так і синхронізацію роботи цих ліній.

На багатьох виробничих підприємствах використовуються просторово розподілені технологічні об'єкти, які можуть мати різні цілі й обмеження, параметри продуктивності та ефективності.Управління їх спільним функціонуванням зумовлює необхідність вирішення проблеми координації. Як правило, розглядаються переважно ітераційні та неітераційні алгоритми координації, проте не розглянута задача розподілу ресурсів у взаємозв'язку з завданням синхронізації паралельних технологічних процесів процесів [1, 2], що зумовлює *актуальність* роботи.

Об'єктом роботи є процес координації технологічних ліній, який забезпечує мінімальні втрати часу при виробництві.

Предметом роботи є методи координації технологічних ліній, які забезпечують мінімальні втрати часу при виробництві.

Метою роботи є підвищення ефективності управління за рахунок удосконалення методу координації роботи технологічних ліній.

Постановка задачі. Розглядається паралельна робота декількох технологічних ліній при спільній останній операції(Рисунок 1). Як критерій ефективності координації роботи системи управління, що враховує як економічні так і технічні параметри, використовується інтенсивність отримання прибутку:

$$E = \frac{P - C}{T}, \tag{1}$$

де P – ціна партії продукції; C – витрати на виробництво партії; T – тривалість процесу виробництва партії.

Розв'язання. Ціна партії товару визначається

$$P_i = S_i \cdot R_i, \tag{2}$$

де S_i – ціна одиниці продукції i-ої лінії, R_i - кількість ресурсу i-ої лінії.

Загальний об'єм сировини розподіляється між технологічними лініями

$$R_\Sigma = \sum_{i=1}^{n} R_i. \tag{3}$$

Час виконання технологічного процесу залежить від кількості сировини, призначеного кожній лінії, і її технічних параметрів. Таким чином, тривалість технологічного процесу t_{p_i} кожної лінії

$$t_{p_i} = \frac{R_i}{p_i}, \tag{4}$$

де p_i– продуктивність i-ої лінії.

З іншої сторони, тривалість технологічного процесу

$$t_{p_i} = t_{f_i} - t_{o_i}, \tag{5}$$

де t_{f_i} - час завершення технологічного процесу (початок), t_{o_i} - час початку технологічної операції (тривалість простою) [3]. Оскільки, продукція всіх технологічних ліній надходить на одну ділянку, ці операції виконуються послідовно. Стрічковий графік представлений на рисунку 2.

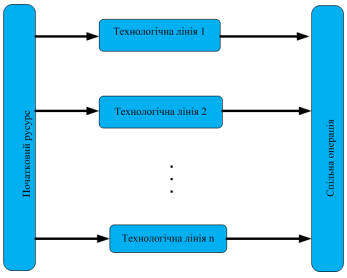

Рис. 1. Схема координації паралельних технологчних ліній з спільною операцією

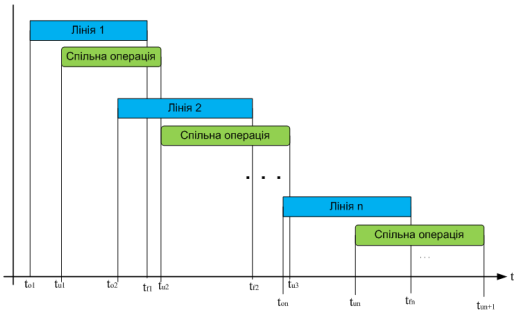

Рис. 2. Стрічковий графік роботи технологічних ліній із спільною операцією

Цільова функція для n технологічних ліній залежить від змінних

$$F(R_1, R_2, \ldots, R_n, t_{o_1}, t_{o_2}, \ldots, t_{o_n}).$$

Розглядаючи паралельну роботу технологічних ліній, використовуються методи: безітеративні алгоритми, ітеративні алгоритми, генетичні алгоритми.

При безітеративному алгоритмі оптимальний результат досягається у результаті одноразового обміну між рівнями. Недоліком алгоритму є висока складність при обчисленнях.

За ітеративного алгоритму оптимальне рішення знаходиться завдяки багаторазовому обміну інформації між центром та елементами. Недоліком є велика кількість ітерацій, і як наслідок високий час обчислень [2].

Генетичні алгоритми – це алгоритм пошуку, що використовується для вирішення задач оптимізації і моделювання шляхом послідовного підбору, комбінування і варіації шуканих параметрів. Головним при координації за генетичним алгоритмом є вибір функції пристосованості [3].

Через нелінійність цільової функції (1), що випливає з формул (2), (4) було обрано генетичний алгоритм.

Висновок. Удосконалено модель управління паралельними технологічними лініями із спільною операцією та метод координації технологічних процесів, який враховує розподіл ресурсів і синхронізацію.

Література. **1.** Месарович М. Теория иерархических многоуровневых систем./ Месарович М., Мако Д., Такахара И. – М., 1973. — 344 с. **2.** Системна задача координації в технологічних комплексах неперервного типу [Електронний ресурс] / А.П. Ладанюк, Д.А. Шумигай, Р.О. Бойко – Режим доступу : URL: http://dspace.nuft.edu.ua/jspui/bitstream/123456789/4444/1/Sh_3.pdf. **3.** Розробка програмного модуля для підвищення ефективності координації виробничих процесів молочного виробництва / В.М. Дубовой, М.Е. Дуда, М.М. Байас // Вимірювальна та обчислювальна техніка в технологічних процесах – 2014 – № 4 - С.128-134.

Дубовой В.М., Пилипенко І.В., Стець Р.С.
Вінницький національний технічний університет, Вінниця, Україна

Моделювання процесу тестування програмного забезпечення як розгалужено-циклічного технологічного процесу

З невпинним розвитком та поширенням інформаційних технологій у всіх сферах людської життєдіяльності з'являються нові та складніші завдання перед розробниками програмного забезпечення. Програмні продукти стають складнішими, багатокомпонентними і вимагають спеціалізованого підходу для їх розроблення та перевірки. У зв'язку з цим, підвищуються вимоги до надійності програм, виникає потреба у скороченні затрат на тестування тау прогнозуванні надійності розроблюваного програмного забезпечення (ПЗ) [1].

Відомо, що для прогнозування надійності програмного забезпечення широко використовують тип моделей, які описують характеристики ПЗ на основі кількості виявлених помилок у програмному продукті [2]. Тому, для оцінки та прогнозування надійності ПЗ необхідно досліджувати, вдосконалювати та аналізувати моделі та методи процесу тестування.

Актуальність. Існує безліч моделей тестування програмного продукту. Даною темою займається багато сучасних дослідників, таких як Яковина В.С., Сенів М.М., Чабанюк Я.М. [3]; Durand J.B., Gaudoin O. [4]; Yamada S., Ohba M., Osaki S. [5] та ін.

Важливою складовою кожної моделі тестування програмного продукту є критерій достатності процесу тестування, який дозволяє керівникам проектів приймати обґрунтовані рішення про завершення даного етапу розробки. Принципова характеристика процесу тестування програмного забезпечення полягає у циклічності процесу, а саме в тому, що в програмі за період виконання кожного циклу тестування виявляються і виправляються помилки, які не були знайдені раніше.

В даний час у переважній більшості IT компаній такі показники носять формальний характер та затрати на тестування намагаються зменшити, що призводить до зниження якості продукту. Необхідно будувати модель процесу тестування, щоб спрогнозувати затрати на процес тестування, враховуючи його циклічну повторюваність. Саме це і обґрунтовує актуальність проблеми моделювання процесу тестування як розгалужено-циклічного технологічного процесу.

Метою роботи є прогнозування часу тестування програмного продукту як розгалуженоциклічного технологічного процесу.

Задачі дослідження.

1. Аналіз процесу тестування програмного забезпечення як розгалужено-циклічного технологічного процесу.
2. Побудова моделі процесу тестування як розгалужено-циклічного технологічного процесу
3. Застосування розробленої моделі для прогнозування часу тестування програмного продукту.

Побудова моделі процесу тестування як розгалужено-циклічного технологічного процесу. В [4] авторами запропонована модель управління циклічними розгалуженими технологічними процесами, яка ґрунтується на невизначених циклічних графах і неоднорідних марковських ланцюгах. Побудуємо таку модель для процесу тестування ПЗ.

Представимо застосування марковської моделі РТП до процесу тестування ПЗ, враховуючи перетворення циклічного графа на еквівалентний йому ациклічний. За основу візьмемо неоднорідну марковську модель [4]:

$$b_v(X_j) = \sum_{i=1}^{m} \sum_{h=1}^{v+1} \{ \sum \dots \sum^{h-1} c_1 c_2 \prod_{r=0}^{h-1} b_{v-h}(X_i) \}$$

де $c_v^{00}(X_i, X_j) = c_{v-1}^{00}(X_i, X_j) = c_1$ - коефіцієнт впливу v-го підпроцеса з i-го в j-й стан без урахування впливу інших підпроцесів;

$c_v^{v-1,h}(X_i, X_j) = c_{v-1}^{v-2,h}(X_i, X_j) = c_2$ - коефіцієнт впливу v-1-го підпроцеса з i-го в j-й стан з урахування попереднього v-2-го підпроцеса з h-м станом;

X_i, X_j - номер стану операції;

$b_v(X_j)$ - вірогідність стану;

m - кількість станів;

v - кількість підпроцесів;

h - номер стану;

b_{v-h} - вірогідність v-2-го підпроцеса з h-м станом.

Висновки. Аналіз процесу тестування програмного забезпечення показав, що даний процес можна віднести до розгалужено-циклічних технологічних процесів, оскільки виконується циклічно з прийняттям рішень на контрольних операціях.

Застосована модель розгалужено-циклічного технологічного процесу до процесу тестування ПЗ дає змогу прогнозувати час на виконання процесу. Складність застосування визначається тим, що існують стани, котрі не можливо одночасно віднести до бажаних чи небажаних. Кожний модуль, що тестується має свої характеристики та потребує окремого прогнозування.

Запропонована методика дозволяє прогнозування для кожного окремого модуля програмного продукту, що призводить для якіснішого прийняття рішень на кожній контрольованій операції всього підпроцесу.

Література. **1.** Процедура стохастичної оптимізації для моделі тестування з напівмарковськими переключеннями http://irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?C21COM=2&I21DBN=UJRN&P21DBN=UJRN&IMAGE_FILE_DOWNLOAD=1&Image_file_name=PDF/Mtkm_fiz_mat_2014_10_13.pdf. **2.** Моделі і методи тестування програмних систем http://dspace.nbuv.gov.ua/bitstream/handle/123456789/289/%D0%9A%D0%BE%D1%80%D0%BE%D1%82%D1%83%D0%BD_%231.pdf. **3.** Яковина В.С., Сенів М.М., Чабанюк Я.М., Федасюк Д.В., Хімка У.Т. Критерій достатності процесу тестування програмного забезпечення. Вісник Національного університету "Львівська політехніка" Комп'ютерні науки та інформаційні технології (2010). **4.** Прийняття рішень в управлінні розгалуженими технологічними процесами : монографія / В. М. Дубовой, Г. Ю. Дерман, І. В. Пилипенко, М. М. Байас. — Вінниця : ВНТУ, 2014. — 216 с. **5.** Дубовой В.М., Пилипенко І.В., Стець Р.С. Застосування марковської моделі для аналізу впливу циклічності на управління розгалуженим технологічним процесом. Наукові праці ВНТУ, 2014, № 4 [Електронний ресурс] – Режим доступу: http://praci.vntu.edu.ua/article/view/3827/5583. **6.** Хенрик Книберг. Scrum и Хр: заметки с передовой [Електронний ресурс] - Режим доступу: www.dbases.ru/cgi-bin/catalog/catalog.cgi?i=1560.

Душин О.С., Сирота С.В.
Національний технічний університет України "КПІ", факультет прикладної математики, Київ, Україна

Модель синхронізації роботи адаптивних систем регулювання дорожнього руху на суміжних перехрестях

Нагальною проблемою багатьох сучасних міст є оптимальне керування дорожнім рухом. В даній роботі пропонується програмна модель мережі керованих перехресть, що базується на багатоагентному підході. Агентами виступають моделі окремих керованих перехресть. Топологія мережі задається графом, який представляється списком ребер.

Модель адаптивної системи регулювання дорожнього руху на чотирьох направленому перехресті з використанням нечіткої логіки запропонована у [1]. Розроблено програмне забезпечення, яке реалізує дану модель. Спроектована модель системи дозволяє розраховувати пропускну здатність перехрестя із різними модулями керування фазами руху як із фіксованим часом перемикання, так і ті, що базуються на принципах нечіткої логіки та нейронних мережах [2-4]. Даний програмний засіб пропонується використовувати для реалізації перехрестя як агента.

Ідея полягає в тому, що перехрестя розглядається за принципом чорної скриньки із 12 входами і 4 виходами (рис. 1). Значення на кожному вході означає кількість транспортних засобів (ТЗ) в черзі на відповідному в'їзді на певному напрямку в момент перемикання фази руху, значення на виході - кількість ТЗ на кожному виході. Значення певного виходу одного перехрестя додаються до значень відповідної групи входів іншого.

Вільні входи відповідають в'їздам на транспортний вузол, а виходи – виїздам. На входи подаються значення із модулю розподілу за стохастичними принципами. Математичним представленням уніфікованої моделі транспортного є орієнтований граф. Такий підхід дозволяє створити уніфіковану модель транспортного вузла.

Для програмної реалізації передбачається використати засоби багатопотокової архітектури. Кожен з агентів повинен працювати в реальному часі.

Рис. 1. Представлення моделі перехрестя за принципом чорної скриньки

Висновки. Спроектовано уніфіковану модель синхронізації роботи адаптивних систем регулювання дорожнього руху на суміжних перехрестях. Дана модель дозволить побудувати віртуальну дорожню мережу, що дозволить досліджувати та покращувати дорожню ситуацію для всієї транспортної інфраструктури в цілому.

Література. **1.** Сирота С.В., Душин О.С. Модель адаптивної систем регулювання дорожнього руху на перехресті з використанням нечіткої логіки // Прикладна математика та комп'ютинг. ПМК, 2014 : сьома наук. конф. магістрантів та аспірантів, Київ, 15—17 квіт. 2015 р. : зб. тез доп. / [редкол.: Дичка І. А. та ін.]. — К. : Просвіта, 2015. **2.** Yan Ge. A Two-Stage Fuzzy Logic Control Method of Traffic Signal Based on Traffic Urgency Degree // Modelling and Simulation in Engi-neering. – 2014. P. 1–6. **3.** Papas C. P., Mamdani E. H. A fuzzy logic controller for a traffic junction // IEEE Transactions on Systems, Man and Cybernetics. – 1977. – Vol. 7, – P. 707–717. **4.** Niittymäki J., Turunen E. Traffic signal control on similarity logic rea-soning // Fuzzy Sets and Systems, 2003. – P. 101-140.

Забелин С.И.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Метод раннего распознавания DDoS атак при помощи SDN

SDN - сети будущего. Программно-конфигурируемые сети (ПКС, англ. Software Defined Networks - SDN) – это актуальный тренд в сетевых технологиях, в которой уровень управления сетью отделён от устройств передачи данных и реализуется программно. Непосредственно пересылку данных выполняют специальные коммутаторы SDN на основе универсальных таблиц потоков, которые формируются и поддерживаются внешним управляющим компонентом - контроллером сети. Ключевая особенность управления SDN в том, что как таблицы потоков в коммутаторах, так и правила функционирования контроллеров доступны для управления внешними программными средствами через специализированные API. [1] Из-за централизованной структуры SDN может стать потенциальной целью злоумышленников. [2]

SDN и сетевые атаки. DDoS-атака (атака типа «отказ в обслуживании», от англ. Distributed Denial of Service) - один из многочисленных видов атак на вычислительную систему. Цель данной атаки - парализовать работу атакуемого веб-узла. [3]

Для эксперимента на стенде использовалась атака насыщением полосы пропускания и на истощение ресурсов. Стандартный контроллер не имеет защит от DDoS-атак и с помощью API был модифицирован и назван ProtoPoxM.

Методика исследования. Для исследования был выбран контроллер SDN – POX. Контроллер имеет поддержку высокоуровневых API SDN, такие как виртуализация и обнаружение топологий. POX реализован на языке Python. [4]

Экспериментальные исследования были организованы на основе эмулятора сетей Mininet и контроллера POX. Организация DDoS-атаки была смоделирована на разных топологиях построения сети и разном количестве сетевого оборудования.

Для эксперимента использовалась атака насыщением полосы пропускания UDP-флуд при различных топологиях сети.

Модифицированный контроллер ProtoPoxM. Стандартный контроллер POX не имеет защит от DDoS-атак и с помощью API был модифицирован. Результатом исследования стал модифицированный контроллер ProtoPoxM, в котором устранены многие уязвимости стандартной версии POX. В ProtoPoxM реализованы функции анализа сетевого трафика внутри сети и на основе полученных данных обнаружения подсети которая начала DDoS-атаку, и впоследствии ProtoPoxM создаёт фильтры которые, предотвращают возможные DDoS-атаки.

Исходя из результатов экспериментальных исследований ProtoPoxM на незащищенном и защищенном контроллерах SDN в среде Mininet был сделан вывод о его полной работоспособности.

Выводы. Используя преимущества SDN, а именно централизацию управления, гибкость, программное конфигурирование позволяют создавать высокозащищенные контроллеры, что является важным шагом в повышении безопасности сетей. Была разработана методика защиты от DDoS-атак и проведены экспериментальные исследования, которые были направлены на подтверждение применения разработанных методов и инструментов в практической работе.

Литература. **1.** T. Nadeau and K. Gray, *SDN: Software Defined Networks*, Washington, O'Reilly Media, 2013. **2.** T. Limoncelli. (2012, June 20). *Openflow: a radical new idea in networking* [Online]. Available: http://queue.acm.org/detail.cfm?id=2305856 **3.** Wikipedia contributors. (2014, February 10). *DoS-атака* [Online]. Available: http://bit.ly/1Hv1eei **4.** Ali Al-Shabibi. (2014, January 25). *POX Wiki* [Online]. Available: https://openflow.stanford.edu/display/ONL/POX+Wiki

Кислий Р.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Моделювання профілю користувача для уникнення проблеми холодного старту в рекомендаційних системах

З рекомендаційними системами (РС) звичайний користувач інтернету зустрічається практично щодня — майже в будь-якому інтернет магазині існує система, яка показує користувачу речі, які б могли його зацікавити. Це може бути як проста система що показує речі, найбіль подібні до тих, які переглядає користувач, так і складна система, що враховує наявну інформацію про користувача: його історію покупок, переглядів, схожість з іншими користувачами.

РС поділяються на 2 типи — рекомендації по контенту, та рекомендації за допомогою колаборативної фільтрації (метод, що використовує оцінки групи користувачів для прогнозування невідомої оцінки іншого користувача) [1]. Проте не всі рекомендації цінні — користувач, що має інші смаки, не є цінним джерелом рекомендацій.

Як показує досвід, перед здійсненням покупки, або використанням сервісу, люди зазвичай радяться з друзями та родичами, тому рекомендації від таких користувачів є цінними. Популярні соціальні мережі, такі як Facebook чи Twitter, створюють безліч нових шляхів для людей, щоб спілкуватися і будувати віртуальні спільноти. Соціальні мережі не тільки спростили можливість для користувачів ділитися своїми оцінками, але й можуть бути використані як додаткове джерело даних для рекомендаційних систем.

Сьогодні, багато соціальних мереж також пропонують концепцію "кіл друзів" та спільнот, які виконують функцію кластеризації. Очевидно, що у користувача різна довіра до друзів з різних "кіл" та спільнот. На жаль, в більшості існуючих мульти-рейтингових датасетів, соціальні зв'язки користувача з усіх категорій змішуються один з одним. В дослідженні проводиться спроба надати різної ваги користувачам, при враховуванні їх рекомендацій, в залежності від їх належності до певного "кола друзів", тої чи іншої спільноти користувача [2].

Це додаткове джерело не тільки допомогає покращити точність рекомендацій, а й частково уникнути проблеми "холодного старту" (за наявності достатньої кількості інформації у профілі користувача), за допомогою передбачення вподобань користувача. Робота над покращенням точності рекомендацій в основному спрямована на мінімізацію середньоквадратичної помилки (СКП). СКП обчислюється за формулою (1), де X — значення, розраховане за допомогою моделі, а X_t — значення, отримане емпірично.

$$\sqrt{\frac{\sum_{i=1}^{n}(x_i - x)^2}{n}} \tag{1}$$

Для дослідження вибраний відкритий датасет з Kaggle[3], мова python для збору статистики з соціальних мереж та мова R для проведення аналізу даних. Такий вибір зумовлений тим, що R надає швидкі та зручні інструменти для обробки даних. Розроблюваний сервіс являє собою рекомендаційну систему, що зменшить похибку при рекомендації послуг користувачам, за рахунок інтеграції з соціальними мережами і побудови профілю користувача.

Подальшим дослідженням буде сентиментальний аналіз дій та дописів користувача в соціальних мережах, що потенційно дозволить ще більше зменшити похибку при рекомендаціях.

Література. **1.** An Introduction to Statistical Learning [Електронний ресурс] / Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani - Режим доступу: http://www-bcf.usc.edu/~gareth/ISL/ISLRFourthPrinting.pdf **2.** Trust-based recommendation systems: an axiomatic approach [Електронний ресурс] / Reid Andersen, Christian Borgs, Jennifer Chayes - Режим доступу: http://research.microsoft.com/en-us/um/people/borgs/Papers/trust.pdf **3.** Kaggle: The Home of Data Science [Електронний ресурс] - Режим доступу: http://www.kaggle.com

Ковальчук-Химюк Л.О., Тавров Д.Ю.
Національний технічний університет України "КПІ", Київ, Україна

Система нечіткого виведення для оцінювання успішності учнів під час білінгвального навчання математиці

Сучасний світ розвинених комунікацій різноманітний і багатомовний, тому система освіти кожної розвиненої країни повинна підготувати своїх учнів до життя в багатомовному середовищі. Сучасний громадянин світу у своїй життєдіяльності регулярно користується мережею Інтернет і володіє принаймні трьома іноземними мовами. Таким чином, можна стверджувати, що головною рисою середньої освіти на початку XXI століття стає багатомовність у поєднанні з комп'ютеризацією навчального процесу на всіх рівнях.

У даній роботі розглядатимемо галузь білінгвального навчання математиці. Окрім оволодіння, власне, математичними знаннями, уміннями та навичками, учні повинні вміти чітко, ясно й грамотно пояснювати вибір того або іншого розв'язку деякої задачі як усно, так і письмово. Для формування таких навичок використовують спеціальні шаблони, які допомагають навчитися використовувати математичну лексику під час усної відповіді або в письмовій роботі.

Можна сформувати такі задачі білінгвального навчання математиці:

- формування та вдосконалення лінгвістичної та математичної компетенцій учня;
- розвиток умінь використання іноземної мови для одержання математичної інформації з різних джерел;
- оволодіння знаннями з математики з використанням двох мов;
- розвиток інтелектуальних здібностей, пізнавальних можливостей, творчого потенціалу;
- формування та вдосконалення рівня мотивації до оволодіння предметними математичними знаннями та іноземною мовою.

З урахуванням вищезазначеного, актуальною є задача розроблення математичних моделей оцінювання успішності учнів під час білінгвального навчання математиці. Такі моделі можна використати для спрощення роботи вчителя та зменшення рівня суб'єктивізму під час оцінювання учнів, а також під час розроблення спеціалізованих експертно-навчальних систем.

Оцінити рівень навчальних досягнень окремо взятого школяра часто буває непросто, оскільки оцінка залежить від низки факторів, оцінювання більшості з яких має суб'єктивний характер. У роботі [1] запропоновано підхід до оцінювання успішності білінгвальних учнів шляхом використання апарату нечіткої логіки. Вхідними змінними є так звані вхідні характеристики, які, на нашу думку, не враховують усіх задач білінгвального навчання математиці. Окрім того, оцінювання таких характеристик не позбавлено надлишкової суб'єктивності.

У даній роботі пропонується оцінювати загальну успішність білінгвальних учнів за допомогою системи нечіткого виведення (СНВ), вхідними лінгвістичними змінними якої [2] є рівні володіння низкою компетенцій, які можна розділити на два класи:

- *математичні компетенції* — уміння пояснити етапи розв'язання іноземною мовою (Y_1), уміння пояснити етапи розв'язання рідною мовою (Y_2), уміння розв'язувати задачу (Y_3);
- *мовні компетенції* — розуміння умови задачі (Y_4), уміння пояснити суть задачі іноземною мовою (Y_5), уміння пояснити суть задачі рідною мовою (Y_6).

Відповідно до ступеня оволодіння знаннями й способами діяльності виокремлюють чотири рівні навчальних досягнень школярів із деякої дисципліни [3]: початковий, середній (С), достатній (В), високий (А). Початковий рівень у пропонованій моделі не розглядається, оскільки навчання в білінгвальной секції потребує щонайменше середнього рівня знань як із математики, так і з іноземної мови. Таким чином, вхідні змінні Y_i мають значення А, В та С, визначені на проміжку $[0,1]$ функціями належності:

$$\mu_A(x) = e^{\left(-\frac{(x-0,0)^2}{2\cdot 0,2^2}\right)}, \quad \mu_B(x) = e^{\left(-\frac{(x-0,5)^2}{2\cdot 0,2^2}\right)}, \quad \mu_A(x) = e^{\left(-\frac{(x-1,0)^2}{2\cdot 0,2^2}\right)}. \tag{1}$$

Вихідною змінною Z такої СНВ є результуюча оцінка учня, яка має значення A, B та C, визначені на проміжку $[4, 12]$ функціями належності:

$$\mu_A(x) = e^{\left(-\frac{(x-4,0)^2}{2\cdot 1,4^2}\right)}, \quad \mu_B(x) = e^{\left(-\frac{(x-8,0)^2}{2\cdot 1,4^2}\right)}, \quad \mu_A(x) = e^{\left(-\frac{(x-12,0)^2}{2\cdot 1,4^2}\right)}. \tag{2}$$

Нечіткі продукційні правила для такої СНВ наведено в Табл. 1.

Табл. 1. Нечіткі продукційні правила для СНВ визначення оцінки учня

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Z	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Z
C	C	A	A	A/B/C	C	B	C	B	C	A	B/C	C	C
C	B	C	B/C	C	C	C	C	B	B	A/B	B	B/C	B
C	B	A	B	B	B	A	C	B	A	A	B	C	A
C	B	A	A	A/C	C	B	C	B	B/C	A	A	A/B/C	B
C	B	A/C	A	B	B	B	C	A	C	C	C	C	C
C	A	B	A/B/C	C	C	C	C	A	A	A	A	A/B/C	A
C	A	A	A/B	B/C	C	B	C	B/C	B	C	C	C	C
C	B/C	B	A/B	C	C	B	C	B/C	A	B	B/C	C	B
C	B/C	A	A	A	A/B	A	C	A/C	C	B	B/C	C	C
C	A/C	C	A	A	A/B	B	C	A/C	C	A	A/B/C	C	C
C	A/C	C	A/B	B	B	B	C	A/C	B	B	B	B/C	B
C	A/C	B	A	A	A	A	C	A/C	B	A	A/B	B/C	B
C	A/C	A	A/B	B	B	A	C	A/B/C	A	C	C	C	B
B	B	C	A	B	B	B	B	B	C	A	A/B/C	C	C
B	B	B	B	B	B	A	B	B	B	B	B/C	C	B
B	B	B	A	B	B/C	B	B	B	B	A	A	A/B/C	B
B	B	B	A/C	C	C	B	B	B	A	A	A	A/B/C	A
B	B	A	A/B	B	B/C	A	B	B	A	A/B/C	C	C	B
B	A	C	A	A/C	C	C	B	A	C	A/B	B	B/C	C
B	A	B	B	B	B/C	B	B	A	B	A	B	B	A
B	A	B	B/C	C	C	B	B	A	A	C	C	C	B
B	A	A	B	B/C	C	B	B	A	A	A/B	B	B	A
B	A/B	C	A	A	A/B	B	B	A/B	C	B/C	C	C	C
A	A	C	B	B	B/C	C	A	A	C	A	B	B	B
A	A	C	A	A	A/B	B	A	A	C	A	A/B/C	C	C
A	A	C	B/C	C	C	C	A	A	A/B	C	C	C	B
A	A	A/B	B	B/C	C	B	A	A	A/B	A/B	B	B	A
B/C	B	C	B	B	B/C	C	A/B	A	A/B	A	A	A/B	A
A/B	A	A/B	A	A/B/C	C	B							

Оцінити рівні володіння компетенціями Y_i безпосередньо майже неможливо. Для мінімізації суб'єктивного впливу вчителя пропонується оцінювати такий рівень непрямим чином шляхом застосування окремої СНВ, вхідними змінними X_i якої є оцінки учня за виконання певних видів робіт, які можна оцінити за звичною 12-бальної шкалою. Такими роботами є: вставити пропущене слово (X_1), переформулювати фразу (X_2), сформулювати властивості іноземною (X_3) та рідною (X_4) мовами, вибрати слово зі списку (X_5), розв'язати задачу (X_6), пояснити розв'язок іноземною (X_7) та рідною (X_8) мовами, виділити зайві дані (X_9), вибрати зі списку питання, на які можна відповісти, виходячи з умов задачі (X_{10}), проаналізувати розв'язок іноземною (X_{11}) та рідною (X_{12}) мовами. Вхідні змінні X_i мають значення A, B та C, визначені на проміжку $[4, 12]$ функціями належності (2). Вихідними змінними Y_j СНВ є рівні володіння відповідними компетенціями (вони ж одночасно є вхідними змінними першої СНВ). Нечіткі

продукційні правила для такої СНВ наведено в Табл. 2.

Табл. 2. Нечіткі продукційні правила для СНВ визначення рівней компетенцій учня

(a) Компетенція Y_1

X_3	X_7	X_{11}	Y_1
C	C	A	B
C	C	B/C	C
C	B	C	C
C	B	B	B
C	B	A	A
C	A	C	B
C	A	A/B	A
A/B	C	C	C
A/B	C	B	B
A/B	C	A	A
A/B	A/B	C	B
A/B	A/B	A/B	A

(b) Компетенція Y_2

X_4	X_8	X_{12}	Y_2
C	A	C	C
C	A	A/B	B
C	B/C	A	B
C	B/C	B/C	C
B	C	A/B/C	C
B	B	C	C
B	B	A/B	B
B	A	A/B/C	B
A	C	A	B
A	C	B/C	C
A	A/B	A/B/C	B

(c) Компетенція Y_3

X_6	X_9	X_{10}	Y_3
C	C	A/B/C	C
C	B	A	B
C	B	B/C	C
C	A	C	C
C	A	A/B	B
B	A	A	A
B	A	B/C	B
B	B/C	A/B/C	B
A	C	C	B
A	C	A/B	A
A	B	A	A
A	A/B	A/B/C	A

(d) Компетенція Y_6

X_4	X_{12}	Y_6
C	A/B/C	C
A/B	C	C
B	A/B	B
A	B	B
A	A	A

(e) Компетенція Y_4

X_1	X_5	X_9	X_{10}	Y_4	X_1	X_5	X_9	X_{10}	Y_4
C	C	B	A	B	C	C	B	B/C	C
C	B	A/B	C	C	C	B	A/B	A/B	B
C	A	B	C	C	C	A	B	A/B	B
C	A	A	A	A	C	A	A	B/C	B
C	A/C	C	A/B/C	C	B	C	B/C	A/B/C	C
B	B	B	C	C	B	B	B	A/B	B
B	B	A	A	A	B	B	A	B/C	B
B	A	C	C	C	B	A	C	A/B	B
B	A	A/B	C	B	B	A	A/B	A/B	A
A	C	C	B	B	A	C	C	A/C	C
A	C	A/B	C	C	A	C	A/B	A/B	B
A	A/B	C	A/B/C	B	A	A/B	A/B	C	B
A	A/B	A/B	A/B	A	B/C	C	A	C	C
B/C	C	A	A/B	B	B/C	B	C	A	B
B/C	B	C	B/C	C					

(f) Компетенція Y_5

X_2	X_3	X_{11}	Y_5
C	C	A/B/C	C
C	A/B	C	C
C	A/B	A/B	B
B	B	A	A
B	B	B/C	B
B	A	C	B
B	A	A/B	A
B	A/B	C	C
B	A/B	A/B	B
A	A/B	C	C
A	A/B	C	B
A	A/B	A/B	B
A	A/B	A/B	A

Таким чином, оцінити загальний рівень успішності білінгвального учня, який вивчає математику, можна за допомогою поєднання двох СНВ. На нашу думку, запропонована структура СНВ дозволяє як урахувати задачі білінгвального навчання, так і мінімізувати рівень суб'єктивності під час оцінювання успішності учня.

Література. **1.** Рзаев Р. Об одном подходе к формированию методики билингвального обучения / Р. Р. Рзаев, З. Р. Джамалов, Н. А. Умарова // Математичні машини і системи. — 2014. — №4. — С. 92–103. **2.** Zadeh L. A. The Concept of a Linguistic Variable and its Application to Approximate Reasoning / L. A. Zadeh // Information Sciences. — 1975. — 8. — P. 199–249. **3.** Про затвердження орієнтовних вимог оцінювання навчальних досягнень учнів із базових дисциплін у системі загальної середньої освіти : Наказ Міністерства освіти і науки України від 21.08.2013 №1222.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Кондіус К.Ю.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Наукові публікації та їх цитування в галузі інтелектуальних технологій

Оцінка та позиційне позиціонування окремих патентів або портфеля патентів є важливою і складною задачею. Проведені дослідження свідчать про те, що одним з найбільш важливих показників, який широко використовується при визначені вартості пакету та його цінності є цитування патентів [1].

Існуючі дослідженні в цій сфері були зосереджені на загальній кількості патентних цитат, які отримав конкретний патент, і практично не відрізняючи якість та значущість таких посилань, їх вплив на вартість патенту. Зокрема, практично не звертається увага дослідників на наступні положення:

1. Дане посилання (цитата) належить одному і тому ж заявнику, тобто автором заявки з якої йде цитата і заявки, в якій приводиться ця цитата, є одна й та сама особа (інакше кажучи, заявник цитує сам себе – самоцитування);
2. Дане посилання належить одному і тому ж класу технологій, чи даний патент цитує патентну заявку з іншого.

В цьому зв'язку можна припустити наступне. Патентне самоцитування означає, що дана фірма або організація намагається розвивати таке технічне рішення, за рахунок нових патентів розширює сферу використання цього патенту, інвестує його науково-технологічний розвиток. Це свідчить про наявну економічну цінність цього патенту для даної організації, що, без сумніву, може зацікавити конкуруючі організації і таким чином вплинути на рівень комерційного потенціалу патенту.

Метою дослідження стало вивчення динаміки розвитку сучасного наукового знання на прикладі збільшення наукових публікацій та їх цитування в галузі інтелектуальних технологій; визначення ролі міжнародних індексів цитування патентів, як ефективного науково метричного інструменту, що впливає на покращення наукового співробітництва, поширення знання, показників престижу наукової установи, вартісної оцінки патенту та потенціалу його комерціалізації.

Пропонується класифікувати патентні цитати з урахуванням наведених вище напрямків на 4 відносно самостійні групи.

1. Просте самоцитування – патент цитується одним і тим же автором в тому ж самому технологічному класі.
2. Стандартне цитування – патент цитується іншим автором в тому ж самому технологічному класі.
3. Широке самоцитування – патент цитується одним і тим же автором в тому ж самому технологічному класі.
4. Інноваційне цитування – патент цитується іншим автором, але вже в іншому технологічному класі.

Пропонована структура цитування патентів, на наш погляд, має велике значення для визначення вартості патенту. Вказані чотири типи цитат мають різні наслідки при визначенні вартості патенту, що передбачає необхідність проводити зважену оцінку цитування. На нашу думку, зважені патентні цитати є більш об'єктивним показником цінності патенту та його рейтингу на ринку ОІВ, чим показник простої кількості цитат.

Література. **1.** Булгаков В.В. Мультидисциплінарні індекси цитування – ефективний інструмент оцінки наукового знання у 21 сторіччі / В.В. Булгаков, М.Ю. Антомонов // Гігієна населених місць. - № 61. – 2013. – С. 415-424. **2.** Горбунов-Посадов М.М. Интернет-активность как обязанность ученого / М.М. Горбунов-Пасадов // Информационные технологии и вычислительные системы. – 2007. - №3. – С. 88-93. **3.** Игами М. Современное состояние сферы нанотехнологій: анализ патентов / м. Игами, Т. Оказаки // Форсайт. – 2008. - №3(7). – С.22-31.

Копычко С.Н., Кулык Р.Ю., Дрозденко А.Н.
Национальный технический университет Украины "КПИ", Киев, Украина

Исследование XML парсеров языка программирования Java

Extensible Markup Language (XML) - простой, гибкий текстовый формат, который имеет четкую структуру, расширяем и легко читается человеком. Первоначально разработан для решения задач крупномасштабных электронных публикаций. XML также играет все более важную роль в обмене разнообразных данных в Интернете и в других местах [1].

В XML файле хранятся данные. Программист очень часто сталкивается с задачей обработки XML файла. Например, для программы, зависящей от курсов валют, информация о состоянии курса на текущий момент может приходить именно в таком файле. Перед ним ставится задача автоматического разбора файла и получение из него данных для дальнейшей обработки.

В языке Java существует несколько XML парсеров. Основные – это SAX (Simple API for XML) и DOM (Document Object Model), используемые чаще всего. Также существуют StAX (Streaming API for XML), JDOM, DOM4j парсеры.

SAX парсер отлично подходит для чтения данных из XML файла. Он рассматривает документ сверху-вниз и вызывает функции, составленные программистом для обработки типов данных файла, передавая им найденные данные. Преимущество этого парсера состоит в том, что он требует мало памяти, так как он динамически просматривает документ, не загружая предварительно документ полностью в память.

В отличие от SAX DOM парсер загружает XML файл полностью. В памяти файл представляется в виде дерева объектов. Поэтому скорость чтения файла и количество памяти напрямую зависит от размера документа. Этот парсер позволяет не только читать, а еще и изменять документ.

StAX фокусируется на поточной обработке. По сути именно возможность приложений обрабатывать данные XML в виде потока событий отличает стандарт StAX от других подходов. Идея обработки XML как набора событий не слишком свежая (по сути, она уже представлена в стандарте SAX); однако в StAX программный код может запрашивать такие события одно за другим вместо того, чтобы предоставлять обработчики, которые получали бы события от синтаксического анализатора в определяемом самим анализатором порядке [2]. Также StAX может производить запись в файл, и как и SAX расходует мало памяти.

JDOM может использоваться как альтернатива пакету DOM для программного манипулирования XML-документами. Это не обязательная замена, фактически JDOM и DOM могут счастливо сосуществовать. Кроме того JDOM не занимается разбором исходного текста XML, хотя он обеспечивает классы-оболочки, которые берут на себя большую часть работы по конфигурированию и выполнению реализации парсера. В JDOM нет иерархий, то есть элемент XML является экземпляром класса Element, атрибут XML является экземпляром класса Attribute, а сам XML-документ является экземпляром класса Document. Поскольку все они представляют разные концепции в XML, они всегда представляются собственными типами, а не как аморфные "узлы" [3].

В данной работе исследован DOM парсер для разбора XML файлов, имеющих небольшой размер, но глубокую древовидную структуру, благодаря чему можно проще добираться до конечных элементов дерева. Он предоставляет куда большие возможности по обработке данных из исходного документа.

Литература. **1.** Extensible Markup Language (XML) [Электронный ресурс] - Режим доступа: http://www.w3.org/XML/ **2.** Использование StAX для обработки XML [Электронный ресурс] - Режим доступа: http://www.ibm.com/developerworks/ru/library/x-stax1/ **3.** Упрощение XML-программирования при помощи JDOM [Электронный ресурс] - Режим доступа: http://www.interface.ru/home.asp?artId=3078.

Королевський Г.Д.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Візуалізація запитів до ресурсів Open Linked Data

Вступ. Об'єм доступної інформації в інтернеті зростає з кожним днем, існує потреба в структуризації та смисловому зв'язуванні розрізнених даних, так з'явилась концепція семантичного веб. Наступний крок – це швидкий та доступний пошук по базі знань. На даний момент семантичний веб включає в себе достатні ресурси для існування великої бази знань (Open Linked Data), яка обмежена двома факторами: семантичний веб містить лише ті зв'язки, які може собі уявити людина, котра їх створює, другий фактор – велика кількість нерозміченої інформації. Проте існує ще й проблема відсутності дружнього до користувача інтерфейсу. Немає можливості візуалізувати запити до бази знань у зручній для користувача формі, результати отримуються у важкому для сприйняття вигляді.

Серед обмалі робочих програмних продуктів по роботі із запитами до семантичних даних у Веб можна виділити Knowledge Graph від Google, який видає точний, але обмежений результат одночасно зі звичайним запитом до пошукової системи.

Задача. Таким чином досі актуальною є проблема адекватної візуалізації запитів до ресурсів LOD та їх результатів. Основні задачі, що мають бути вирішені в роботі над нею, це: дослідження існуючих рішень відображення запитів до баз знань та їх ефективності з точки зору користувача, розробка ПЗ у вигляді плагіну до браузеру Google Chrome, який дозволить проводити аналіз сторінок на наявність семантичної розмітки, показувати зв'язки елементів, а також виконувати швидкий пошук по базі знань будь-якого слова у тексті, результатом якого будуть короткі відомості про шукане поняття, а також семантично зв'язана з ним інформація та фото.

Рис. 1. Семантичний плагін встановлений у Google Chrome

Рис. 2. Інформація про шуканий термін з семантичної бази знань FreeBase

Засоби вирішення задачі. В роботі розглядаються наступні підходи для обробки запитів:

- Латентно-семантичний аналіз текстового запиту. Для пошуку елементів однієї тематики. При цьому підході враховується не тільки наявність ключевих слів з запиту, а й змістова близкість до запиту (виміряється за допомогою ЛСА)
- Алгоритми перетворення звичайного тексту у запит до бази знань. Після ЛСА буде сформовано запит на мові MQL (Metaweb Query Language) до бази знань FreeBase

Висновок. Підсумовуючи, слід відмітити, що використання сучасних алгоритмів аналізу запитів користувача та відкритих ресурсів зв'язаних даних (LOD) дає можливість створення додатку до браузеру, що дозволить проводити пошук потрібної інформації набагато швидше та зручніше, отримуючи максимум від пошуку.

Література. **1.** Toby Segaran, Colin Evans, Jamie Taylor, "Programming the Semantic Web", 2009. **2.** Thomas Landauer, Peter W. Foltz, Darrell Laham. "Introduction to Latent Semantic Analysis», 1998.

Котенко Д.А.
Національний технічний університет України "КПІ", Київ, Україна

Шляхи оптимізації роботи з деревоподібними структурами даних

Актуальність. На сьогоднішній день обсяги інформації, яку необхідно зберігати, швидко зростають [1]. Відповідно, необхідний швидкий та ефективний механізм, який дозволить мати можливість за короткий проміжок часу отримати необхідну інформацію. Як правило, великі масиви інформації впорядковані та структуровані, що додає певних обмежень.

Для обробки та зберігання великих об'ємів інформації створені системи керування базами даних (далі СКБД), які дозволяють досить швидко отримувати необхідну інформацію. Одним із ключових механізмів СКБД для вирішення даної задачі слугують індекси, які будуються на основі даних, що зберігаються. Індекси потребують додаткового простору для зберігання та часу для їх побудови, а також потребують перебудови при зміні інформації, на основі якої вони створені. Індекси будуються та зберігаються у вигляді деревоподібних структур даних, які і забезпечують високу швидкість доступу до даних. Безпосередньо від моделей побудови індексів залежить швидкість доступу до інформації; час, який потребується для побудови та актуалізації масивів індексів. З причин росту об'ємів інформації та загострення питання швидкості доступу до неї дана тема набуває актуальності. Однією з поширених деревоподібних структур даних у комп'ютерній галузі є B-дерева, саме вони розглядаються у даній роботі.

B-дерева. Для побудови індексів в СКБД дуже широко використовуються такі структури даних, як B-дерева. Це збалансоване дерево з дуже високою ступеню розгалуження, яке оптимально підходить для зберігання структурованих даних у зовнішній пам'яті, адже для доступу до інформації необхідна мінімальна кількість звернень. Існує декілька поширених модифікацій B-дерева, а саме:

- B+ дерево. У даній модифікації B-дерева всі записи зберігаються на рівні листових вузлів [2], а у внутрішніх вузлах зберігаються лише ключі; кожен листовий вузол містить покажчик на наступний. Ці особливості дозволяють значно збільшити швидкість доступу до послідовних даних (блоків). B+ дерево має широке коло застосування, а саме:
  - Файлові системи, такі як: NTFS, XFS, ReiserFS, тощо;
  - Реляційні СКБД, такі як: Microsoft SQL Server, Oracle, iSQLite та інші;
  - Нереляційні СКБД (NoSQL), такі як: CouchDB, Tokyo Cabinet.
- B* дерево. Ключовою відмінністю цього дерева є вимога, щоб вузли були заповнені на 2/3, а не на 1/2, як це потребується у B-деревах [3]. Більш висока щільність даних призводить до більш високої ступені розгалуження, і отже, меншої висоти дерева як результат. Таким чином збільшується продуктивність у порівнянні з B-деревом, адже з вторинної пам'яті потребується зчитувати меншу кількість вузлів [4].

Висновки. У роботі запропонована модифікація B-дерева, яка об'єднає переваги вже існуючих B+ та B* дерев. Це дозволить водночас мати швидкий послідовний доступ до даних, заощадити дисковий простір і підвищити швидкість доступу до інформації. Також можливе збільшення коефіцієнту заповнення до значень, що перевищують 2/3. Така зміна може значно збільшити час додавання та видалення даних, але забезпечить більшу компактність даних.

Література. **1.** Wikipedia contributors (2015, April 21). Big Data [Online]. Available: https://en.wikipedia.org/wiki/Big_data **2.** P. E. Black (2007, November 6). B+-tree in Dictionary of Algorithms and Data Structures [Online]. Available: http://xlinux.nist.gov/dads/HTML/bplustree.html **3.** P. E. Black (2007, November 6). B*-tree in Dictionary of Algorithms and Data Structures [Online]. Available: http://xlinux.nist.gov/dads/HTML/bstartree.html **4.** D. E. Knuth, "Multiway trees" in *The Art of Computer Programming*, vol. 3, *Sorting and Searching*, R. S. Varga, Eds. Reading, MA: Addison-Wesley, 1973, pp. 471-476.

Круш І.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Використання графових баз даних для оптимізація швидкості обробки складних запитів

Ще з 80 років XX сторіччя традиційною моделлю збереження даних є реляційна - така модель, яка сприймається користувачем як набір нормалізованих відношень різного ступеню. Інформація у цих базах структуровано зберігається у таблицях з наперед визначеними колонками конкретних типів. Це змушує розробників створювати програми, строго структуруючи дані, які використовуються у їхньому додатку.

В реляційних базах даних зв'язок одного запису на інший запис таблиці ідентифікується за допомогою посилання на первинний ключ першого запису через зовнішній ключ другого. Для того, щоб вибрати певні дані спільно з двох таблиць, потрібно застосувати механізм JOIN. JOIN виконуються під час вибірки пошуком співпадінь між первинним і зовнішнім ключами багатьох записів таблиць, які ми хочемо об'єднати. Ця операція дуже затратна і в обчисленнях, і в памяті і її складність зростає експотенціально.

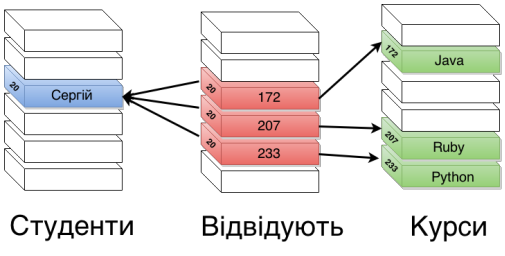

Рис. 1. Зв'язок "багато-до-багатьох" у реляційній моделі

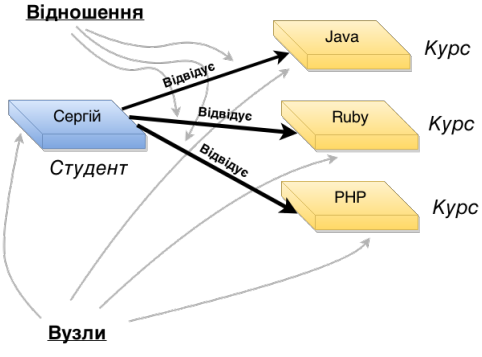

Рис. 2. Зв'язок "багато-до-багатьох" у графовій моделі

Якщо у базі є відношення *багато до багатьох*, тоді для ідентифікації зв'язку між двома таблицями створюється спеціальна JOIN-таблиця, що містить у собі зовнішні ключі таблиць, які потрібно об'єднати. Це ускладнює виконання нашого запиту і відповідно збільшує час очікування відповіді. Через складніть операції JOIN розробники часто змушені денормалізовувати таблиці з метою зменшення кількості об'єднань, що руйнує чітку і логічну структуру сутностей.

Для побудови систем з великою кількістю зв'язків, запити до яких вимагатимуть об'єднання більше ніж двох таблиць, використовувати реляційну базу даних нераціонально. В минулому відсутність життєздатних альтернатив і величезне ком'юніті підтримки цього виду баз майже унеможливило широке росповсюдження інших реалізацій БД. Саме для вирішення таких типів задач використовуються графові бази даних.

У графових БД основними структурами є вузли і відношення, в той час як у інших базах зв'язки треба описувати через спеціальні властивості, як, наприклад, *механізм первинних/зовнішніх ключів* або *map-reduce* обробку. Механізм, який пропонують графові бази зручний тим, що, абстрагуючи все в одну цілісну структуру, ми можемо будувати насправді грамотно продумані моделі, який найточніше характеризують реальний прототип.

Для прикладу розглянемо сайт пошуку роботи. Роботодавець розміщує вакансію, у якої є вимоги до знань потенціального працівника та місто, куди він потрібен. У традиційній реляційній моделі схема такої моделі може мати вигляд, показаний на рис. 3.

Для того, щоб користувачу знайти собі вакансію у певному місті з певними навиками, нам потрібно буде написати з кількома підзапитами та операціями JOIN. Очевидно, швидкість

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

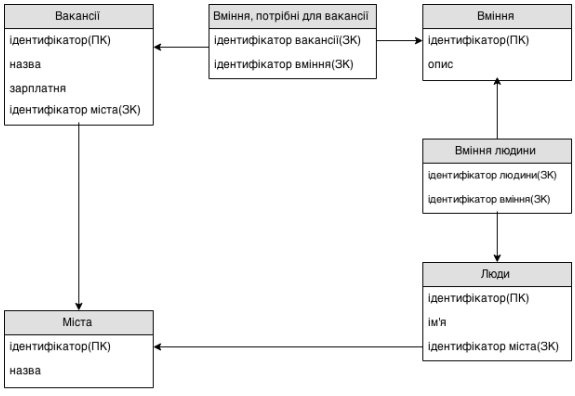

Рис. 3. Реляційна схема бази даних сайту пошуку роботи

роботи такого запиту буде вкрай малой. Спробуємо побудувати схему у графовій моделі.

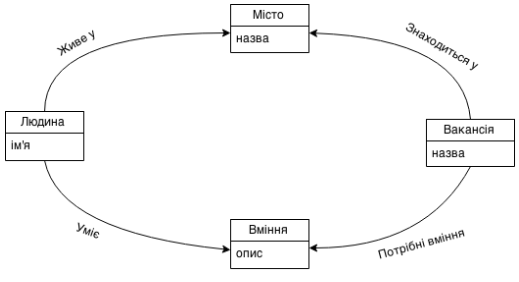

Рис. 4. Графова схема бази даних сайту пошуку роботи

У такому варіанті для того, щоб виконати запит користувача на пошук роботи, нам досить через відношення **Живе у** знайти вакансію, яка має відношення **Знаходиться у** з тим самим **Містом**, а також через відношення **Уміє** знайти вакансію, яка має відношення **Потрібні вміння** через **Вміння**.

У графових база даних кожний вузол(сутність або атрибут) напряму фізично містить у собі список записів-відношень, що представляють їх зв'язки з іншими вузлами. Відношення структуровані за типом і напрямком та можуть містити додаткові атрибути. Коли в реляційній базі використовується операція JOIN, то в графовій базі просто напряму просканується список відношень початкового вузла, що, очевидно, є набагато швидшою операцією, ніж об'єднання таблиць, адже це не вимагає розширеного пошуку.

Саме можливість матеріалізувати відношення в структури самої бази дозволяє графовим БД забезпечити продуктивність на кілька порядків вищу, ніж у реляційних, особливо для важких JOIN-запитів. У таких випадках час відповіді можна зменшити від кількох хвилин до кількох мілісекунд.

Виходячи зі сказаного вище, можемо зробити висновок, що у випадках систем з великою кількістю зв'язків раціональніше використовувати графову модель бази даних, адже вона є набагато простішою, інтуїтивно зрозумілішою і продуктивнішою, ніж реляційна та дозволяє максимально пришвидшити вибірку.

Література. **1.** Your Database Is So Retro: Old Data, New Databases, http://www.texasenterprise.utexas.edu/2014/08/14/innovation/your-database-so-retro-old-data-new-databases. **2.** GraphDB vs RDBMS, http://neo4j.com/developer/graph-db-vs-rdbms. **3.** Graph Databases, http://graphdatabases.com/. **4.** Graph Database Use Cases, http://www.slideshare.net/maxdemarzi/graph-database-use-cases.

Куліш З.І.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Забезпечення інформаційної безпеки в системі дистанційного навчання LON-CAPA

На сьогоднішній день однією з найбільш розповсюджених систем дистанційного навчання в Україні є MOODLE, яка надає розробнику навчальних матеріалів можливість використовувати всі необхідні ресурси та засоби контролю, однак вона має свої недоліки: розширюваність(можливість розширення кола слухачів системи дистанційного навчання), масштабованість (додання програм і курсів навчання), відсутність інструменту для керування групами студентів, проблеми з адмініструванням користувачів та з захищеністю і обробкою даних.[1]

Проте після детального дослідження різних систем дистанційного навчання (гнучкість інтерфейсу, персоналізація, розширюваність/відкритість, можливість налаштування вмісту) [2] для дослідження була обрана система LON-CAPA. Це система дистанційного навчання з відкритим вихідним кодом (Open-Source), яка реалізована для користувачів у вигляді сервісу, що надає зручний інтерфейс для віддаленого доступу до засобів (програм) та матеріалів навчання.

У системі дистанційного навчання LON-CAPA ефективність та якість навчального процесу визначаються використанням детально продуманих програмних засобів для керування вмістом навчальних програм, які реалізуються у вигляді LCMS (Learning Content Management System).[3]

В основі LCMS лежить концепція представлення навчання як сукупності багаторазово використовуваних навчальних об'єктів зі своєю цільовою аудиторією і визначеним контекстом використання. Ця концепція значно покращує навчальний процес, однак захищеність навчальних матеріалів (об'єктів) суттєво знижується.

Для забезпечення інформаційної безпеки була використана особливість даної системи дистанційного навчання, а саме можливість підключення та підтримки служби каталогів Active Directory. З її допомогою було налаштувано адміністрування користувачів на основі групових політик.[4] Вони визначають, які зміни можуть вносити користувачі(додавання нового матеріалу, редагування), хто має доступ до навчальних матеріалів та інше. До того ж завдяки ієрархічній структурі Active Directory користувачі, котрі знаходяться в певних групах служби каталогів наслідують дозволи і налаштування безпеки, що значно полегшує роботу і покращує захищеність та конфіденційність навчальних матеріалів.

Висновок. З використанням служби каталогів Active Directory у системі LON-CAPA був реалізований високий рівень захисту інформації, яка зберігається та обробляється.

Література. **1.** А. І. Гладир, к.т.н., доц., Н. В. Зачепа, фахівець,О. О. Мотруніч, асп. Кременчуцький національний університет імені Михайла Остроградського – Системи дистанційного навчання – огляд програмних платформ – Режим доступу: http://www.uk.xlibx.com/4mehanika/23929-11-kremenchuckiy-nacionalniy-universitet-imeni-mihayla-ostrogradskogo-mizhnarodna-naukovo-tehnichna-konferenciya-mo.php **2.** 2007 Distance Education Survey. A Report on Course Structure and Education Council Member Institutions. 1601 18 th Street, NW, Suite 2, Washington, DC 20009 – 2529 – Режим доступу: http://www.detc. Org **3.** Готская И.Б., Жучков В.М. Кораблев А.В. , РГПУ им.А.И Герцена – Аналитическая записка « Выбор системы дистанционного обучения» – Режим доступу:http://rakurs.spb.ru/2/0/2/1/?id=13 **4.** С. Реймер, М. Малкер Active Directory для Windows Server 2003. Справочник администратора / Пер, с англ. — М.: «СПЭКОМ», 2004.— 512 с

Мазурик А.Ю.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Создание интерактивной инфографики средствами библиотеки Highcharts.js

Инфографика - это графический способ подачи информации, данных и знаний, целью которого является быстро и четко преподносить сложную информацию. Именно интерактивная подача информации (динамические графики, таблицы и т.п.) позволяет отобразить удобно и четко, а главное - наглядно, колоссальные объемы информации.

В инфографике можно выделить три основных принципа, следуя которым можно эффективно отображать данные. Первый из них - информационное кодирование. Суть информационного кодирования заключается в том, чтобы показать какой либо атрибут информации с помощью формы, цвета или размера. В таком случае любые графики станут нагляднее и смогут нести больше информации. Смысл кодирования исходит именно из того, как наш мозг воспринимает входную информацию. Используя данный принцип можно за гораздо меньшее время передавать большее количество информации человеку.

Следующий принцип инфографики - макро- и микроуровни. На мой взгляд, этот принцип является самым сильным. Микроуровни - это так называемые "точки основания", с помощью которых строится общая картина или макроуровень. Так, например, на карте всех зафиксированных землетрясений на Земле микроуровнем будут точки катастроф. И посмотрев на них можно заметить некоторую тенденцию того, что в основном все землетрясения возникают в области границ тектонических плит.

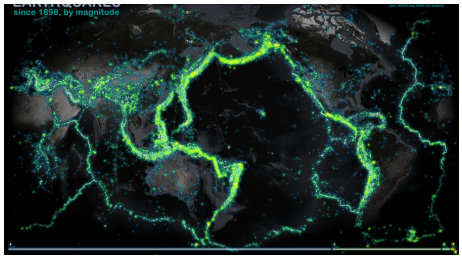

Рис. 1. Инфографика землетрясений на Земле (1988-2012)

Последним принципом инфографики является интерактивность. Время статической инфографики плавно подходит к концу и каждому, кто смотрит ваши графики хочется с ними как-то повзаимодействовать. И вот эта самая динамика позволит отобразить еще нагляднее и еще больше данных в ваших работах.

Для большего распространения данных их размещают в Интернете. Для этого есть бесчисленное множество ресурсов и плагинов, которые позволяют достаточно быстро и легко связать данные и необходимые графики. Большинство плагинов как движок используют D3.js. Его синтаксис схож с известными JS-библиотеками как jQuery и Prototype, поэтому порог входа можно назвать достаточно низким. Но для этого необходимо потратить много времени, потому что D3 - это движок, в котором нет заготовленных графиков, гистограмм и других визуализаций.

Одной из самых популярных и достаточно полных библиотек визуализации является Highcharts.js, которая написана на D3. В ней реализованы все необходимые элементы инфографики, а также на каждый из графиков можно накладывать собственные стили. Для примера рассмотрим график изменения курса доллара по отношению к евро с 2007 по 2009 год.

Тут, как можно заметить реализованы все основные принципы построения инфографики. В качестве "точки основы" был выбран курс доллара в определенный день, а на макроуровне можно увидеть динамику падения его курса. Интерактивность реализована за счет временной шкалы на всей площади графика, на которой можно выбрать необходимый промежуток

Рис. 2. График изменения курса доллара в 2006 - 2008 гг.

времени для детального изучения.

Удобство реализации данного графика заключается в том, что данные и представление разделены полностью, а сама визуализация создается автоматически, используя Highcrats.js. В функцию обработчик необходимо лишь передать нужные данные и способ обработки осей:

```
detailChart = $('#detail-container').highcharts({
        chart: {
            marginBottom: 120,
            reflow: false,
            marginLeft: 50,
            marginRight: 20,
            style: {
                position: 'absolute'
            }
        },
        credits: {
            enabled: false
        },
        title: {
            text: 'Historical USD to EUR Exchange Rate'
        }, ...);
```

Рис. 3. Пример кода на Highcharts.js

А потом нужно добавить элемент detailChart к необходимому DOM элементу. Интерактивность будет реализовано за счет функции, которая будет перерисовывать основную область визуализации.

Как можно отметить, создание непростой визуализации графическими библиотеками JS оказывается очень простой задачей если подобрать правильный плагин для нужной визуализации.

В заключение хочеться сказать, что придерживаясь основных принципов построения инфографики и используя Highchart.js, можно создавать наглядные и понятные визуализации. Самое основное - определится с типом представления данных, который сможет лучше всего донести до людей вашу информацию. А Highcharts.js предоставит весь необходимый инструментарий для создания любых инфографиков.

Литература. **1.** Курс "Визуализация данных. Основы", http://infogra.ru/infographics/onlajn-kurs.vizualizatsiya-osnovy. **2.** Framework D3.js, http://d3js.org/. **3.** Library Highcharts.js, http://www.highcharts.com/.

Малишев А.І.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Документування RESTful API

Вступ. Важливим етапом розробки будь-якої клієнт-серверної інформаційної системи є правильно оформлене за усіма вимогами RESTful API. Однією з основних проблем створення такого API є його документування. Оскільки на сьогоднішній день не існує чітких правил по створенню документації RESTful API, то у даних тезах буде висвітлено оптимальні рекомендації по цій проблемі.

Основні положення. Найперше, що потрібно зробити - це проаналізувати основні, високорівневі структури даних (моделі), якими оперує програма. Далі проводиться аналіз дій, що здійснюються над цими компонентами. Як приклад, візьмемо документацію по foursquare API [1]. В цьому прикладі є набір високорівневих об'єктів, таких як місця, користувачі тощо. Об'єктом може бути, наприклад, JSON-опис [2] об'єкта з параметрами. Також там є набір дій, які можна зробити над цими об'єктами. Завдяки такій декомпозиції, створення структури викликів API стає простішим і зрозумілішим. Наприклад, для додавання нового місця логічно буде викликати метод типу "/venues/add". Правильним є те, що:

1. Потрібно документувати всі високорівневі об'єкти.
2. Для кожного об'єкту документувати запити і відповіді на них. Замість простих типів даних використовуються високорівневі об'єкти.
3. Замість того, щоб писати "Цей виклик повертає три строкових поля, перше містить id, друге ім'я, а третє опис" пишіть "Цей виклик повертає структуру (модель), що описує місце".

Документування параметрів запиту. Для прикладу візьмемо API, що дозволяє користувачу входити, використовуючи Facebok token [3]. Викличемо цей метод як /login.

```
Request
/login
Headers
Authorization: Token XXXXX
User-Agent: MyGreatApp/1.0
Accept: application/json
Accept-Encoding: compress, gzip
Parameters
Encoding type - application/x-www-form-urlencoded
token - ‘‘Facebook Auth Token’’ (mandatory)
profileInfo = ‘‘json string containing public profile information from Facebook’’
```

Де profileinfo високорівневий об'єкт. Оскільки вже задокументовано внутрішню структуру цього об'єкта то такої простої згадки достатньо.

Такі параметри як Accept, Accept-Encoding, Encoding type потрібно документувати окремо, замість того, щоб повторювати їх у всіх розділах.

Документування параметрів відповіді. Відповіді на виклики API повинні бути також задокументовані, ґрунтуючись на високорівневій моделі об'єктів. Посилаючись на приклад foursquare, виклик методу "/venue/venueid" поверне структуру даних (модель), що описує місце проведення заходу.

Висновки. В даних тезах було піднято актуальну проблему документування RESTful API для сучасних клієнт-серверних інформаційних систем і наведено рекомендації по правильному документуванню API.

Література. **1.** Foursquare API, https://developer.foursquare.com **2.** JSON, http://www.json.org/ **3.** Facebook token, https://developers.facebook.com/docs/facebook-login/access-token

Малишев А.І.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Розробка способів управління доступом в корпоративному Java-додатку

Управління доступом є однією з основних частин безпеки веб-додатків. Контроль доступу гарантує, що тільки аутентифіковані і авторизовані особи можуть мати доступ до конфіденційної інформації, і тільки користувач з допустимою роллю може виконувати надані йому дії. Формування ролей покликане визначити чіткі і зрозумілі для користувачів інформаційної системи правила розмежування доступу. Рольовий поділ дозволяє реалізувати гнучкі, що змінюються динамічно в процесі функціонування додатка, правила розмежування доступу [1]. Розглянемо кілька способів реалізації системи управління доступом в корпоративному Java-додатку.

Перший спосіб реалізації передбачає обмеження доступу до частин проекту. Крім цього накладається обмеження на число невдалих спроб входу в систему, після чого аккаунт блокується на певний час, що ускладнить застосування методу «грубої сили» (brute force [2]). Реалізується клас менеджера користувачів, який повинен забезпечувати додавання, видалення, зміну і читання облікового запису користувача, перевірку на існування логіна користувача та виконання входу користувача в систему. У класі інформації про користувача необхідно зробити незмінними ідентифікатор користувача і логін. Цей клас також повинен містити реєстраційну інформацію про користувача (обов'язково логін / пароль і роль користувача) і час його останнього входу. Також необхідно зберігати інформацію про кількість невдалих спроб аутентифікації і про час / дату зняття блокування з облікового запису користувача, якщо він заблокований. Необхідно реалізувати фільтр, що обмежує доступ до URL-адреси сайту відповідно до заданих обмежень. Обмеження повинні задаватися у файлі XML [3], шлях до якого буде вказуватися в конфігураційних параметрах програми. При накладенні обмежень безпеки необхідно вибрати, які ресурси захищати, задавши обмеження з шаблоном URL і ролями, яким заборонено відвідувати вказану частину сайту. Фільтр повинен перехоплювати всі запити від клієнта, які приходять. Якщо URL запиту HTTP і роль користувача відповідають шаблону і зазначеній ролі в обмеженні, доступ до ресурсу обмежується.

Альтернативний спосіб полягає в застосуванні захисту на рівні методів контролера, використовуючи підтримку анотацій [4]. Буде проводитися перевірка прав користувача для виконання методів, над якими вказана анотація. Функція, над якою стоїть анотація, виконуватиметься тільки якщо у користувача (даного потоку виконання) є права, зазначені в параметрі анотації. Можливий варіант, коли анотується весь клас. Тоді всі методи класу потребують відповідних прав, і ми можемо перевизначити ті методи, на виконання яких потрібні особливі права. Об'єкт, який конфігурується за допомогою анотації, створюється за допомогою фабрики-контейнера, тому що, створюючи об'єкт за допомогою оператора new, довелося б провести додаткові дії для перевірки прав користувача.

Запропоновані способи управління доступом повинні вибиратися залежно від вимог предметної області і не завжди є взаємозамінними. Безпеки, яка базується на URL-адресах, виявляється недостатньо, коли кілька ролей має доступ до одного і того ж ресурсу (URL-адреса), але при цьому ролі повинні виконувати різні дії. Для подібного сценарію необхідно реалізувати більш тонку політику безпеки, застосовуючи різні права доступу до різних частин ресурсу.

У роботі були розглянуті кілька способів проектування політики управління доступом, які були засновані на URL-адресах і виклику методів контролера, а також випадки, в яких найкраще застосовувати дані способи.

Література. **1.** Role-based access control, http://en.wikipedia.org/wiki/Role-based_access_control **2.** Brute force attack, http://techopedia.com/definition/18091/brute-force-attack **3.** XML, http://w3.org/XML **4.** Annotations, http://tutorials.jenkov.com/java/annotations.html

Маматов В.И.
Национальный технический университет Украины "КПИ", факультет информатики и вычислительной техники, Киев, Украина

Электронная коммерция

Под определинием электронной коммерции(e-commerce) подразумеваются любые формы деловых сделок меджу сторонами, когда взаимодействие этих сторон осуществяеться електронным способом, нежели физическим контактом, в результате которого право владения или право использования рассмотренной собственности или услуги передается между сторонами. В докладе анализируются различные категории електронной коммерции, а также рнжирование секторов рынка електронной комерции по денежному обьему с количству продаж [2].

Электронную коммерцию принято делить на следующие [1] категории:

- **business-to-consumer (B2C).** Основу этой категории являет электронная розничная торговля;

- **business-to-administration (B2A).** Взаимодействие администрации и бизнеса включает деловые связи коммерческих структур с государственными организациями;

- **business-to-business (B2B).** Направление включает в себя все уровни взаимодействия между компаниями. Могут использоваться специальные стандарты и технологии электронного обмена данными;

- **consumer-to-administration (C2A).** Это направление наименее развито, однако имеет достаточно высокий потенциал, который может быть использован для организации взаимодействия государственных структур и потребителей, особенно в социальной и налоговой сфере;

- **consumer-to-consumer (C2C).** Подразумевается возможность взаимодействия потребителей для обмена коммерческой информацией. Это может быть обмен опытом приобретения того или иного товара, обмен опытом взаимодействия с той или иной фирмой и многое другое. Среди этапов использования електронной коммерции [3] можно выжелить такие : маркетинг, производство, продажи и платежи, а также степень использования информационных и коммуникационных технологий и систем, которая служит мерой, по которой бизнес может считаться электронным. Вся необходимая информация находиться на web-серверах которые принадлежат организациям, осуществляющим интернет-услуги. Доступ к информации осуществляется по запросам из программ-браузеров клиентов сети. Продвижение e-commerce в Интернет обеспечивает доступ производителей к максимальному числу потребителей и их многообразным предпочтениям и предоставляет возможность клиентам вводить свои заказы в систему управления предприятием.

Таблица 1. Распределение мест секторов рынка электронной коммерции

Место	Ранжирование по денежному объему	Ранжирование по количеству продаж
1	недвижимость	программное обеспечение
2	компьютеры и комплектующие	звуковоспроизводящая аппаратура
3	программное обеспечение	разное
4	туристическое обслуживание	компьютеры и комплектующие
5	звуковоспроизводящая аппаратура	туристическое обслуживание
6	финансовые услуги	финансовые услуги

Литература. **1.** Беляевский, И. К Маркетинговое исследование: информация, анализ, прогноз [Текст] / И. К Беляевский. – М.: Финансы и статистика, 2001. **2.** Курс ЦИТ "Internet-технологии в проектах с пластиковыми карточками". В. Завалеев, Центр Информационных Технологий. **3.** Алексунин, В. А., Родигина В. В. Электронная коммерция и маркетинг в Интернете [Текст]: Учебное пособие / В. А. Алексунин, В. В. Родигина. – М.: Дашков и К, 2005.

Медведський А.М.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Індексація сторінок побудованих за допомогою JavaScript фреймворків

Сьогодні існує велика кількість JavaScript фреймворків, що часто використовуються і для побудови HTML. Серед найвідоміших - AngularJS, EmberJS, Backbone та інші. Але при їх використанні виникає проблема індексації сторінок. ЇЇ суть полягає в тому, що одні пошукові роботи взагалі не знають JavaScript, інші - в дуже обмеженому об'ємі. Що означає, коли сторінка генерується за допомогою скріпта, робот отримує лише вашу сторінку з підключеним кодом, який не виконується. [4]

Для вирішення цієї проблеми існує велика кількість підходів, але у всіх суть полягає у різній видачі даних. Наприклад, якщо сторінку відвідує звичайний користувач, він отримує повноцінний додаток, у випадку пошукових роботів, вони отримують згенерований статичний HTML. Все це відбувається на сервері [1].

```
var express = require('express');
var app = express();

var UAParser = require('ua-parser');

var routes = require('./routes/index');
var staticRoute = require('./routes/test');

var UACheck = function (req, res) {
    var ua = req.headers['user-agent'];
    ua = UAParser.parseUA(ua)['family'];
    if (ua == 'FacebookBot' || ua == 'TwitterBot' || ua == 'Googlebot' || ua == 'Other'
    ) {
        staticRoute(req, res);
    } else {
        routes(req, res);
    }
}

app.use('/', UACheck);
```

Рис. 1. Приклад коду на NodeJS

Одним із популярних способів є використання Prerender. Для цього потрбіно:

1. Middleware на Ruby on Rail або NodeJS перевіряє користувацького агента при кожному запиті.
2. Якщо було виявлено плазуна пошукової машини, то відправляється GET запит до Prerender.
3. Prerender викликає ваш додаток для тієї сторінки, яку запитує плазун.
4. Сторінка відмальовується за допомогою PhantomJS.
5. Отриманий HTML відправляється назад.
6. Middleware повертає HTML краулеру.

[5]

Цей підхід вимагає додаткових ресурсів та часу розробника, щоб правильно налаштувати роботу. Але через відсутність хороших альтернатив та наявність готового інструментарію для реалізації, підхід набув широкого розповсюдження.

Рішенням проблеми індексації є методологія, що має лише один канал подачі і для пошукових

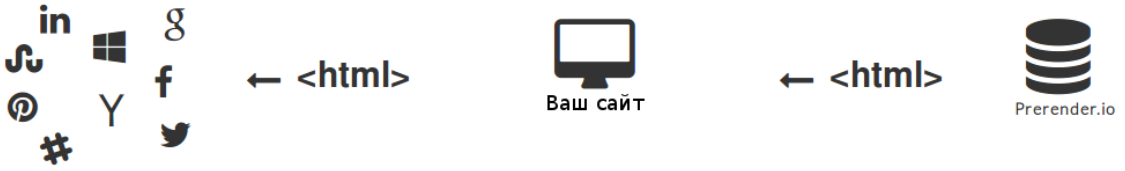

Рис. 2. Приклад роботи методу

систем, і для реальних користувачів. Такий варіант означав би, що сторінки відображаються одразу після їх надання. Непотрібні були б очікування подій готовності сторінки та завантаження всього JS. Додатки могли б генеруватись на сервері за допомогою NodeJS.

ReactJS - бібліотека, що зазвичай використовується в поєднанні з методологією Flux. Вона використовує для своєї роботи технологію Virtual DOM та Diff алгоритм. Він полягає в порівнянні змін у віртуальному дереві з існуючим, що призводить до зменшення кількості запитів до DOM. Це робить бібліотеку не тільки дуже швидкою про роботі з DOM елементами, а також SEO friendly [2].

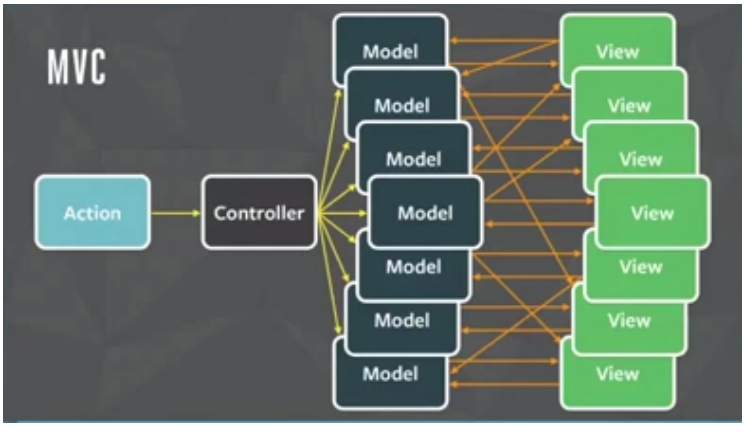

Рис. 3. Діаграма Flux+React MVC

Саме в ReactJS реалізований підхід одноканального надання. Цю бібліотеку можна можна використовувати для подачі статичної розмітки, без клієнтського коду взагалі. Використання ReactJS - найкраще вирішення проблеми індексації сторінок. Завдяки цьому та багатьом іншим перевагам бібліотека швидко набирає популярності. [3].

Література. **1.** BEKK Open www.evilbloodydemon.ru/blog/2014/11/diy-angularjs-seo-with-phantomjs. **2.** SEO Для сайтов www.open.bekk.no/easier-reasoning-with-unidirectional-dataflow-and-immutable-data. **3.** Хабрхабр www.habrahabr.ru/company/hexlet/blog/249579/. **4.** Node.js + Express4 + Angular.js індексація www.toster.ru/q/171773. **5.** Хабрахабр http://habrahabr.ru/post/196718/.

Мельничук М.М.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Модель віддаленого керування мікроконтролерними системами

Вступ. Одним із основних напрямків розвитку мікроконтролерних систем являється дослідження та розробка роботів, які виконують власні задачі за допомогою складених алгоритмів та їх реалізацій. Зазвичай ці системи не є ідеальними, тому вони вимагають втручання оператора керування. Саме тому тема віддаленого управління є досі актуальною.

Недоліки зв'язку. Одними з найважливіших характеристик систем віддаленого керування є відстань дії та практичність керування системою. Розглянемо декілька видів віддаленого зв'язку, які існують на даний момент.

1. Інфрачервоний канал. Його явними проблемами являються обмеженість дальності дії (всього до кілька десятків метрів) та неможливість працювати через фізичні перешкоди.
2. Провідний канал. Досить непрактичний, але найбільш надійний.
3. Радіоканал. Більша дальність дії, можливість керування через перешкоди.

Краща модель зв'язку. Найкращим з представлених видів є радіоканал, але що робити, коли потрібні набагато більші відстані у сотні чи тисячі кілометрів.

Тут ми згадаємо, певно, найдивовижніший винахід людства, мережу інтернет. Це є найпростіший та найкращий метод віддаленого керування мікроконтролерними системами, який, у поєднанні з радіоканальним зв'язком, дає доступ до пристрою у будь-якій точці планети, необділеної хоча б мобільним покриттям.

Приклад моделі. Прикладом такої системи керування є робот, який побудований на наступній схемі керування.

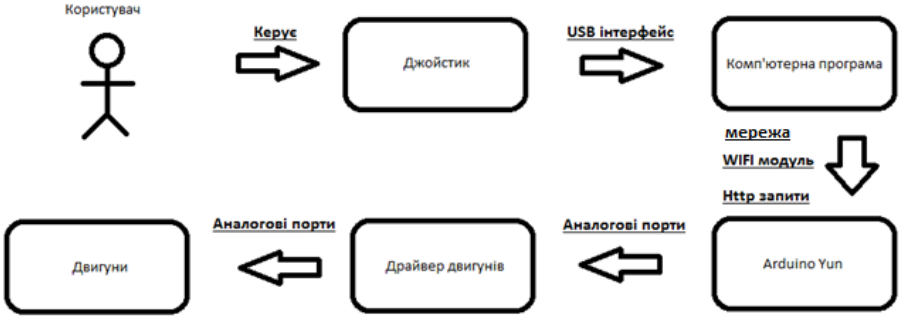

Рис. 1. Схема керування

Поясненя моделі керування. *Оператор* керує *джойстиком*, відповідно до потрібного напряму руху та потужності (швидкості). Дані з джойстика зчитуються за допомогою *комп'ютерної програми* та певним чином редагуються. Далі в цій же програмі формується спеціальний *Http запит* та відсилається через мережу *Інтернет* до точки доступу, що через *Wi-fi* зв'язок передає інформацію на*Arduino Yun*. В середині Arduino запит обробляється певним чином; згідно цих даних посилається сигнал на *драйвер двух двигунів*. Далі драйвер, інтерпретуючи сигнали, подає струм та напругу на *сервоприводи*, які передають обертальну енергію колесам, що рухають платформу, на якій знаходяться Arduino Yun, драйвер двигунів, живлення та самі двигуни з колесами.

Література. **1.** Інтернет-ресурс. - Режим доступу: http://www.arduino.cc/en/Main/ArduinoBoardYun. **2.** Інтернет-ресурс. - Режим доступу: http://uk.wikipedia.org/wiki/Зв'язок. **3.** Інтернет-ресурс. - Режим доступу: http://uk.wikipedia.org/wiki/Робот

Мирошникова І.Ю.
Фізико-технічний інститут НТУУ "КПІ", Київ, Україна

Вибір провайдера хмарного сервісу за умови динамічного середовища

Розвиток мережевих технологій сприяє поширенню парадигми хмарних обчислень, яка дозволяє інформаційним системам використовувати для виконання обчислювальних задач ресурси сторонніх провайдерів. В таких системах виникає задача вибору провайдера серед доступних, характеристики надання сервісів якого задовольнять вимоги користувача.

Із зростанням кількості провайдерів хмарних сервісів набувають поширення інформаційні системи, в яких вибір провайдера здійснюється у окремій підсистемі [1] автоматизовано, за стратегією на основі машинного навчання. Зокрема, використовується навчання з підкріпленням [2], оскільки система тоді не потребує виконання провайдерами тестових завдань. У роботі [3] показано, що додаткове врахування репутації провайдерів при використанні стратегій на основі навчання з підкріпленням призводить до покращення якості сервісів, наданих обраними провайдерами. Систему було змодельовано для середовища зі сталими характеристиками провайдерів, тому доцільно дослідити поведінку системи за умов середовища, в якому характеристики провайдерів можуть змінюватись. **Метою роботи** є дослідження поведінки системи вибору провайдера хмарного сервісу на основі навчання з підкріпленням та урахуванням репутації провайдерів за умов повної та тимчасової відмови одного з провайдерів.

Налаштування системи. У даній роботі змодельовано систему з одним брокером, з налаштуваннями, аналогічними до наведених у роботі [3]. Брокер обслуговує 100 користувачів та 10 провайдерів. Поріг репутації для множини авторитетних провайдерів: 0,8. Провайдер обирається з множини авторитетних на основі максимізації очікуваної цінності.

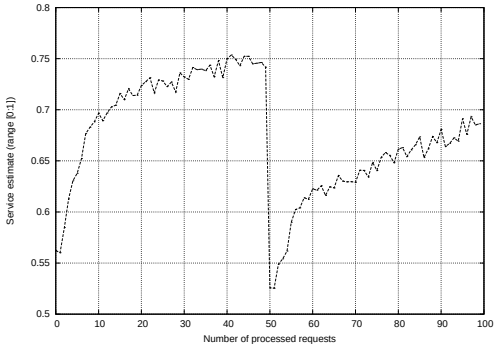

Рис. 1. Відмова обраного провайдера після 50 оброблених запитів

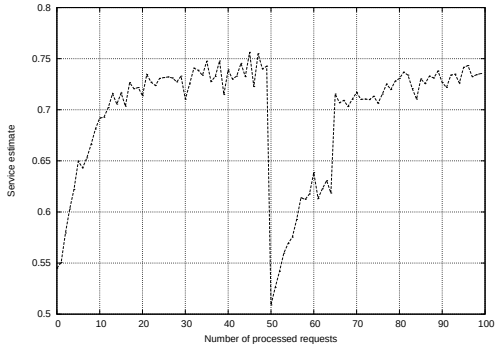

Рис. 2. Відмова обраного провайдера з 51й по 65й оброблені запити

На рис. 1 та 2 відображені результати моделювання системи, усереднені за 1000 прогонів по 100 запитів у кожному, у разі відмови найкращого за репутацією (на момент відмови) провайдера - повної та тимчасової відповідно. Час адаптації системи до відмови провайдера, приблизно рівний часу її адаптації до невідомого середовища.

Результати моделювання показують працездатність використаного в системі підходу на основі навчання з підкріпленням та репутації провайдерів в динамічному середовищі.

Література. **1.** Wang W., Niu D., Li B. & Liang B. (2013). Dynamic Cloud Resource Reservation via Cloud Brokerage. *Proceedings of the 2013 IEEE 33rd International Conference on Distributed Computing Systems*, 400-409. **2.** Wu J., Xu X., Zhang P. & Liu C. (2011) A novel multi-agent reinforcement learning approach for job scheduling in Grid computing. *Future Generation Computer Systems*, 27(5), 430-439. **3.** Мирошникова, І. Ю., Новіков, О.М. (2014). Вибір провайдера хмарного сервісу на основі навчання з підкріпленням та репутації провайдерів. *Реєстрація, зберігання і обробка даних*, 16(4), 54-63.

Михалько В.Г.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Використання кешування і попередніх розрахунків для пришвидшення часу відгуку веб-сервісів

При зростанні популярності сучасних веб-сервісів, їх розробники неодмінно стикаються з такими речами як:

- Великі і стрімко зростаючі об'єми даних
- Складні запити до бази даних
- Висока частота оновлень

При цьому час на обробку запиту від користувача має бути мінімальним і тривати не довше кількох мілісекунд. Це постає однією з найбільших проблем. При зростанні об'єму даних і кількості користувачів швидкість обробки запитів до бази даних стає низькою. Запити можуть тривати декілька секунд або більше, а це часто є неприйнятним для користувачів.

Для прикладу розглянемо веб-сервіс, метою якого є можливість створення постів, а також перегляд їх за різними критеріями, такими як рейтинг, тег чи категорія. Частина схеми бази даних може мати такий вигляд, як зображено на рис.1.

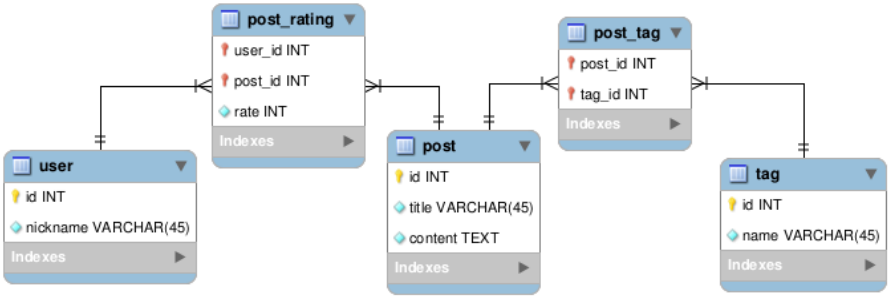

Рис. 1. Схема бази даних

Якщо користувач хоче завантажити сторінку постів з тегом "cats", відсортованими за рейтингом, то запит до БД буде виглядати таким чином:

```
SELECT post.id, post.title, post.content, SUM(post_rating.rate) AS total_rate
FROM post JOIN post_tag ON post_tag.post_id=post.id
JOIN tag ON post_tag.tag_id=tag.id
JOIN post_rating ON post_rating.post_id=post.id
WHERE tag.name="cats" GROUP BY post.id, post.title, post.content
ORDER BY total_rate DESC LIMIT 0, 10;
```

При невеликій кількості постів і користувачів цей запит не викликатиме проблеми. Але якщо в базі даних 10 мільйонів постів і при цьому 10 тисяч користувачів одночасно намагаються завантажити цю сторінку, то відповіді будуть надходити дуже довго. Частково це можна вирішити, використовуючи індекси. Індекси значно пришвидшують пошук по одній таблиці, але вони все одно не рятують від великої кількості JOIN-операцій, які є вкрай складними і потребують часу на обробку. Іншим варіантом є денормалізація, але при цьому збільшується ймовірність, що дані втратять свою цілісність.

Можливим рішенням цієї проблеми є використання кешування, тобто запис результатів частих запитів в оперативну пам'ять. Це дозволить значно швидше отримати дані. Проте деякі проблеми все одно залишаються. Кеш, в більшості випадків, працює за таким принципом: після запиту на отримання деяких даних спочатку йде перевірка на їх наявність в кеш-сховищі; якщо вони там знайдені - то це називається попаданням (cache hit) і користувач швидко їх отримує, в іншому випадку система має шукати дані в основному сховищі, потім перемістити

їх в кеш і надіслати користувачу - це називається промахом(cache miss), і час його обробки значно вищий. Але в такому підході все ж є недолік - завжди знайдеться користувач, який перший буде робити запит на дані. В цьому випадку відбудеться промах, і запит може тривати дуже довго.

Для вирішення цієї проблеми можна поєднувати попередні обчислення (precomputations) і кешування, тобто кешувати дані не під час першого запиту від користувача, а завчасно. Для цього треба прорахувати всі, або, принаймні, більшість можливих запитів від користувачів, виконати їх передчасно і помістити результати в систему кешування "ключ-значення" типу Redis [1] або Memcached [2].

При використанні даного підходу треба враховувати, що система вимагає періодичного оновлення кеша, щоб він не був застарілим (cache-invalidation). Стратегія оновлення кеша часто є досить специфічною і залежною від конкретного веб-сервісу.

Застосуємо кешування і попередні обчислення для прикладу веб-додатку наведеного вище і порівняємо результати зі звичайними запитами до БД. Для наповнення сервісу тестовими даними використаємо скрипт `datagen.py` [3]. Одним із основних запитів до цього веб-сервісу, буде той, що наведено вище і подібні, через це є сенс їх попередньо виконувати і кешувати. Це робить фонова програма `redis_cache_update.py` [4]. Основна її ідея полягає в тому, щоб для кожного тегу витягнути відсортовані за рейтингом дані, розділити їх на частини і помістити в систему кешування Redis. Ключ буде мати такий вигляд *"тег:зміщення:кількість"*, а значенням буде інформація про відповідні пости, закодована у форматі JSON. Наприклад, якщо користувач хоче отримати пости з тегом "cats", які займають з 20-ї по 30-ту позиції у рейтингу, то потрібний ключ у Redis буде виглядати таким чином: *"cats:20:10"*.

Для тестування цього підходу і порівняння зі звичайними запитами до БД було використано тестові програми `mysql_fetch.py` [5] і `redis_fetch.py` [6]. Експеримент був повторений декілька разів над різними об'ємами даних. Результати наведено у таблиці 1.

Табл. 1. Результати експериментів

Кількість даних		Середня тривалість звичайних запитів, с	Середня тривалість попередньо розрахованих і кешованих запитів, с
Користувачі	Пости		
100	2500	0.057	0.000163
200	5000	0.196	0.000159
500	10000	1.084	0.000161
1000	25000	6.069	0.000163
2000	50000	39.185	0.000197

З табл. 1 видно, що при кількості користувачів більше за 500 і кількості постів більше за 10000 звичайні запити до БД є вкрай повільними і неприйнятними для клієнтів веб-сервісу. Тривалість завчасно розрахованих запитів, навпаки, майже не підвищується, незважаючи на збільшення кількості даних.

При використанні кешування існує ймовірність, що користувач отримає дещо застарілі дані. В програмі [4], наприклад, оновлення відбувається кожні 5 хвилин, а не одразу. Але, зважаючи на інші показники, цією затримкою можна знехтувати.

Отже, як видно з дослідів, кешування і попередні обчислення дозволяють значно зменшити час відгуку і тому вони є дуже корисними для великих веб-сервісів. Але при розробці таких систем треба враховувати специфіку конкретного веб-додатку і, виходячи з неї, балансувати між швидкими відгуками на запити і частотою оновлень даних, занесених у кеш.

Література. **1.** Redis docs, www.redis.io/documentation. **2.** About Memcached, www.memcached.org/about. **3.** datagen.py, http://goo.gl/bpdBHw. **4.** redis_cache_update.py, http://goo.gl/exe6Jd. **5.** mysql_fetch.py, http://goo.gl/O3jRJr. **6.** redis_fetch.py, http://goo.gl/pGVsnM.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Наріжна Н.С.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Рівні тестування розподілених систем

Широке впровадження розподілених інформаційних систем (РІС) є характерним сьогодні майже для всіх галузей людської діяльності, де на них покладається вирішення все важливіших задач. Вони відповідають за виконання технологічних процесів на виробництві, за зв'язок і телекомунікації, енергетичне постачання й транспортне обслуговування, виконання соціальних і медичних програм, за підтримку державної, оборонної, наукової діяльності. З появою й розвитком розподілених систем, постало питання, яким чином можливо протестувати такі складні системи. Одним з найкращим підходом тестування РІС є використання рівнів тестування. Рівні тестування – це методологія поділу системи, що буде тестуватися на логічно-зрозумілі рівні програмного забезпечення. Виділяються такі рівні: компонентне (або модульне), інтеграційне, системне та приймальне тестування [1].

Модульне тестування (компонентне) – тестується мінімально можливий для тестування компонент, наприклад, окремий клас або функція. Часто модульне тестування здійснюється розробниками ПЗ. Метою модульного тестування є ізоляція кожної частини програми та впевненість у тому, що кожна окрема частина є коректною. Модульний тест забезпечує жорсткий «контракт», за яким має працювати тестований код. Модульне тестування допомагає знайти помилки раніше в циклі розробки ПЗ, що робить розробку дешевшою та швидшою.

У розподілених системах головне - це взаємодія між компонентами. За взаємодію між компонентами відповідає інтеграційне тестування. Інтеграційне тестування призначено для перевірки зв'язку між компонентами, а також взаємодії з різними частинами системи [2]. Метою інтеграційного тестування є верифікувати вимоги з функціональності, продуктивності, надійності до основних компонентів програми. Ці компоненти, тобто групи модулів, тестуються методом чорної скриньки, успішні та неуспішні тест-кейси симулюються відповідними вхідними параметрами. Інтеграційне тестування також має два рівні: компонентний інтеграційний рівень – перевірка взаємодії між компонентами системи після проведення компонентного тестування; системний інтеграційний рівень – перевіряється взаємодія між різними системами.

Системне тестування – перевірка інтегрованої системи на відповідність всім вимогам. Перевірка повноти та правильності документації користувача є важливою частиною системного тестування. Всі тестові комбінації повинні розроблятися тільки з використанням документації користувача. Метою системного тестування є виявлення протиріч між розробленою системою та первісними цілями її створення. Компонентами системного тестування є розроблена система ПЗ, кінцеві цілі і вся документація, яка додається до системи [3].

Приймальне тестування – це тестування ПЗ, яке проводять для перевірки на відповідність до вимог замовника. Приймальне тестування - це фінальний етап тестування програми перед публічним запуском. Як тільки прийнято рішення, що продукт повністю готовий до використання - приймальне тестування має підтвердити це. Часто приймальне тестування означає залучення команди незалежних тестерів для аналізу стану продукту. Як правило, подібне тестування виконується, коли замовник хоче дізнатися незалежну думку про якість замовленого ним продукту.

Використання рівнів тестування є сильним помічником при тестуванні розподілених інформаційних систем. За допомогою поділу РІС для тестування, можливо виявити проблемні місця продукту, який тестується, ще на ранній стадії розробки, що впливаэ на ціну та швидкість розробки.

Література. **1.** K. C. Tai, R. H. Carver: Testing of distributed programs; in A. Zomaya (ed.): Handbook of Parallel and Distributed Computing; McGraw Hill; 1995; pp. 956-979. **2.** A. Ulrich, S. T. Chanson: An approach to testing distributed software systems; 15th PSTV 1995; Warsaw, Poland; pp. 107-122; 1995. **3.** R. Binder. Testing object-oriented software: a survey. Software Testing, Verification and Reliability, 6:125–252, 1996.

Науменко Т.О.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Дослідження адаптивних технологій контролю знань

Найслабшою ланкою в навчальному процесі, як правило, є явно недостатня кількість інформації у викладача про якість засвоєння матеріалу, про правильне розуміння його, про самостійну роботу студентів. Подання про це може бути отримано лише наприкінці навчання, та й то лише за деякими курсами і далеко не повне, так як за кілька годин, які відводяться для оцінки знань викладачем під час поточного і рубіжного контролю з цілою групою студентів, важко з'ясувати справжні знання, отримані студентами за весь період навчання.[1]

Основними недоліками існуючих систем контролю знань виступають: перевірка великих блоків матеріалу; суб'єктивна оцінка результатів навчання; одноманітність використання комп'ютерних та електронно-обчислювальних засобів; використання результатів контролю, що не дозволяє отримати цілісної системної картини стану освітнього процесу; відсутність можливості найбільш повно враховувати індивідуальні особливості студентів.

Адаптивна технологія навчання враховує особливості кожного етапу переходу на новий рівень знань і уточнення успішності навчальної діяльності на кожному такому етапі, тобто моделі адаптивного навчання відбивають циклічність навчального процесу.[2] Однією з головних складових адаптивної технології навчання є методи проведення контролю знань – адаптивне тестування.

У західній літературі виділяється три варіанти адаптивного тестування. **Пірамідальне тестування** - всім претендентам дається завдання середньої складності і вже потім, залежно від відповіді, кожному дається завдання легше чи важче. На кожному кроці застосовується правило ділення шкали складності завдань навпіл; **Flexilevel** - коли контроль починається з будь-якого рівня складності, а потім відбувається поступове наближення до реального рівня підготовленості; **Stradaptive** - коли тестування проводиться з банком тестових завдань, де завдання розділені за рівнями складності. Після вірного виконання, наступне завдання береться з більш високого рівня складності, після невірного - навпаки.[3] Аналіз літератури визначив більш детальний розподіл методів проведення контролю знань: 1 - не адаптивні (строга послідовність; випадкова вибірка; комбінований метод), 2 – частково адаптовані(випадкова вибірка з урахуванням окремих параметрів моделі студента; контроль на основі відповідей студента; контроль на основі моделі навчального матеріалу; модульно-рейтинговий метод), 3 – повністю адаптивні(контроль по моделі студента; контроль по моделях студента і навчального матеріалу).[4]

В результаті досліджень були знайдені такі способи вирішення проблем:

- Застосування одразу декількох адаптивних методів контролю знань для виявлення більш повної картини розуміння студентом предмету вивчення.
- Впровадження штучного інтелекту, що полегшить покращення вже існуючих та розробки нових методів контролю знань
- Ширше використання новітніх технологій(наприклад, мобільні додатки, тестування за допомогою кмп'ютерних ігор тощо)

Література. **1.** Юсупов Д. Ф. Нейросетевые технологии адаптивного обучения и контроля знаний студентов по курсу основы программирования [Текст] / Д. Ф. Юсупов // Молодой ученый. — 2014. — №6. — С. 779-783. **2.** Роберт И. Новые информационные технологии в обучении: дидактические проблемы, перспективы использования //Информатика и образование. -1991. -№4. - с. 18-25 **3.** Weiss D. J.(Ed.) New Horizons in Testing: Latent Trait. Test Theory and Computerised Adaptive Testing. N.Y., Academic Press, 1983. – 345 pp. **4.** Зайцева Л.В., Прокофьева Н.О. Модели и методы адаптивного контроля знаний // Educational Technology & Society. - Nr.7(4), 2004 ISSN 1436-4522 (Международный электронный журнал). / Интернет. - http://ifets.ieee.org/russian/periodical/journal.html

Негода К.В., Ковальчук-Химюк Л.О.
Національний технічний університет України "КПІ", Київ, Україна

Дослідження динаміки ендогенного економічного зростання в Україні протягом 2007–2012 рр. за моделлю Ромера

Світова економічна криза актуалізувала проблеми, пов'язані з економічним зростанням. Ендогенні теорії економічного зростання, крім праці та фізичного капіталу, ураховують чинники, від яких залежить продуктивність праці: людський капітал, навчання працівників у процесі діяльності, витрати на науково-дослідні роботи.

Основною метою даної роботи є аналіз особливостей моделей ендогенного економічного зростання [1] та дослідження можливості існування стійкого росту з постійним темпом приросту на основі технічного прогресу, який є наслідком навчання робітників у процесі своєї діяльності.

Для аналізу динаміки ендогенного економічного зростання було обрано модель Ромера навчання в процесі діяльності (моделі з постійною нормою заощадження). Основна ідея П. Ромера полягає в тому, що «існує обмін між споживанням сьогодні та знаннями, які можна буде використовувати для поширення споживання завтра» [2].

Виробнича функція в моделі П. Ромера має наступний вигляд [3]:

$$Y_t = K_t^{\alpha} \left(A_t L_t\right)^{1-\alpha} , \tag{1}$$

де K_t — капітал, під яким розуміють суспільний капітал знань у періоді t; A_t — функція навчання робітників на практиці; α $(0 < \alpha < 1)$ — технологічний параметр за фактором фізичного капіталу. Функція A_t в (1) залежить від загального обсягу капіталу в економіці:

$$A_t = K_t^{\phi} ,$$

де ϕ — параметр ефективності навчання.

У моделі П. Ромера темп економічного зростання знаходиться в прямій залежності від величини людського капіталу, зосередженого у сфері отримання нового знання, передбачається вільне розповсюдження знань між робітниками.

Для проведення розрахунків було створено програмне забезпечення в середовищі програмування R-studio [4]. У якості вихідних даних було взято статистичні дані економічного зростання України за 2002–2012 рр. У результаті проведення розрахунків було отримано значення функції Y за кожен рік залежно від параметру ефективності навчання ϕ. Розглянуто два випадки: постійна віддача від навчання $\phi = 1$ та спадна віддача — $0 < \phi < 1$.

Аналіз результатів (табл. 1) показав, що постійне екзогенне економічне зростання за базових припущень даної моделі можливе тільки в ситуації спадної віддачі від навчання, а ендогенне — за постійної віддачі, якщо відсутнє зростання чисельності населення в країні.

Табл. 1. Значення виробничної функції

	2007	2008	2009	2010	2011	2012
Y за $\phi = 1,0$	17193560,46	26510100,11	31996023,02	54636752,84	60906300,72	75402593,33
Y за $\phi = 0,7$	812858,323	1144924,771	1320947,697	2017009,563	2198675,486	2603200,51
Y за $\phi = 0,5$	106279,022	140935,589	157789,719	223626,539	240156,166	276008,112

Література. **1.** Сеницц П. Сутність та види економічного зростання: історико-аналітичний аспект / П. Сеницц // Вісник ТНЕУ. — 2013. — № 2. — С. 100–107. **2.** Дагаев А. Новые модели экономического роста с эндогенным технологическим прогрессом / А. Дагаев // Мировая экономика и международные отношения. — 2001. — № 6. — С. 43. **3.** Шараев Ю. В. Теория экономического роста: учеб. пособие для вузов; Гос. ун-т – Высшая школа экономики / Ю. В. Шараев. — М. : Изд. дом ГУ ВШЭ, 2006. — С. 77–81. **4.** Наглядная статистика. Используем R! / Шипунов А.Б., Балдин Е. М., Волкова П. А. и др. — М. : ДМК Пресс. — 2012. — 297 с.

***Перевертайло В.В.*[1], *Майкут С.О.*[1], *Чернацький А.Ю.*[2]**
[1] *Національний технічний університет України "КПІ", факультет електроніки, Київ, Україна;*
[2] *Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна*

Система автоматичного моніторингу параметрів приміщення на базі Arduino Mega 2560

У сучасному інформаційному світі з кожним днем зростає можливість для створення безпечніших умов життя людини, місця роботи чи іншого місця перебування. Це досягається різноманітними методами реєстрації та передачі інформації для подальшої її обробки та відповідного реагування. Отримувана інформація дозволяє проводити аналіз умов середовища з метою періодичного тестування і, у разі необхідності, підтримання необхідних параметрів мікроклімату.

Все частіше керування багатьма функціональними процесами відбувається за допомогою програмних пристроїв, які характеризуються високою швидкістю реагування, багатозадачністю, достатньо довгим часом безвідмовної роботи та ін.

У нашій роботі представлено систему, яка дозволяє дистанційно проводити реєстрацію параметрів приміщення, в якому знаходиться котельня житлового будинку.

Параметри, які необхідно контролювати впродовж роботи котельні: температура, вологість, витоки газу і тиск.

З метою побудови відповідної системи було вивчено сучасні пропозиції, які пропонуються на ринку інформаційних технологій. У результаті, було отримано висновок, що системи такого рівня достатньо дорогі для пересічного користувача, тому в основі нашої розробки ми поклали мету: конструювання дешевого та простого у розробці технологічно-інформаційного комплексу, який дозволить проводити дистанційний контроль теплозабезпеченості будівлі.

Для реалізації поставленої задачі необхідно було опрацювати певні джерела [1–3], які допомогли нам у подальшому спроектувати цифрову систему на кристалі, зв'язати її з комп'ютером для програмування та обробки інформації від датчиків раніше згадуваних вимірюваних параметрів.

Arduino на сьогодні у достатньому обсязі забезпечує користувача певним набором уже готових мікроконтролерних плат. Використовуючи такі плати та необхідні датчики, можна створити систему моніторингу та розумного аналізу типу «Розумний дім». Апаратні засоби із можливістю розширення та відкритими принциповими схемами (мікроконтролери ATmega8, ATmega168, ATmega2560 та ін.) є основою Arduino [4]. Такий підхід дозволяє самостійно розробляти прототипи систем на мікроконтролері з метою економії коштів та розуміння принципу її роботи.

У нашій роботі розроблена платформа на базі Arduino, яка буде проводити моніторинг мікроклімату в приміщенні. При цьому обробка інформації та керування може виконуватись як на локальному комп'ютері, так і на віддаленому сервері, зв'язаному через Інтернет (див. рис. 1).

Arduino Mega 2560 – це пристрій, основою якого є мікроконтролер (МК) ATmega2560. У склад Arduino Mega 2560 входить усе необхідне для зручної роботи з МК: 54 цифрових входи/виходи, із яких 15 можуть використовуватись у якості ШІМ-виходів; 16 аналогових входів, 4 UART (апаратних прийомопередавачів для реалізації послідовних інтерфейсів); USB-роз'єм; роз'єм для живлення; ICSP-роз'єм для внутрішньо схемного програмування та кнопка скидання. Крім того, Arduino Mega сумісний з багатьма платами розширення, що розроблені для Arduino Duemilanove и Diecimila. Мікроконтролер ATmega2560 має 256 кБ флеш-пам'яті, 8 кБ SRAM, 4 кБ EEPROM , тактову частоту 16 МГц, напругу живлення 4,5 – 5,5 В.

Для живлення плати була вибрана модульна схема джерела живлення постійної напруги 5 В. Для усіх елементів створеного пристрою необхідна постійна напруга, яку вони споживають від відповідних виводів плати Arduino Mega 2560.

Датчик для вимірювання температури було вибрано DS18B20, оскільки цей цифровий

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

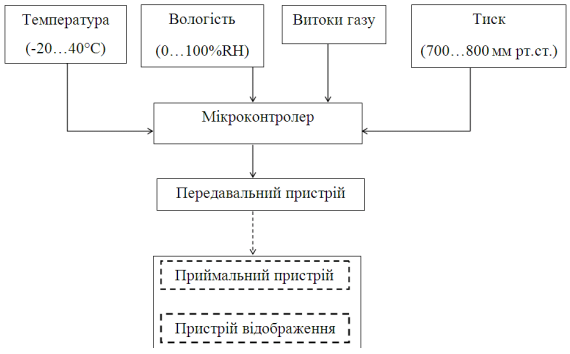

Рис. 1. Структурна схема вимірювального пристрою

термометр є доволі дешевим за собівартістю, простим у під'єднанні до мікроконтролера за допомогою шини 1-Wire і надійним, оскільки температурні дані передаються на МК відразу у цифровому форматі. Робочий інтервал температур лежить у межах від 55°C до +125°C, а діапазон напруги живлення: 3,0 В. . . 5,5 В.

Датчик для вимірювання вологості був вибраний DHT22. Крім вологості, цей датчик дозволяє проводити вимірювання температури, що у подальшому дозволило нам порівнювати отримувані данні від обох датчиків: DS18B20 та DHT22. Він здатний вимірювати вологість від 0 до 100%RH при вхідній напрузі живлення 3,3 В. . . 6 В.

Для визначення витоків газу використано датчик MQ4 – це аналоговий датчик, який високочутливий до витоків природного газу, зрідженого природного газу, метану у повітрі, також сигаретного та кухонного диму. Цей датчик став одним із елементів нашої системи завдяки своїм параметрам, простоті регулювання та низькій вартості. Чутливість датчика можна регулювати за допомогою потенціометра, напруга живлення 5 В.

Для вимірювання атмосферного тиску ми визначили датчик BMP180, що покриває діапазон від 225 до 825 мм рт.ст. Даний датчик доволі дешевий, хоч і охоплює вельми великий діапазон.

Для передачі інформації, отриманої з датчиків, до користувача через Internet до мікроконтролера під'єднується плата розширення Ethernet Shield W5100. Вона дозволяє дуже легко підключити Arduino до мережі Internet. При цьому для веб-доступу реалізована відповідна аутентифікація користувача, також є можливість запису даних, що надходять від кожного датчика, на карту пам'яті microSD.

Розробка програми для мікроконтролера проводилась у середовищі "Середовище розробки Arduino", яке містить вбудований редактор програмного коду, області повідомлень, вікна виведення тексту, панелі інструментів з кнопками часто використовуваних команд та декількох меню. Для завантаження програми та зв'язку середовище розробки підключається до апаратної частини Arduino. "Середовище розробки Arduino" компілює програмний код та завантажує його до пристрою Arduino.

Як результат, ми отримали систему на базі мікроконтролера ATmega2560 фірми AVR для реєстрації температури, вологості, витоку газу і тиску. Створена система повністю відповідає поставленій задачі, є простою у реалізації, конструюванні, програмуванні, збереженні та відображенні інформації, також є дешевою у порівнянні із сучасними аналогами серійного виробництва.

Література. **1.** М. Э. Сопер. Практические советы и решения по созданию "Умного дома". - НТ Персс, 2007. - 432с. **2.** В.Н. Гололобов. "Умный дом" совими руками. - М.: НТ Пресс, 2007. - 416с. **3.** Роберт К. Элсенпитер, Тоби Дж. Велт. Умный Дом строим сами. - КУДИЦ-Образ, 2005. - 384с. **4.** Программирование Ардуино, arduino.ru/Reference.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Попеляев Д.П.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Применение событийно-ориентированного подхода к разработке веб-сервисов

Современные веб-проекты представляют собой высоконагруженные веб-сервисы, количество запросов к которым за день измеряется тысячами. Такие проекты стремятся предоставить пользователям как можно больше возможностей, большую производительность и надежность. В этих условиях необходимо учитывать такие параметры, как сложность поддержки, высокую связность модулей, которая усложняет их модификацию, надежность компонентов системы и другие.

Классический подход к компоновке веб-приложений. В наши дни сформировался «классический» подход к компоновке веб-приложения, который представляет собой такую схему:

1. Сервера, которые обслуживают frontend предназначены для отдачи статических ресурсов (страниц, изображений, таблиц стилей, скриптов) и распределения сетевого трафика по нижестоящим узлам.
2. Нижестоящие узлы (backend) занимаются динамическими вычислениями.
3. Также существует группа серверов которые хранят данные (например, база данных MySQL).
4. Сам код веб-сервиса скомпилирован на все узлы системы и одинаково обрабатывает запросы, пришедшие на тот или иной узел.

Такое решение влечет за собой некоторые недостатки, основными из которых являются следующие:

- расширение структуры новыми серверами приводит к необходимости изменять схему распределения сетевой нагрузки всего ресурса;
- поставка новой версии продукта занимает длительное время, так как есть необходимость физической загрузки информации на все машины системы, а также одновременного завершения обработки запросов неактуальной версией продукта на всех серверах (в противном случае возникнут конфликты, при которых часть запросов пользователей всё еще будет обрабатываться уже неактуальной версией продукта, а часть будет обработана новой версией);
- использование протокола HTTP, как технически устаревшего, является недостатком. Протокол HTTP со стороны клиента требует ответа от сервиса в ограниченное время. Это обязывает сервер вычислить и подготовить ответ в строгом временном промежутке. Если сервер в этот момент занят – запрос теряется [1].

Более современным подходом к компоновке веб-приложений является применение сервисно-ориентированной архитектуры.

Сервисно-ориентированная архитектура (SOA). При построении проекта в СОА, каждая независимая логическая часть обрабатывается отдельным модулем – сервисом. Сервисы соединены между собой независимой асинхронной шиной обмена. При разбиении системы на модули возрастает ее надежность, простота в поддержке и расширении [2].
В классическом подходе к программированию веб-сервисов проектирование начинается от моделей, их свойств и операций над ними. Модели отображают объекты реального мира, а операции – действия над ними. Однако, как показывает современная практика, мир намного эффективней описывается языком событий и реакций на них. На базе сервисно-ориентированной архитектуры возможно применение событийно-ориентированного подхода.

Событийно-ориентированный подход. Событийно-ориентированный подход предполагает, что модули (сервисы) общаются между собой путем отправки асинхронных событий. Событие – это сообщение, которое имеет название и набор параметров. Событие предназначено для того, чтобы сообщить системе о том, что что-то произошло, или дать запрос в систему о

необходимости получить информацию.

Стоит отметить, что событие отправляется в систему безадресно (или широковещательно) – то есть конкретные узлы-получатели не указываются. Напротив, конкретные узлы (компоненты, сервисы) слушают общую шину на предмет тех или иных событий, на которые они готовы прореагировать. Это может означать, что сервис готов услышать какое-то событие и выполнить соответствующее действие, или же сервис располагает какими-то знаниями (например, владеет базой данных со сведениями о пользователях) и готов сообщить их в ответ на запрос. В обоих случаях, результат реакции на событие порождает новое событие (с другим именем и другими параметрами), которое так же могут услышать другие заинтересованные в этом сервисы. Краткий список основных характеристик событийно-ориентированной архитектуры изложен в таблице 1.

Таблица 1. Основные характеристики событийно-ориентированной архитектуры

Характеристика	**Описание**
Отсутствие строгой связи взаимодействий модулей	Модуль, который сообщает о событии, не должен иметь информацию об узлах-получателях
Тип связи “многие ко многим”	Общение посредством сообщений о событиях / запросов на получение информации, в котором каждое событие получается многими узлами одновременно
События как основной элемент управления	Поведение модулей регулируется получаемыми и исходящими событиями
Асинхронность работы	Поддержка асинхронных операций посредством общения событиями

Преимущества событийно-ориентированного подхода. При правильной организации общей шины, сервисы отправляют и получают события асинхронно, не ожидая друг друга. Это значит, что сервис-отправитель может отправить сколько угодно событий в общую шину без временных задержек и перейти к решению следующих задач. Сравнивая это с классическим протоколом HTTP, подразумевающим ожидание ответа в текущем цикле обработки, становятся явными преимущества событийно-ориентированного подхода. Сервисы-получатели будут получать новые события также асинхронно, независимо друг от друга, сразу же по завершению обработки предыдущего события. Также исчезает проблема с поставкой новой версии продукта, так как в данном подходе система разделена на модули и есть возможность проводить обновление ПО на каждом модуле поочередно, не прерывая работы всей системы. Это производится путем отключения модуля на необходимое количество времени, на протяжении которого все запросы к этому сервису будут ждать в очереди, пока сервис снова не приступит к работе. Данное мероприятие никак не отобразится на работе всего ресурса.
Таким образом, можно сказать, что применение событийно-ориентированного подхода является оптимальным и прогрессивным решением в области разработки веб-сервисов.

Применение. Событийно-ориентированный подход будет использован для оптимизации взаимодействия модулей системы GridALLTED, которая предназначена для моделирования и анализа динамических систем, состоящих из компонентов различной физической природы – электронных, гидравлических, пневматических, механических и других.

Литература. **1.** IBM Technical library. Patterns: Service-Oriented Architecture and Web Services. April 2004. http://www.redbooks.ibm.com/redbooks/pdfs/sg246303.pdf. **2.** Petrenko A. Service-Oriented Architecture for Grid-Enabled Computer Simulation Software / Anatolii Petrenko, Volodymyr Ladogubets, Oleksii Finogenov, Bogdan Bulakh // Perspective Technologies and Methods in MEMS Design:8-th International Conference MEMSTECH’2012,18-21 April 2012, Polyana-Svalyava (Zakarpattya),Ukraine:proc.–Lviv:Publishing House Vezha&Co,2012.–P.179–181.

Прогонов Д.А., Паничева Д.О., Кущ С.Н.
Физико-технический институт НТУУ "КПИ", Киев, Украина

Анализ точности статистических методов обнаружения стеганограмм с данными, встроенными в области преобразования контейнера

При проведении атак на информационные ресурсы частных компаний и государственных структур широко используются скрытые (стеганографические) системы связи (ССС). Особенностью ССС является встраивание канала обмена сообщениями в существующие информационные потоки, что затрудняет их обнаружение при использовании современных систем контроля трафика. Актуальной является задача поиска новых методов пассивного стегоанализа, позволяющих повысить точность обнаружения скрытых сообщений.

Наиболее часто в качестве файлов-контейнеров для встраивания стегоданных используются мультимедийные данные, в частности цифровые изображения (ЦИ), что объясняется значительной избыточностью их цифрового представления. В зависимости от области скрытия сообщений, известные методы стеганографии ЦИ возможно разделить на две группы — встраивание сообщений в пространственной области ЦИ (LSB-методы) и в области преобразования контейнера (ОПК). Для распознавания стеганограмм, сформированных согласно LSB-методам, были предложены эффективные методы статистического стегоанализа [1], в то время как для методов скрытия в ОПК в настоящее время не известны универсальные алгоритмы обнаружения, что обусловлено разнообразием методов обработки изображения-контейнера (ИК) и стегоданных.

Ранее авторами была показана высокая эффективность использования статистической модели (СМ) SPAM ЦИ для распознавания стеганограмм с данными, встроенных в спектральной области ЦИ с использованием двумерного дискретного вейвлет преобразования (ДДВП) [3]. Для повышения точности распознавания стегоданных, скрытых в ОПК, представляет интерес применение специализированных СМ ИК в частотной области .

Целью работы является сравнительный анализ точности обнаружения стеганограмм с данными, встроенными в ОПК с применением как стандартных (ДДВП), так и специальных преобразований (сингулярное разложение (СР)) ЦИ, в случае использования СМ ИК в пространственной и частотной областях.

Материалы и методы. В работе для скрытия сообщений в ОПК были использованы методы Дея (ДДВП) [4] и Агарваля (СР) [5]. Формирование стеганограмм проводилось путем сложения коэффициентов разложения изображения-контейнера K_{cover} и стегоданных K_{data}, представленных в виде ЦИ, с весовым коэффициентом G:

$$K_{stego} = K_{cover} + G \cdot K_{data},$$

где K_{stego} — коэффициенты разложения стеганограммы. Для получения стеганограмм в пространственной области к полученным коэффициентам K_{stego} применялось обратное ДДВП (метод Дея) либо производилось умножение матриц сингулярных чисел и сингулярных векторов (метод Агарваля).

Для обнаружения стеганограмм были использованы модели ИК в пространственной (SPAM модель [3]) и частотной (CC-PEV модель [6]) областях. Модель SPAM основана на использовании теории марковских процессов для анализа изменений разниц яркостей смежных пикселей при скрытии сообщений в ЦИ. Модель CC-PEV применяется при стегоанализе ЦИ, представленных в формате JPEG, и основана на исследовании корреляционных связей, как между коэффициентами дискретного косинусного преобразования (ДКП), внутри отдельного блока, так и между блоками разбиения ИК.

Для анализа изменений CC-PEV модели ЦИ при встраивании стегоданных были использованы следующие характеристики:

- Гистограмма H значений коэффициентов ДКП;
- Гистограммы $h^{i,j}$ значений низкочастотных коэффициентов ДКП;

- Вариация $g^d_{i,j}$ значений коэффициентов ДКП по всем блокам разбиения ЦИ;
- Вариация V значений коэффициентов ДКП в смежных блоках разбиения ЦИ;
- Вариация B_α значений смежных пикселей по всем блокам разбиения ЦИ;
- Матрицы сочетаемости $C_{s,t}$ значений коэффициентов ДКП.

Для повышения точности обнаружения стеганограмм производилась калибровка модели CC-PEV [6] — оценивание параметров исходного ИК на основании анализа заданной стеганограммы. Процесс калибровки состоял из нескольких этапов: декомпрессии ЦИ, кадрирования полученного изображения и повторного JPEG-сжатия.

Полученные результаты. Исследование проводилось с использованием псевдослучайной выборки 2500 ЦИ из тестового пакета MIRFlickr-25k. В качестве стегоданных были использованы ЦИ с различной степенью детализации: чертеж, карта и портрет. Степень заполнения контейнера стегоданными (количества измененных коэффициентов разложения ИК к общему количеству коэффициентов разложения) изменялась от 5% до 25%, с шагом 5%, и от 25% до 95% с шагом 10%. Значения весового параметра G варьировались от $G_{min} = 0.02$ (нижняя граница восстановления стегоданных) до $G_{max} = 0.08$ (появление визуальных искажений ИК) с шагом $\Delta_G = 0.02$.

Распознавание стеганограмм проводилось с использованием ансамбля линейных дискриминантов Фишера. Для оценки ошибки распознавания (ОР) стеганограмм, тестирование стегодетектора повторялось 10 раз при различных разбиениях тестовой выборки ЦИ на обучающую и контрольную подвыборки. По результатам анализа построены зависимости ОР от степени заполнения контейнера стегоданными и весового параметра G.

Модель CC-PEV была специально разработана для обнаружения стеганограмм с данными, встроенными в частотной области, поэтому ожидалось, что её применение позволит существенно уменьшить значение ОР по сравнению со случаем использования SPAM модели. Однако, по результатам анализа было установлено, что ОР для SPAM модели оказалась существенно меньше в области средних и сильных заполнений (от 25%) ИК в диапазоне от средних до больших значений G ($G \geqslant 0.04$). Например, величина ОР для метода Дея/Агарваля при использовании CC-PEV модели существенно выше (27% / 41%), чем для SPAM модели (14% / 28%) [3].

Выводы. В работе впервые проведен сравнительный анализ эффективности использования СМ ИК в пространственной (SPAM модель) и частотной области (CC-PEV модель) для обнаружения стегоданных, встроенных в ОПК. Установлено, что точность распознавания стеганограмм с использованием модели ЦИ в частотной области при среднем и сильном заполнении контейнера стегоданными существенно ниже, чем в случае применения СМ ЦИ в пространственной области. Полученные результаты могут быть использованы при адаптации СМ ЦИ для распознавания стеганограмм с данными, встроенными в ОПК.

Литература. **1.** Kodovsky J. and Fridrich J. “Calibration revisited”. Proc. of the 11th ACM Multimedia and Security Workshop, Princeton, NJ, September 7–8, 2009. **2.** Progonov D.O. and Kushch S.M. On the Multifractal Analysis of Steganograms. Proc. of the 4th International Scientific Conference “Theoretical and Applied Aspects of Cybernetics”. – Kyiv: Bukrek, 2014. – pp. 32-38. **3.** Прогонов Д.О., Кущ С.М. Статистичний аналіз стеганограм з даними, прихованими в області перетворення контейнеру. – Збірник матеріалів наук.-практ. конференції “Актуальні проблеми управління інформаційною безпекою держави”. – К.: Центр навч., наук. та період. видань НА СБУ, 2015. – с. 329-332. **4.** Dey N. et al. A novel approach of color image hiding using RGB color planes and DWT. International journal of computer application, 2011. Vol. 36, No.5. pp.19-24. **5.** Agarwal R. and Santhanam M.S. Digital watermarking in the singular vector domain. International Journal of Image and Graphics, 2008. Vol. 8. pp. 351-362. **6.** T. Pevny and J. Fridrich. Merging Markov and DCT features for multiclass JPEG steganalysis. Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX. — San Jose, CA, 2007. — Volume 6505, pages 1–14.

Рева К.В., Родіонов А.М.
Національний технічний університет України "КПІ", Київ, Україна

Керування пулом ресурсів із застосуванням машинного навчання з підкріпленням

Вступ. Задача динамічного розподілу ресурсів між багатьма клієнтами є досить актуальною [1], [2], особливо у зв'язку зі стрімким розвитком "хмарних обчислень" (Cloud computing), які пропонують користувачам сервіси типу PaaS (Platform-as-a-Service), SaaS (Software-as-a-Service), IaaS (Infrastructure-as-a-Service) [3]. Кожен з цих сервісів, зазвичай, представляє собою пул певних ресурсів, підтримку та прогнозування динаміки використання яких має забезпечувати провайдер "хмарних обчислень".

Розглядається проблема динамічного розподілу ресурсів між декількома пулами (сукупність ресурсів та системи обробки завдань, що використовує дані ресурси). Для вирішення даної проблеми розглядається модель на основі машинного навчання. Аналізуються результати її моделювання та описуються можливі варіанти покращення ефективності роботи моделі.

Постановка задачі. Загальна ідея базової моделі в тому, що кожен пул вираховує значення ефективності своєї роботи та найбажанішу для нього дію (з урахуванням, що трансфер ресурсів не безкоштовний і бездіяльність може бути вигідніша для конкретного пулу в деякий момент часу) і відправляє ці значення деякому арбітру. Арбітр на основі значень ефективності обирає пул, дія якого принесе найбільшу користь системі, та виконує відповідну дію, що пропонує пул (забираючи, якщо потрібно, ресурси з менш ефективних пулів).

Даний підхід не бере до уваги розподілення часу роботи між конкретними завданнями в конкретному пулі, що при наявності завдань з різними параметрами (час на виконання, необхідні ресурси) може привести до їх нерівномірного опрацювання. Введення внутрішнього агента, який буде розподіляти ресурси між завданнями на основі їх часу зберігання в черзі та загального часу обробки завдань, може збільшити ефективність роботи всієї системи.

Загальний опис підходу для базової моделі. Для кращого розрахування ефективності роботи кожного пулу, має сенс використати метод машинного навчання з підкріпленням, хоче існує безліч інших методів та підходів до вирішення даної проблеми [4], [5]. Машинне навчання з підкріпленням - підрозділ машинного навчання, що вивчає як агент повинен діяти в деякому оточенні, щоб максимізувати довготривалу винагороду [6]. В даній задачі, агент - пул, що намагається мінімізувати чергу завдань та максимально ефективно використати наявні ресурси для завдань, що потрібно опрацювати. Оточення, в якому працює агент, зазвичай описується як марковський процес прийняття рішень (S, A, r, T), де

- S – фінітна множина станів
- A – фінітна множина дій
- $r : S \times A \times S \mapsto \mathbb{R}$ – функція винагороди
- $T : S \times A \mapsto PD(S)$ – функція переходу станів, що відображає поточний стан та дію агента у набір ймовірнісних розподілів на S

На кожному кроці t, агент спостерігає стан $s_t \in S$, обирає дію $a \in A$, змінює свій стан згідно з функцією T, що залежить тільки від s_t і a_t, та отримує винагороду $r(s_t, a_t, s_{t+1})$. Ґрунтуючись на такій взаємодії з навколишнім середовищем, агент, який навчається з підкріпленням, повинен виробити стратегію $\pi : S \mapsto A$, яка максимізує величину $R = r_0 + r_1 + ... + r_n$. Один із підходів - використання TD(0) (temporal difference(TD) learning). При такому підході, функція довготривалої цінності має вигляд [7]:

$$V^{\pi}(s_t) \leftarrow V^{\pi}(s_t) + \alpha_t(r(s_t, \pi(s_t), s_{t+1}) + \gamma V^{\pi}(s_{t+1}) - V^{\pi}(s_t))$$

При умові, що агент буде слідувати стратегії π:

$$\pi'(s) = argmax E[r(s_t, a, s_{t+1}) + \gamma V^{\pi}(s_{t+1})]$$

Але, беручи до уваги, що простір станів занадто великий для даної задачі, потрібно ввести ряд спрощень, що зменшить складність обчислення функції цінності.

Основні положення базової моделі. Ідея в тому, щоб наблизити функцію цінності. Базуючись на [7], намагаємось наблизити істину функцію цінності за допомогою:

$$\hat{V}(s,p) = \sum_{i=1}^{M} p^i \phi^i(s)$$

де: $p^i_{t+1} = p^i_t + \alpha_t[r(s_t, \pi(s_t), s_{t+1}) + \gamma \hat{V}(s_{t+1}, p_t) - \hat{V}(\bar{s_t}, p_t)]\phi^i(\bar{s_t})$
є параметрами навчання, що налаштовуються з часом, а

$$\phi^i(s) = \frac{w^i(x)}{\sum_{i=1}^{M} w^i(x)}$$

являє собою базис для розкладу параметрів. M - кількість категорій, на які можна класифікувати кількість необхідних ресурсів для завдання, w^i - деякий базисний набір функцій. Також, якщо вважати перехід між станами миттєвим, то можна задати стратегію вибору найліпшої дії як [7]:

$$\pi'(s) = argmax[r(s_t, a, \bar{s_t}) + V^{\pi}(\bar{s_t})]$$

Функція миттєвої винагороди повинна залежати від значення передбачуваної довжини черги для даного пулу (B), а також від ціни трансферу ресурсів між пулами (TC):

$$r(s_t, s_{t+1}) = B(s_t, s_{t+1}) + TC(s_t, s_{t+1})$$

Попереднє моделювання. Використовуючи вищенаведену модель, було отримано ряд результатів для різного набору параметрів. Далі розглянемо результати моделювання одного з пулів з початковим розподілом одного з ресурсів (6,5,7) відповідно для трьох пулів та 10000 ітерацій. Зміна довжини черги для 2-го пулу показана на рис.1, а відповідна їй кількість доступних ресурсів для даного пулу показана на рис.2. Як можна побачити, різке збільшення довжини черги для даного пулу, зумовлене відбиранням ресурсів з цього пулу, було подолано досить швидко відносно загального часу роботи. Система при виявленні нестабільності змогла відновити нормальний стан для даного пулу.

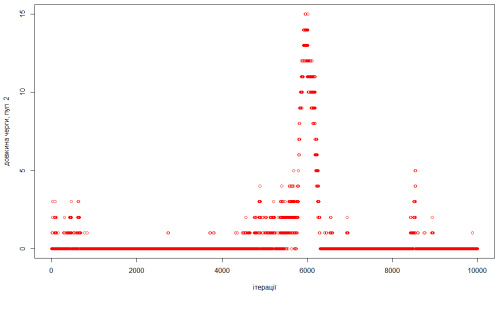

Рис. 1. Залежність довжини черги в другому пулі від номеру ітерації

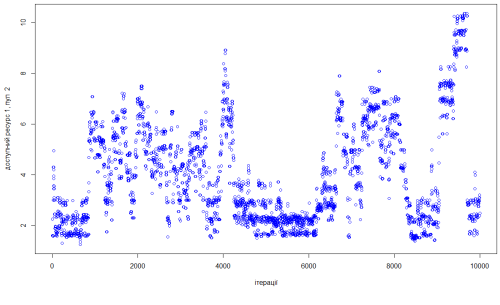

Рис. 2. Залежність кількості доступного ресурсу для другого пулу від номеру ітерації

Агент внутрішнього стану. Описана вище модель дозволяє досить ефективно розподіляти ресурси між пулами, але при даному підході не береться до уваги ефективність розподілу часу на обробку завдань всередині самого пула. Введення окремого агента для вирішення даної задачі повинно збільшити ефективність роботи системи в цілому. Основна ідея: спостерігати за виконанням завдань в пулі і, базуючись на часі опрацювання кожного завдання та часі знаходження завдань в черзі, обирати ліпшу стратегію дій для максимально ефективної

обробки всіх проектів.

Введемо агента, навчання якого теж базується на підході машинного навчання з підкріпленням, де за функцію винагороди візьмемо:

$$r(s_t, s_{t+1}) = PT(s_t, s_{t+1}) - \theta QT(s_t, s_{t+1})$$

де $PT(s_t, s_{t+1})$ - сумарний час, витрачений на обробку завдань, а $QT(s_t, s_{t+1})$ - сумарний час знаходження завдань в черзі, θ - коефіцієнт, який потрібно підібрати експериментальним шляхом.

Даний підхід дасть змогу досягти більшої ефективності у виділенні ресурсів для завдань, зважаючи на час та ресурси для їх виконання.

Висновки. Було проведено моделювання планувальника ресурсів, в основі якого лежить машинне навчання з підкріпленням. Результати моделювання показали життєздатність описаної моделі, а також необхідність подальших експериментів для знаходження оптимальних значень початкових налаштувань та граничних умов застосовності даної моделі. Крім того, було запропоновано покращення загальної ефективності роботи систем, шляхом введення додаткового внутрішнього агента для кожного пулу, що буде намагатися оптимально розподілити час на обробку завдань для кожного пулу.

Література. **1.** Oracle Cloud's Platform as a Service: Innovate, Integrate, and Collaborate — http://www.oracle.com/us/solutions/cloud/platform-as-a-service/oracle-cloud-paas-brief-2411348.pdf —

2. Шмойлов Дмитрий Викторович — ОБЛАЧНЫЕ ВЫЧИСЛЕНИЯ: АКТУАЛЬНОСТЬ И ПРОБЛЕМЫ — Электронное научное издание "Электроника и информационные технологии" — ГОУВПО "Мордовский государственный университет им. Н.П. Огарева" — 2011

3. Why "PaaS" is seen and considered as the future of cloud computing? — http://marketrealist.com/2014/07/why-paas-is-considered-the-future-of-cloud-computing/ —

4. А.В. Прохоров, Е.М. Пахнина — Мультиагентные технологии управления ресурсами в распределенных вычислительных средах — Second International Conference "Cluster Computing" CC 2013 — с. 184-190

5. С.В. Минухин, С.В. Баранник — Метод планирования ресурсов в распределенной гетерогенной системе и исследование его практической реализации — Міжнародна конференція "Високопродуктивні обчислення" HPC-UA'2012 — с. 247-254

6. Обучение с подкреплением / Р.С. Саттон, Э. Г. Барто; пер. с англ — М. : БИНОМ. Лаборатория знаний, 2011. — 399 с.

7. D. Vengerov. A Reinforcement Learning Approach to Dynamic Recource Allocation. Eng. Appl. of AI, 20 (3) — 2007. — pp.383-390

Ротенберг А.В.
Национальный технический университет Украины "КПИ", факультет информатики и вычислительной техники, Киев, Украина

Выбор модели данных СУБД: специфика примененинния SQL и NoSQL

Сегодня, в современных условиях, тяжело представить приложение, не имеющее базы данных. То ли это простой сайт, то ли сложное корпоративное приложение, всегда есть необходимость записи, чтения и хранения данных.

Для работы с данными используются СУБД (Система Управления Базой Данных). На данный момент СУБД основывается на двух основных моделях данных – SQL (structured query language — «язык структурированных запросов») и NoSQL (Not Only SQL или No SQL).

Прежде, чем приступать к написанию программного кода проекта, перед разработчиком стоит задача выбора средств разработки, в том числе и выбор модели базы данных. Для выбора модели построим сравнительную таблицу между SQL и NoSQL (табл.1).

Таблица 1. Сравнительная таблица SQL и NoSQL

Критерий	SQL	NoSQL
Хранение данных	Данные хранятся в строках и столбцах. Строка содержит информацию о одной сущности, а столбец об определенной характеристике. Например, в строке хранится информация о работнике, а столбцы хранят информацию о зарплате, специализации и т.д.	Термин NoSQL включает в себя множество баз данных, имеющие различные модели хранения данных. Основные модели это : документно-ориентированные, ключ-значение, колоночно -ориентированные.
Гибкость	Каждая запись соответствует фиксированной схеме, это значит, что каждый столбец должен быть создан до записи, и каждая строка должна иметь данные в каждой колонке.	Используются динамические схемы. Записывать данные можно без создания столбцов, и каждая строка не должна содержать данные в каждой колонке.
Масштабируемость	Используется вертикальное масштабирование. Больше данных означает использование большего сервера, который может быть очень дорогим. Также можно масштабировать на несколько серверов, но это тяжелый и занимающий много времени процесс.	Используется горизонтальное масшабирование, то есть между серверами. В основном, это недорогие серверы, или облачные решения, что делает масштабирование более дешевым по сравнению с вертикальным.
ACID соответствие (A-атомарность, C – согласованность, I – изолированность, D - надежность)	Большинство реляционных баз данных соответствуют требованиям ACID.	Большинство решений NoSQL жертвуют требованиям ACID для повышения масштабируемости и производительности.

Выбор технологии зависит от сценария использования. Если структура данных сильно меняется, необходима быстрая масштабируемость и высокая производительность, для того что бы справиться с быстрорастущим количеством данных, то необходимо выбирать NoSQL. Но если структура данных мало меняется и рост данных умеренный и управляемый, то лучше использовать SQL подход.

Литература. 1. John Roijackers http://alexandria.tue.nl/extra1/afstversl/wsk-i/roijackers2012.pdf

Рубець М.Г., Бідюк П.І.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Класифікація типів операційних ризиків та можливості їх математичного опису

Надається спрощена класифікація типів операційних ризиків та методів їх математичного опису з метою їх подальшого використання для оцінювання і прогнозування.

Операційні ризики у банківській сфері привертають все більшу увагу дослідників, оскільки регулятор постійно збільшує контроль у цій області, а самі ризики ще не отримали ґрунтовного математичного апарату для їх дослідження, наприклад, як кредитні або ринкові ризики. Тож у роботі надається класифікація типів операційних ризиків та існуючих підходів до їх моделювання та оцінювання для банківської сфери і пропонується модифікація комбінації існуючих методів.

Головні положення про операційні ризики для банківської сфери надає Базельський Комітет з питань банківського нагляду при Банку міжнародних розрахунків у документі Базель II [1,2]. Документ містить означення суті операційного ризику, а саме «операційний ризик – це ризик збитку в результаті неадекватних або помилкових внутрішніх процесів, дій співробітників і систем або зовнішніх подій», дає класифікацію подій операційного ризику (рис. 1), а також рекомендує підходи до оцінювання операційних ризиків (рис. 2).

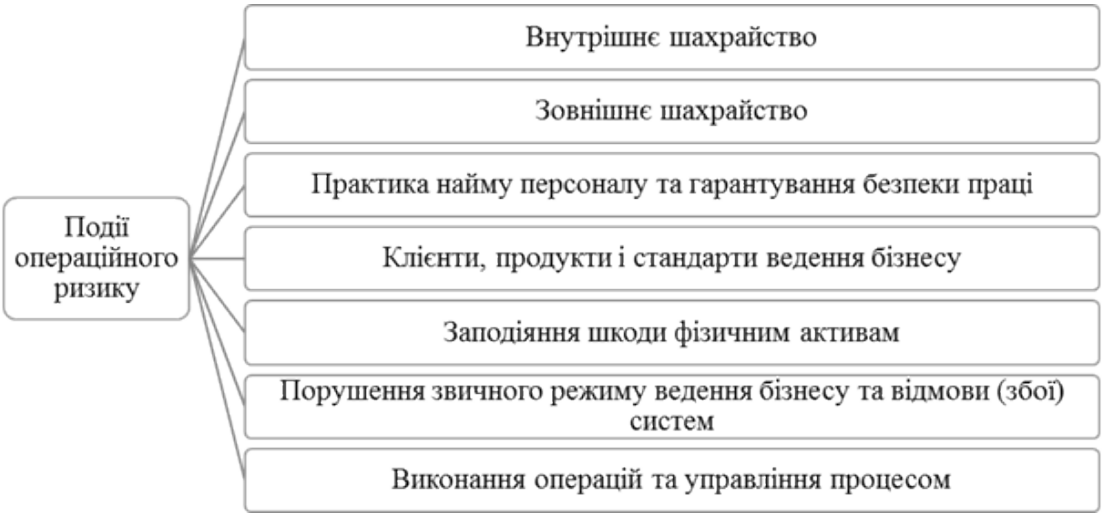

Рис. 1. Категорія подій операційного ризику згідно Базель II

Основну увагу дослідників привертають сучасні методи оцінювання операційних ризиків, оскільки вони дають розробляти унікальні системи для аналізу операційних ризиків для кожного банку [3,4]. Такі системи дають можливість скоротити капітал на покриття операційних ризиків за умов підтвердження своєї функціональності та адекватності вимогам регулятора [5].

Операційні ризики найчастіше пов'язані з людським фактором (шахрайство, помилки персоналу, розробників банківських систем тощо), тому вони не піддаються чіткому математичному опису, і в той же час кількість учасників діяльності у банківській сфері постійно зростає, що призводить до зростання частки операційних ризиків серед інших, більш досліджених, ризиків. Ці тенденції свідчать про високу актуальність досліджень у цьому напрямі, яка буде постійно зростати. У той же час, огляд існуючої літератури показав, що на території України увага операційним ризикам майже не надається.

До спрощених методів оцінювання операційних ризиків відносяться метод на основі базового індикатора (BIA) та стандартизований підхід (SIA). Методи поглибленого аналізу включають внутрішнє оцінювання (IMA), оцінювання кривої розподілів збитку (LDA), метод скорингових

Рис. 2. Класифікація методів оцінювання операційних ризиків

карт та сценарний аналіз. Кращі результати оцінювання досягаються, як правило, за допомогою комбінованого підходу, який грунтується на спільному використанні двох і більше взаємно доповнюючих методів.

У подальших дослідженнях будуть проаналізовані та використані існуючі сучасні методи оцінювання операційних ризиків, а також застосовано поєднання методів штучного інтелекту, а саме Байєсових мереж (БМ), нечіткої логіки, нелінійних регресійних моделей типу логіт і пробіт, дерев рішень, нейронних мереж різних структур з методами оцінювання кривої розподілу збитків та сценарним аналізом. Пропонується побудова БМ на основі поділу діяльності банків за бізнесовими лініями та поділом ризиків за причинами виникнення, котрі пропонує Базель II. Перевагами БМ є встановлення фактичних причинно-наслідкових зв'язків між процесами та можливість використання різних типів змінних в межах однієї багатовимірної моделі. Наявність множини методів формування ймовірнісного висновку дає можливість генерування альтернатив стосовно остаточного результату і вибору кращої з них за допомогою статистичних критеріїв якості.

Для дослідження та подальшого практичного використання проектується та реалізується програмно спеціалізована система підтримки прийняття рішень (СППР), яка забезпечує отримання високоякісного остаточного результату оцінювання завдяки системному підходу до проектування та реалізації. Відмінними рисами такої системи є ієрархічність її структури, функціональна повнота, адаптивність процедур оцінювання структури і параметрів моделей, а також контроль обчислювальних процесів на кожному етапі обробки статистичних даних, побудови математичних моделей та оцінювання прогнозів стосовно можливих втрат.

Література. **1.** Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework. [Електронний ресурс] Режим доступу: http://www.bis.org/publ/bcbs118.htm **2.** Фаррахов И.Т., Розанова Е.Ю. Оценка операционных рисков: возможны варианты. [Електронний ресурс] Режим доступу: http://inec.ru/it/reporting-institutions/brief-description-risk/publication/risk-ko-1-2-2014.php **3.** Operational Risk Regulation, Analysis and Management edited by Carol Alexander. – New York: Pearson Education Limited, 2003. – 369 p. **4.** Pavel V. Shevchenko, Modelling Operational Risk Using Bayesian Inference. – New York: Springer, 2011, 302 p. **5.** Yew Khuen Yoon, Modelling operational risk in financial institutions using bayesian networks. – London: University of London, 2003. – 83 p.

Савчук А.Ю.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Створення сучасної системи навігації з елементами доповненної реальності для мобільних платформ

Введення. У сучасному світі досить часто зустрічаються ситуації коли інформації у навколишньому середовищі не вистачає. Наприклад, проблема навігації, коли надані лише карти, що не завжди дають достатньо інформації для знаходження маршруту, а навіть для пошуку свого точного місцезнахождення. Доповнена реальність (AR) дозволяє створити новій рівень візуалізації простору навколо нас, доповнюючи реальний світ об'єктами віртуальної реальності. Така технологія інтегрована з сучасними системами навігації у мобільному додатку, дозволить доповнити реальне середовище, в режимі реального часу необхідними для навігації підказками, мітками.

Постановка задачі. Створення мультиплатформенного мобільного додатку, з інтеграцією системи побудови маршрутів, з можливістью додавання елементів доповненної реальності, у якості підказок в процесі навігації.

Інструменти вирішення. Використання ігрової платформи Unity 5, надає розробнику безкоштовний інструментарій для створення мультиплатформенних додатків, котрі однаково добре працюють на широкому спектрі пристроїв, не вимагаючи складання окремої версії для кожної апаратної і програмної платформи [1]. Важливим аспектом є те, що платформа Unity активно розвивається, реалізує інтеграцію з багатьма сервісами та плагінами, для вирішення широкого спектру проблем.

Для роботи із доповненою реальністю існує ряд готових рішень. Одним із таких є інструментарій для розробника доповненної реальності (AR SDK) Vuforia, який також надає плагін для Unity - Vuforia Unity Extension. Завдяки алгоритмам, розробленим компанією Qualcomm, Vuforia має змогу розпізнавати образи, які фіксує камера, і, базуючись на них, наповнювати реальність, яка відображається на екрані смартфона, елементами, які створює розробник програмного забезпечення[2].

Для побудови маршруту вибираємо карти Google, адже ця система постійно оновлюється, карти є максимально актуальними та мають максимальне покриття у порівнянні з конкурентами (Яндекс.Карти, 2ГИС), також доступна сучасна GPS-навігація по маршрутах міського транспорту з підказками для кожної пересадки. Для інтеграції з Unity необхідно завантажити плагін від Google Maps[3]. Координати місця положення смартфону знаходяться завдяки стандартним методам Unity.

Отже, всю роботу можна розбити на такі етапи:

1. Завантаження Unity.
2. Завантаження Vuforia та Vuforia Unity Extension.
3. Завантаження Google Maps Unity plugin.
4. Імпорт Vuforia Unity Extension та Google Maps Unity plugin в Unity.
5. Створення мобільного додатку.

Висновок. Використання сучасних платформ побудови мобільних мультиплатформенних програм, дозволяє створити навігаційний додаток з елементами доповненої реальності, що дозволить дивитися на навколишнє середовище через камеру мобільного пристрою і отримувати різноманітну корисну інформацію. Неможливо переоцінити значення цієї технології: це міст між середовищами онлайн і оффлайн, абсолютно новий рівень візуалізації простору.

Література. **1.** http://unity3d.com/ru/unity/multiplatform **2.** https://www.qualcomm.com/products/vuforia **3.** https://www.assetstore.unity3d.com/en/#!/content/3573

Сарапулов А.С.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Сравнение и анализ протоколов передачи мультимедиа в реальном времени

Целью данного исследования является выбор оптимального протокола для мультимедийной передачи информации.

Со стремительным развитием современного общества растёт запрос на скорость и качество передачи информации в реальном времени. В частности, мультимедийной информации.

Способы передачи мультимедиа. Стандарт ISO. Сегодня существует множество способов передачи мультимедиа. Среди них: потоковая передача (streaming), последовательная загрузка (progressive download), адаптивная трансляция (adaptive streaming). Они представляют общую идею: трансляции потока мультимедиа, с некоторыми отличиями. Последовательная передача потока предусматривает загрузку с возможностью использования информации до полного её получения. Адаптивная трансляция проводит коррекцию обьема передаваемого потока, на основе данных по скорости передачи данных клиенту. В то же время формат передачи данных определяют протоколы передачи данных.

OSI Model			
	Data unit	Layer	Function
Host layers	Data	7. Application	Network process to application
		6. Presentation	Data representation, encryption and decryption, convert machine dependent data to machine independent data
		5. Session	Interhost communication, managing sessions between applications
	Segments	4. Transport	End-to-end connections, reliability and flow control
Media layers	Packet/Datagram	3. Network	Path determination and logical addressing
	Frame	2. Data link	Physical addressing
	Bit	1. Physical	Media, signal and binary transmission

Рис. 1. Стек модели OSI

Специально для оптимального распределения функций передачи информации ISO был разработан стандарт: *модель OSI* (Open System Interconnection), которая определяет 7 логических уровней (слоев) для обмена информацией. Протоколы передачи мультимедиа занимают *приложенческий* (application) уровень модели OSI.

Стек доступных протоколов. С ростом важности передаваемой мультимедийной информации взамен известного HTTP, возникли альтернативные протоколы: RTSP, MMS, RTMP и т.п.

Протокол *Real-Time Transport Protocol* (RTP). Некоторые протоколы разработаны опираясь на конкретные транспортные протоколы. Протокол RTP и является их примером, построенный, как правило, на базе UDP. Разработан специально для передачи в реальном времени. Реализуемый на верхних слоях UDP (TCP), рассматривается, как часть транспортного уровня.

Протокол *Real Time Control Protocol* (RTCP) отражает возможности RTP, оперируя, в то же время, на уровне сессии. Основное свойство данного протокола: обеспечивать обратную информацию касательно качества передачи данных. Данная функция, впоследствии, дает возможность корректировки объёма передаваемых данных.

Протокол *Real Time Streaming Protocol* (RTSP) – это протокол передачи мультимедиа приложенческого уровня, который использует несколько протоколов в транспортном уровне для передачи своих пакетов, включая UDP и TCP. Очень тесно используется с RTP, но

Исследование выполнено в рамках работы над магистрской работы

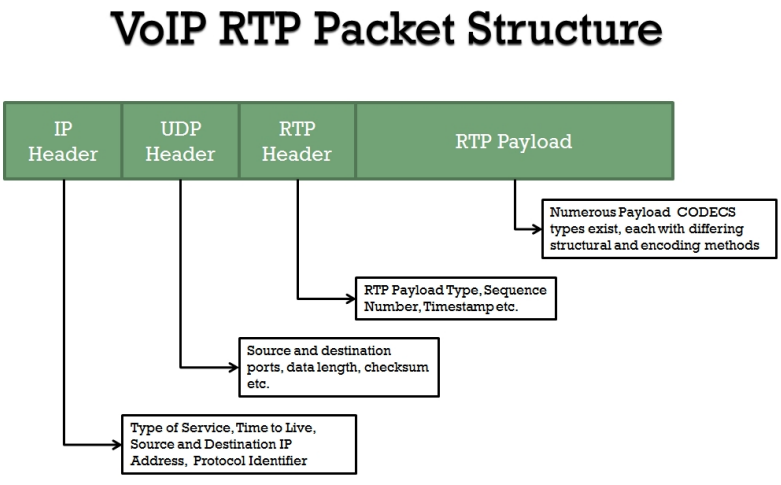

Рис. 2. Структура пакетов RTP

не связан с ним вплотную. Данный протокол разработан, как формат удаленного сетевого контроля. Чем-то напоминает HTTP: передает запрос для инициализации действий, таких как проиграть, пауза и запись.

Протокол *Resource Reservation Protocol* (RSVP) работает на транспортном уровне, хотя используется для установки сессий.

Сравнение протоколов. Стек протоколов RTP, RTCP и RTSP часто объединяют в RTSP. На верхнем уровне эти альтернативные протоколы объединяет несколько общих элементов: наличие сервера для передачи потока, которые работают в связке с традиционным HTTP сервером. В таких случаях роль HTTP сервера заключается в установке соединения между потоковым сервером и клиентами. Соединение существует до тех пор, пока клиент обрывает его. За счет этого, данные протоколы считаются стабильными в сравнении с HTTP, который, к тому же, не имеет связи между сервером и клиентом. Данный вопрос является одним из самых важных недостатков HTTP. К примеру, нет возможности обратиться к отдельному фрагменту мультимеди файла. Потоковые сервера более эффективно решают вопрос обслуживания множества пользователей.

Заключения. Исследование, на основе сравнения технических свойств, и теоретических показателей установило преимущество связки современных протоколов стека RTSP при решении таких, задач, как передача видео и аудио потоков в реальном времени. В тоже время, как и любая альтернатива, данное решение имеет как преимущества, так и недостатки в определенных аспектах относительно своих предшественников. В данном случае, эти недостатки могут выражаться, например, в надежности передачи информации и объёмов нагрузки на потоковые сервера. Поэтому, данное решение нельзя назвать однозначно преимущественным над стандартными протоколами, а производить выбор проткола для решения той или иной задачи необходимо исходя из неспосредственных локальных требований. В случае необходимости передачи мультимедия потока на максимальной скорости, при возможности в использовании мощных серверов и без особых требований к безопасности, даное решение можно считать оптимальным.

Литература. **1.** Tanenbaum A.S., Wetherall D.J, Computer Networks (5th Edition), 2010. **2.** Fall K.R., Stevens W.R., TCP/IP Illustrated, Volume 1: The Protocols (2nd Edition), 2011. **3.** Buford J., Heather Yu, Eng Keong Lua, P2P Networking and Applications, 2008.

Саржинский В.В., Дидковская М.В.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Методы машинного обучения в задаче идентификации диктора по голосу

Введение. В настоящие время, в связи с распространением компьютеризации, особо остро встаёт вопрос ограничения доступа пользователей. При этом всё чаще стандартные парольные системы защиты заменяются или дополняются биометрическими системами идентификации пользователей, которые имеют ряд преимуществ: уникальные человеческие характеристики трудно подделать, они не могут быть забыты или потеряны.

Цель работы. Изучение акустических характеристик речевых сигналов. Реализация алгоритма идентификации диктора по голосу при наличии посторонних шумов.

Основная часть. Для построения модели идентификации дикторов необходимо извлечь данные, характерные конкретному диктору. В зависимости от выбора тех или иных характеристик существенно зависит качество модели идентификации диктора [1-2].

С помощью микрофона записываются звуковые файлы в формате wav, далее звуковой сигнал разбивается на фреймы длины N с половинным перекрытием. При дальнейшей обработке для устранения нежелательных эффектов каждый элемент фрейма умножается на особую весовую функцию - окно Хэмминга. Далее применяется быстрое преобразование Фурье и логарифмируются амплитуды спектра. Поскольку звуковой сигнал записывается в шумной среде, необходимо максимально удалить шумы, и тем самым в минимальной степени повредить спектр речи, произнесённый диктором. Существуют несколько известных методов по удалению шумов: Noise Gate, слепая аддитивная фильтрации (LMS, RLS) и др. Для очистки звуковых данных от посторонних шумов на основе формантного анализа предложен собственный модифицированный алгоритм Noise Gate. Суть данного алгоритма состоит в обнулении значений амплитуд, чьи спектральные частоты не лежат в окрестности значений формант.

Для каждого фрейма получаем следующие характеристики: значения 4 первых формант, 14 мел-частотных кепстральных коэффициента MFCC, по 12 коэффициентов первых и вторых их производных (без краевых точек). Также в качестве ещё одной характеристики взято площадь, лежащую под амплитудой на спектре. Данное значение будет характеризовать усреднённую энергию звукового сигнала, соответствующее одному фрейму. Всего получили 43 характеристики для каждого фрейма, по которым идентифицируется тот или иной диктор. Таким образом, имеется небольшой набор значений, которые при распознавании успешно заменяют тысячи отсчетов речевого сигнала. На этапе постобработки признаков также применяются методы нормализации. Наиболее распространённым методом нормализации, предназначенным для снижения влияния канала, является метод вычитания кепстрального среднего.

После нормализации найденных характеристик применяются методы машинного обучения для идентификации диктора по голосу. В работе были применены следующие методы: метод опорных векторов SVM с подходом один-против-всех, Multinomial logistic regression и Random Forest, а также проанализировано качество их применения к задаче идентификации диктора. Наилучшие результаты были получены с использованием метода Random Forest.

Результат работы. В данной работе предложен и реализован алгоритм идентификации диктора по голосу в шумной среде с использованием методов машинного обучения. Проведены исследования, как мозг, получив информацию о характере изменения звукового давления во времени, извлекает информацию об идентификации диктора.

Литература. **1.** Рабинер Л.Р. Цифровая обработка речевых сигналов / Л. Рабинер, Р. Шафер. – Москва: Радио и связь, 1981. – 495 с. **2.** B. Gold, N. Morgan, Speech and Audio Signal Processing. New York: Wiley, 2000.

Семироз К.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Побудова системи рекомендацій за допомогою графової бази даних

Вступ. Найбільш поширеним в наш час рішенням для зберігання даних в системах різного типу та розміру є реляційні бази даних та системи для їх управління, наприклад, Oracle, MySQL тощо. Незважаючи на значний розвиток реляційної моделі даних та технологій побудованих на її основі, в даний момент виникає ряд завдань в галузі обробки інформації, при вирішенні яких розробник може стикнутися з труднощами безпосередньо написання коду (напр., побудова ієрархічних структур, робота з неструктурованими даними), а також з потребою значних ресурсів сервера, необхідних на обробку, [1]. Поява та поширення таких проблем сприяли розвитку NoSQL (Not only SQL) СКБД, які прийнято ділити на наступні типи [2]:

- Бази орієнтовані на поле (column-oriented store) – кожен блок даних складається з пар ключ-значення та має унікальний ідентифікатор, який також називають основним ключем. Прикладом такої СКБД є HBase (використовується у Facebook, Yahoo!);
- Бази ключ-значення (key-value store) – зберігають дані у вигляді асоціативного масиву. Приклад:Cassandra (використовується у Facebook);
- Бази документів (document database) – об'єднують різноманітні документи в єдину колекцію, дозволяючи індексувати їх не лише на основі ідентифікатора цілого документу, але й окремих його властивостей. Приклад: MongoDB (використовується у GitHub);
- Графові бази даних (graph database) – зберігають дані у вигляді орієнтованих графів з властивостями. Приклад: Neo4j (використовується у CrunchBase).

В даній роботі розглядається досить типова задача автоматизації: створити систему рекомендацій для інтернет-магазину одягу. Для вирішення завдання слід відповісти на наступні питання:

1. Які товари купував даний клієнт?
2. Які ще клієнти купували дані товари?
3. Що ще купували ці клієнти?

Реляційна база даних з легкістю дасть відповідь на перше питання, а також на друге (за умови невеликої кількості даних у базі). Для відповіді на третє питання необхідно створювати складні ієрархічні SQL-запити, які реалізуються шляхом побудови рекурсивних джойнів, що в свою чергу вимагають значних обчислювальних затрат.

Значно простіше отримати подібну інформацію, якщо дані зберігаються у вигляді орієнтованого графа. Для реалізації такої моделі даних було обрано СКБД Neo4j, оскільки вона є безкоштовною, легко інтегрується з сторонніми застосунками за допомогою REST API, має зручний веб-інтерфейс, а також в ній реалізовані нехарактерні для більшості NoSQL-баз даних механізми як підтримка транзакцій [3,4]. Реальним прикладом такої реалізації служить магазин eBay, де перехід з MySQL на Neo4j дозволив скоротити розмір коду запитів в 10-100 разів, а також значно пришвидшив час отримання результатів [5].

В рамках даної роботи БД інтернет-магазину в спрощеному вигляді (лише товари, покупці, замовлення) було реалізовано в MySQL та Neo4j. Дані були згенеровані за допомогою онлайн-ресурсу . Далі був створений відповідний програмний код, що для обраного клієнта, який зробив хоча б одну покупку, повинні бути визначені в якості рекомендації 5 товарів, які:

- Клієнт не купував раніше;
- Купували інші клієнти, які також купували й товар(и) куплені заданим клієнтом.

Література. **1.** I. Robinson, J. Webber, E. Eifrem. Graph Databases. Sebastopol: O'Reilly Media, Inc., 2013. **2.** S. Tiwari. Professional NoSQL. Indianapolis: John Wiley & Sons, Inc., 2011. **3.** G. Vaish. Getting Started with NoSQL. Birmingham: Packt Publishing, 2013. **4.** R. Van Bruggen. Learning Neo4j. Birmingham: Packt Publishing, 2014. **5.** Powering Recommendations with a Graph Database [Електронний ресурс]. – Режим доступу : http://neo4j.com/use-cases/recommendations/.

Сливка Т.П.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Розпізнавання нот при акапельному співі

Цифрова обробка людського голосу є однією з основних частин у нереалізованих перспективних напрямках розвитку Інформаційних технологій, у майбутньому. На даний момент уже існує певна множина алгоритмів, що дозволяють чинити різні маніпуляції над аудіо-сигналами. Безумовно, важливим являється розробка нових та модернізація уже відомих метод. Але, попри все, дуже цікавим та практичним є, саме, застосування, раніше уже відомих алгоритмів у абсолютно інших, нових сферах науки.

Основною задачею даної роботи являється розробка та налагодження програмного продукту згідно до наступних вимог. Програма повинна визначати основну (з максимальним звучанням) ноту, в кожний момент часу, при акапельному виконанні музичного твору користувачем. В зручному вигляді візуалізувати отримані дані (будувати частотно-часовий графік). А також надавати ряд допоміжних функцій, таких як: побудова "ідеального" графіку з файлу нотного формату для можливості візуального порівняння та корекції процесу співу; збереження та візуалізація статистичних даних; ведення персонального щоденника для користувача; тренувальні вправи різних рівнів важкості.

На даний час існує низка продуктів, що в тій чи іншій мірі вирішують поставлену задачу. Програма для тренування у відтворенні звуків певної частоти в ігровому форматі - «Aprendendo a Cantar» [1]. Програма для візуалізації голосу співака та визначення музичної ноти, що звучить - «Sing & See», основні недоліки: продукт платний, не реалізоване порівняння з "ідеалом" (нотами певного музичного твору) [2]. Програмний продукт для візуалізації, редагування та налаштування записаного вокалу - «Celemony», недоліки: дороговизна, відсутність можливості в онлайн режимі спостерігати за візуалізованим вхідним сигналом [3].

Після аналізу ринку програмних продуктів бачимо, що вони не задовольняють умови поставленої задачі, тому беремось за розробку персонального продукту, який в першу чергу виконуватиме пряме - основне, призначення: розпізнання нот при акапельному співі, а також володітиме наступними перевагами: безкоштовність, крос-платформенність, легкий у використанні інтерфейс, розвиваючі уроки для початківців. Забезпечити програмний продукт широким спектром налаштувань, що дозволять краще адаптувати його під конкретного користувача та особисті параметри його голосу. Можливість зберігати, порівнювати досягнуті результати.

Основним шляхом реалізації обираємо використання середовища Qt та мови програмування C++, що забезпечить крос-платформенність, здатність запускати на наступних операційних системах: Android, iOS, Windows, Linux. Для реалізації алгоритму "Швидкого Перетворення Фур'є" використати, теж крос-платформенну бібліотеку fftw3. Також слід зауважити: обравши Qt, як середовище розробки, ми отримуємо широкий вибір графічних елементів для побудови GUI, що в свою чергу значно полегшить та пришвидшить процес розробки та налагодження.

Серед етапів розробки можна виділити декілька основних:

1) Розробка прототипу;
2) Налагодження;
3) Доповнення функціоналу;

Запропонований, вище розглянутий, програмний продукт дозволить, з легкістю та задоволенням, пересічному користувачу, без особливих пізнань у музиці та IT здобути або покращити свої навички при акапельному співі. Цільова аудиторія надзвичайно широка, починаючи від любителів-аматорів і закінчуючи професіоналами.

Література. **1.** Aprendendo a Cantar, https://play.google.com/store/apps/details?id=com.projeto.learnsing. **2.** Sing&See, http://www.singandsee.com/get-singandsee.php. **3.** Celemony, http://www.celemony.com/en/start.

Смага М.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Система передачі даних з використанням матричних штрихкодів

З кожним роком все більшого розповсюдження набувають системи доповнення реальності, які застосовуються в сфері реклами, ігровій індустрії, системах забезпечення безпеки, системах керування транспортними засобами та ін.

Технологія під назвою «доповнена реальність», яку ще називають «покращена реальність» або «збагачена реальність» - це візуальний додаток до реального світу, який створюється шляхом проектування і введення будь-яких віртуальних, уявних об'єктів на даний простір (на екран комп'ютера, телефон, планшет і подібні пристрої) [1].

В даний момент сфери використання доповненої реальності ще не до кінця зрозумілі і оцінені споживачами, тому що технологія тільки входить у життя людей.

Для певних задач доповнення реальності необхідно не лише виявляти об'єкти на зображеннях (мітки), на які будуть накладатися віртуальні об'єкти, але й отримувати певну інформацію про зміну станів віртуальних об'єктів, що дозволить зробити їх динамічними, тобто з'явиться можливість маніпулювати їх поведінкою в режимі реального часу.

Завдяки наявності фото та відеокамер в сучасних мобільних пристроях та достатній якості отриманих зображень та швидкості їх обробки, в задачах доповнення реальності з'явилась можливість не лише шукати мітку для накладання віртуальних об'єктів а й отримувати додаткові дані, що будуть впливати на віртуальні об'єкти, декодуючи послідовності матричних штрихкодів, які можуть бути використані як мітки для побудови системи координат віртуальних об'єктів [2].

В роботі створено систему доповнення реальності на основі запропонованого алгоритму закодовування інформації, яка необхідна для впливу на віртуальні об'єкти, побудови та поступового виведення послідовності матричних штрихкодів на екран пристрою-передавача даних, а також алгоритму пошуку та декодування штрихкодів на зображенні з камери пристрою-отримувача, який зможе об'єднати всі отримані декодовані дані в блок даних ідентичний тому який передавався.

Представлена система має суттєву практичну цінність - можливість передавати додаткові дані в системи доповнення реальності без використання додаткових засобів передачі інформації (Bluetooth, Wi-Fi та ін.), вона дає можливість впливати на поведінку (просторову позицію, напрямок, кадр анімації, що програється) динамічних віртуальних об'єктів, а також, синхронізувати поведінку віртуальних об'єктів між багатьма спостерігачами системи доповнення реальності.

На основі створеної в роботі системи доповнення реальності, в якій буде використовуватись запропонований алгоритм передачі даних, можливо розробити наступні рішення:

- системи презентацій, де в якості слайдів будуть динамічні сцени, побудовані із віртуальних об'єктів;
- систему трансляцій спортивних подій;
- в ігровій індустрії запропоновану систему можна використати як для розробки нових жанрів, так і для трансляцій подій вже існуючих у новому вигляді;
- таку систему можна застосувати в сфері реклами, наприклад, в якості вуличної реклами;

Література. **1.** en.wikipedia.org/wiki/Augmented_reality. **2.** en.wikipedia.org/wiki/QR_code

Смаковський Д.С., Ляшенко М.М.
Національний технічний університет України "КПІ", теплоенергетичний факультет, Київ, Україна

Система пошуку кандидатів на вакансії роботодавців

У XXI сторіччі комп'ютери стали невід'ємною частиною життя кожної людини та вже зайняли почесне місце у багатьох галузях життя. Багато заводів і великі компанії впроваджують у себе комп'ютерну техніку і створюють інформаційну інфраструктуру. Але для кожної частини цього інформаційного процесу потрібні висококваліфіковані кадри. Тому на навчання персоналу витрачаються великі кошти. Проте в деяких випадках дешевше виходить найняти людину вже з потрібними знаннями. На даний момент поміж постачальників хороших кадрів на IT-ринку можна розглянути два напрямки:

1. Рекрутингове агентство. Можуть знайти кадри на вакансії будь-якої складності.
2. Інтернет-сервіси. Пошук IT-спеціалістів полегшується у тому плані, що кожен з них знаходиться у веб-просторі, а, отже, кожному з них можна написати у соціальній мережі чи на пошту.

При побудові роботи з агентством треба розуміти, що процес пошуку потребуватиме деяку кількість часу і агентства беруть дві річні зарплати спеціаліста, якого шукають. У інтернет-сервісів цієї проблеми немає, але у них є інша проблема. Основним важелем у прийнятті рішення між сервісами є швидкість оновлення інформації про спеціалістів. А отже перед розробником постає питання як зібрати потрібну інформацію про спеціалістів у режимі реального часу. Потрібна інформація складається з кваліфікації співробітника, його досвіду, фінансове становища і інших речей, які впливають на самого спеціаліста та його рішення у виборі місця роботи. У такій системі можна виділити основні модулі:

1. модуль вводу інформації – відповідає за надання можливості користувачу системи викласти матеріал, верифікує матеріал, слідкує за цілісністю даної інформації відповідно всього матеріалу сервісу;
2. модуль списків – виводить матеріал користувачу у порядку, який підходить для даного користувача або є стандартним для цього списку, відповідає за усі списки в системі, так як список вакансій, список резюме, списки користувачів;
3. модуль збереження інформації – відповідає за цілісність інформації у джерелах її збереження, верифікує поступаючу інформацію на предмет цілісності на рівні байтів, представляю поступаючи інформацію у вигляді, який може бути представлений у базі, займається операціями переводу інформації з одного представлення до іншого, зберігає інформацію про зв'язок з усіма джерелами інформації;
4. модуль виводу інформації – показує усю отриману від інших модулів інформацію користувачам системи у потрібному користувачеві або стандартному вигляді.

У двох останніх модулях використовується механізм парсингу резюме – представлення файлу резюме у вигляді заданої структури інформації. Без даного модуля розробникам системи потрібно буде зберігати файл резюме, у випадку використання даного механізму можна буде покласти лише потрібну інформацію. Модуль виводу переглядає, яку інформацію потрібно вивести користувачу, що дає значну гнучкість у питаннях отримання кінцевим користувачем лише інформації, яка цікавить.

Перелік посилань.

Література. 1. С.М. Іванова. Мистецтво підбору персоналу. /С.М. Іванова. - М.: Юрайт, 2012. – 674 с.

Сутула О.В.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Побудова архітектури Android додатку на основі патернів проектування

Стандартні Java патерни проектування не доцільно використовувати при розробці Android додатку через те, що Android архітектура має тісну зв'язаність між активністю, візуалізацією та доступом до даних, що невід'ємно порушує інкапсуляцію класів та спричиняє небажаний зв'язок між ними. При створенні Android додатку також виникають проблеми тестування, тому що такі засоби як Unit тести недоступні, а отже відлагодження програми, при відсутності гарної архітектури, відбувається переважно на реальному пристрої, що значно гальмує процес.[3]

У роботі проектується архітектура Android додатку для відділення логіки від реалізації, тим самим спрощуючи тестування та підтримку проекту, вводячи можливість модульного тестування.

Для вирішення даних проблем можна розділити додаток на три шари, кожен із яких буде ефективно використовувати особливості фреймворків для вирішення конкретних завдань поставлених перед додатком, інкапсулюючи реалізацію та не створюючи небажаний зв'язок між собою.

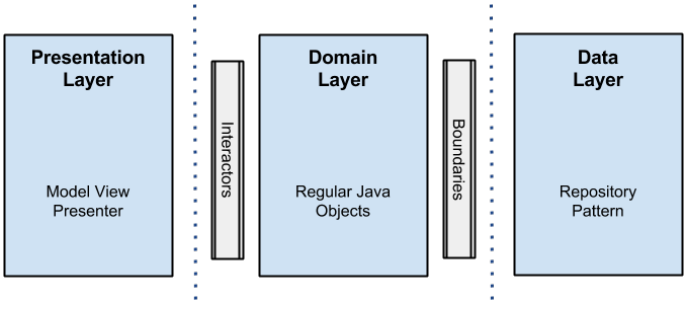

Рис. 1. Схема архітектури додатку

Перший шар - Presentation Layer, що відповідає за інтерфейс та базову логіку, які не можуть функціонувати без доступу до активності. На цьому шарі резонно використовувати патерн MVP (Model View Presenter). Цей патерн дозволяє нам розділити відображення від роботи інтерфейсу, а саме зручно розділити наші views від джерел інформації. За допомогою нього ми можемо відділити майже всю логіку від активності, що дозволить нам тестувати додаток навіть на цьому рівні без використання реального пристрою (емулятора).

Другий шар - Domain Layer. Це шар бізнес логіки що відповідає за обробку всіх видів взаємодії користувача й програми. Має складатися лише з java класів без зв'язку з Android. Цей шар є ядром додатку, яке не має сильних зв'язків з інтерфейсом що дозволяє додатку бути більш гнучким, дає можливість повторного використання коду.

Третій шар - Data Layer. Реалізується патерном Repository і дозволяє нам отримувати дані через єдиний інтерфейс, абстрагуючись від реалізації доступу до кожного джерела інформації (пам'ять, диск, хмара).[1]

Таким чином, даний розподіл функціональності по шарам, дозволяє нам більш гнучко тестувати додаток, застосовувати прийоми модульного тестування, збільшує абстрактність коду, гнучкість, а також дозволяє виокремити всю бізнес логіку й вирішує проблему повторного використання коду на різних платформах.

Література. **1.** Architecting Android... The clean way? http://fernandocejas.com/2014/09/03/architecting-android-the-clean-way/. **2.** Android Architecture, http://www.eazytutz.com/android/android-architecture/. **3.** MVP for Android: how to organize the presentation layer, http://antonioleiva.com/mvp-android/.

Тавров Д.Ю.
Національний технічний університет України "КПІ", Київ, Україна

Схема інформаційної технології забезпечення групової анонімності даних

Сучасні інформаційні технології дозволяють аналізувати великі обсяги первинних неагрегованих даних, оприлюднення яких може становити загрозу розкриття чутливої інформації. Задачі захисту чутливої інформації про індивідуальних респондентів розглядають у межах напряму схоронення приватності публічних даних (privacy-preserving data publishing, [1]). Не менш важливими є задачі захисту чутливої інформації про групи осіб. Наприклад, екстремальні значення територіального розподілу військовослужбовців, отриманого на основі аналізу публічних даних, можуть указувати на місцерозташування військових баз. Задачі такого роду розглядають у межах напряму забезпечення групової анонімності даних (group anonymity, [2]).

Як правило, дані, групову анонімність яких потрібно забезпечити перед оприлюдненням, можна представити у вигляді *мікрофайлу* $\mathbf{M}$ — файлу з записами про респондентів, кожен елемент яких містить значення атрибутів. *Групу* респондентів, розподіл інформації про яку потрібно захистити, визначають сукупність *сутнісних атрибутів* (значення яких дозволяють однозначно визначити факт належності кожного респондента групі) та *параметризуючий атрибут* (за значеннями якого потрібно здійснити розподіл інформації про групу). Розподіл інформації про групу називатимемо її *цільовим представленням* (ЦП). *Задача забезпечення групової анонімності* (ЗЗГА) полягає в модифікації мікрофайлу $\mathbf{M}$ у такий спосіб, щоб замаскувати чутливі особливості ЦП. Розв'язати ЗЗГА можна за допомогою двох підходів:

- вилучити один або декілька сутнісних атрибутів із мікрофайлу перед його оприлюдненням; у цьому випадку збудувати ЦП буде неможливо;
- виконати перерозподіл ЦП шляхом модифікації мікрофайлу [3]; у цьому випадку чутливі особливості ЦП, побудованого на основі модифікованого мікрофайлу, буде масковано.

Ризик порушення групової анонімності залишається [3], навіть якщо вилучити з мікрофайлу сутнісні атрибути. На основі додаткових даних чи експертних міркувань можна збудувати *наближені моделі* груп, аналіз ЦП яких може призвести до порушення групової анонімності.

Таким чином, розв'язання ЗЗГА на практиці потребує створення нової інформаційної технології, під час розроблення якої потрібно враховувати дві принципово відмінні ситуації:

- для мікрофайлу $\mathbf{M}$ можна підібрати один або більше близьких за структурою атрибутів *допоміжних мікрофайлів*, які містять *допоміжні сутнісні атрибути*, що мають інтерпретацію, аналогічну сутнісним атрибутам в $\mathbf{M}$; на основі допоміжного мікрофайлу можна збудувати систему нечіткого виведення (СНВ) для визначення ступенів належності респондентів групі; із допомогою такої СНВ можна збудувати *допоміжне цільове представлення* (ДЦП) групи, на основі якого можна порушити групову анонімність;
- для мікрофайлу $\mathbf{M}$ неможливо підібрати допоміжний мікрофайл; у цьому випадку відповідну СНВ можна збудувати, використовуючи тільки експертні міркування та дані з інших доступних джерел; із допомогою такої СНВ можна збудувати *узагальнене цільове представлення* (УЦП) групи, на основі якого можна порушити групову анонімність.

У процесі забезпечення групової анонімності потрібно спочатку проаналізувати, чи існують загрози порушення групової анонімності з використанням допоміжного мікрофайлу чи експертних міркувань, а потім безпосередньо розв'язати ЗЗГА для визначеної групи:

- якщо для забезпечення групової анонімності ухвалено рішення вилучити принаймні один сутнісний атрибут, модифікацію мікрофайлу достатньо проводити для ДЦП (УЦП);
- якщо ухвалено рішення не вилучати сутнісних атрибутів, потрібно проводити *сумісну* модифікацію мікрофайлу як для ЦП, так і для ДЦП (УЦП).

На рисунку 1 представлено UML-діаграму активності (activity diagram) користувача під час забезпечення групової анонімності даних відповідно до визначених вище міркувань.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

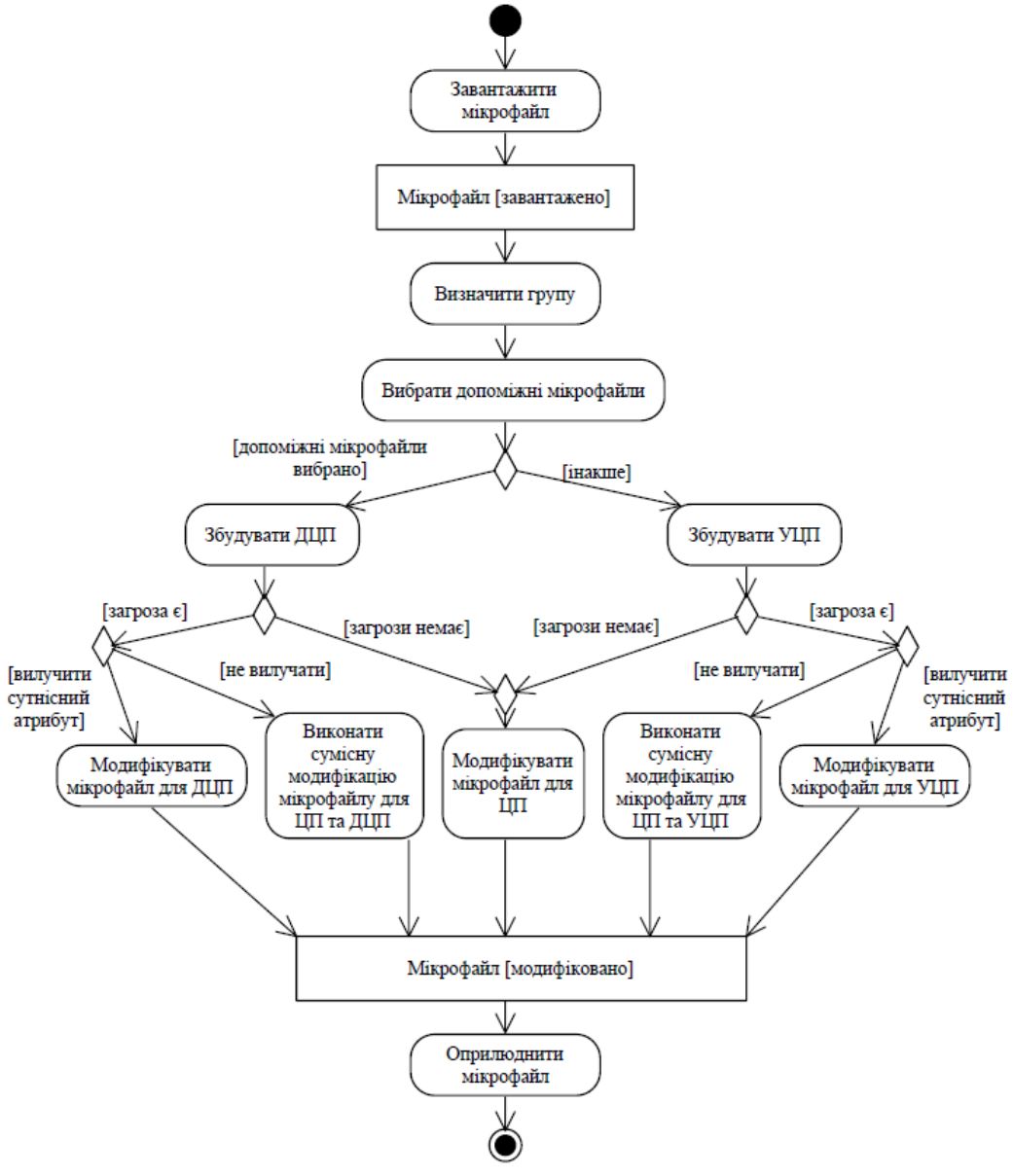

Рис. 1. UML-діаграма активності користувача під час забезпечення групової анонімності

Література. **1.** Fung B. Privacy-Preserving Data Publishing: A Survey on Recent Developments / B. Fung, K. Wang, R. Chen, P. Yu // ACM Computing Surveys. — 2010. — Vol. 42(4). — P. 1–53. **2.** Chertov O. Data Group Anonymity: General Approach / O. Chertov, D. Tavrov // International Journal of Computer Science and Information Security. — 2010. — Vol. 8(7). — P. 1–8. **3.** Chertov O. Microfiles as a Potential Source of Confidential Information Leakage / O. Chertov, D. Tavrov // Intelligent Methods for Cyber Warfare [ed. R. R. Yager, M. Z. Reformat, N. Alajlan]. — Springer International Publishing Switzerland, 2015. — P. 87–114. — (Studies in Computational Intelligence, vol. 563).

Тимошенко Ю.А., Романко А.Р.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Системный анализ и персональная медицина. Brain reading.

Вступление. Развитие современной цивилизации стремительно и бесповоротно движется в сторону ускорения и качественной оптимизации взаимодействия между человеком и машиной. Современные *информационно-коммуникационные технологии (ИКТ)* становятся повсеместными. От средств, которые позволяют решать научно-технические задачи, обрабатывать информацию в самых различных формах и проявлениях, они перешли в сферу незаменимых помощников человека. Смартфоны с широким спектром программ, бытовые устройства, технологии "smart house" - компьютер окружает человека, делая его жизнь лучше.

Однако в начале 21 века все большее внимание привлекает такая сфера применения электроники, как здоровье человека. Здоровье – немаловажный экономический фактор. Когда человек болен – он выпадает из привычного русла жизни, производства, научной и творческой деятельности. Поэтому сохранение его здоровья и качества жизни – важная проблема, которая широко исследуется в последнее время.

Современные технологии для каждого. Благодаря ИКТ медицина становится персонализированной. При помощи различных приложений стало возможным подсказывать человеку, как рационально и правильно спланировать свой день, контролировать свою дневную активность. Одни приложения измеряют калории пищи, которую мы употребляем, другие меряют пульс, третьи снимают электрокардиограмму и определенным образом анализируют информацию и дают нам советы или указания. Это направление позволяет человеку, не посещая врача, контролировать свое состояние здоровья. С точки зрения системного анализа, это направление относится к той области, которая в широких кругах описывается словом "reading".

Например *"heart reading"* - направление исследований в телемедицине, позволяющее контролировать работу сердца, находясь в любой точке земного шара и на любом расстоянии от медицинского учреждения. Мобильные устройства, называемые "холтеровские мониторы ЭКГ" позволяют считывать сердечные ритмы, а специальное программное обеспечение - расшифровывать данные и диагностировать заболевания.

Чтение мозга-новое направление в ИКТ. Более сложным объектом исследования, требующим системного подхода, является мозг человека. Проблеме исследования мозговой деятельности посвящено множество трудов. Огромной препоной на пути исследований была неточность, громоздкость и дороговизна устройств, позволяющих регистрировать активность головного мозга. Современные устройства позволили улучшить и удешевить процесс считывания мозговой активности. Необходимость точной расшифровки сигналов мозга на основе одной только электрической активности нейронов повлекла за собой цепь исследований и разработок специальных приборов (стационарных – магнитно-резонансные томографы, электро-энцефалографы(*EEG*), спектроскопы ближнего инфракрасного диапазона (*NIRS*); мобильных – нейроимпланты), а также аналитических, программных и математических методов преобразования и расшифровки сигналов.

Совокупность методов распознавания, идентификации и реконструкции сигналов нейронных клеток носит название *"Brain reading"* или декодирование мозга. Процесс считывания мозга на современном этапе состоит в: нахождении принципов кодирования мыслительных процессов посредством электрических импульсов и получения сигнала из мозга (input); расшифровки информации и разделения сигналов, приходящих с различный областей головного мозга и накладывающихся друг на друга вследствие сложных связей между нейронами, а также проблем передачи сигналов в мозг (output) .

Нейрокомпьютерные интерфейсы. Одно из перспективных направлений исследований в области взаимодействия человека и машины – нейрокомпьютерные интерфейсы, которые позволяют отслеживать деятельность головного мозга и передавать данные напрямую в компьютер, для последующего анализа мозговой активности. Результаты исследований выводят

персональную медицину на качественно новый уровень, позволяя бороться со всевозможными болезнями (лечение депрессии, эпилепсии, болезни Паркинсона, глухоты, слепоты) или же заменяют часть человеческой деятельности (протезы, роботы). Создаются тренажеры, использующиеся при лечении широкого спектра заболеваний, не только физических, но психологических (депрессии, синдром дефицита внимания).

В докладе описаны значимые результаты исследований в следующих областях:

- Лечение глухоты: первый успешный нейроимплант, полностью распознается 80 процентов предложений, пациент даже может разговаривать по телефону, слушать музыку. Лидер индустрии - Cochlear Americas.
- Лечение слепоты: визуальные импланты трех классов: стимулирующие сетчатку, стимулирующие зрительный нерв и стимулирующие непосредственно зрительную кору мозга. Передовой производитель и разработчик – немецкая компания Retinal Implants.
- Deep brain stimulation(*DBS*): импланты которые используются для лечения и контроля периферийных нервов, функциональной электрической стимуляции,если по какой-то причине нервы отказали (например при повреждениях C5-C6 позвонков очень трудно пользоваться кистями рук), стимуляция вагус нерва - лечение от эпилепсии, депрессии, обжорства. Лечение хронических болей путем нейростимуляции, лечение удушья во время сна. В настоящий момент есть 190,000 пациентов с имплантированной DBS системой Activa производства Medtronic.
- Мозго-машинный интерфейс (*Brain-computer Interface, BCI*). Исследования, влючающие вживление електродов (*braingate*) для взаимодействия с определенными участками нейронов. Неинвазивное считывание информации, с целью создания интерфейса человек-машина (контроль курсора, интерфейс P300, интерфейс передачи информации из мозга одного человека в мозг другого человека посредством интернет соединения, регистрация ошибок совершаемых оператором, технология BMI от компании Honda, позволяющая человеку контролировать движения робота ASIMO).
- Распознавание картинок и мыслей: путём сканирования мозга и зрительной зоны коры несложно определить, на каком из нескольких изображении остановился взгляд человека. Самый последний эксперимент учёных включал просмотр двумя участниками двухчасового видео и последующую его «реконструкцию» с помощью сканирования мозга и работы с искусственной нейронной сетью. Другое перспективное направление – расшифровка планирования людьми будущего.
- Система *C3Vision*. Она сочетает сильные стороны компьютера и мозга для ускорения сортировки трудно формализуемых данных вроде фотографий или видео. При этом результат достигается быстрее, чем могли бы выполнить задачу компьютер или человек по отдельности.

Учреждения, занимающиеся исследованиями и развитием направления. Bernstein Center for Computational Neuroscience; University of Washington; Society for Neuroscience; Medtronic Inc.; Лаборатория нейрофизиологии и нейроинтерфейсов МГУ им. М.В Ломоносова; Department of Neuroinformatics CNS ATR(advanced communications research institute international); Department of Computational Brain Imaging CNS ATR; Honda Japan Co., Ltd.

Литература. **1.** Ганин И.П., Шишкин С.Л., Кочетова А.Г., Каплан А.Я. (2012). Интерфейс мозг-компьютер «на волне P300»: исследование эффекта номера стимулов в последовательности их предъявления. Физиология человека.Т. 38. № 2. С. 5—13. **2.** Collinger J. L. et al. High-performance neuroprosthetic control by an individual with tetraplegia // The Lancet. 2012. № 6736(12). P. 61816—61819. **3.** Kaplan A. Y. et al. Adapting the P300-based brain- computer interface for gaming: a review // IEEE Transactions on Computational Intelligence and AI in Games, 2013.. **4.** Медицинский факультет Северо-Западного университета (США). «New Brain-Machine Interface Could One Day Help Paralyzed Individuals». **5.** The Wadsworth Center Brain-Computer Interface System.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Тригубишин І.Л.
Національний технічний університет України "КПІ", Київ, Україна

Виявлення ботів у соцмережах з використанням закону Бенфорда

Бот – спеціальна програма, що виконує автоматично або за заданим розкладом які-небудь дії через ті ж інтерфейси, що й звичайний користувач. Зазвичай боти призначаються для виконання роботи, одноманітною і повторюваною, з максимально можливою швидкістю. Часто у соцмережах боти використовуються зі шкідливою метою, наприклад для завантаження інтернет-каналів потоком рекламної інформації або інформації, що використовується політичними силами для маніпулювання суспільною думкою. Тому одним з важливих завдань модераторів соцмереж є виявлення ботів і блокування їх сторінок.

Одним з методів виявлення ботів є використання закону Бенфорда. Закон Бенфорда або закон першої цифри — статистичний закон відповідно до якого перша цифра чисел із багатьох (але не всіх) типових джерел інформації у повсякденному житті вживається нерівномірно. Відповідно до цього закону, найчастіше першою цифрою чисел є одиниця, у той час як наступні цифри з'являються в найвищому розряді все рідше й рідше. Закон має місце тоді, коли логарифми множини чисел розподілені рівномірно, що наближено справджується для багатьох реальних показників. За законом Бенфорда ймовірність випадання кожної цифри можна визначити за формулою:

$$P(d) = log10(1 + 1/d) \tag{1}$$

Це означає , що цифра «1» стоїть на першій позиції у близько 30% чисел, у той час як «9» є першою в менш ніж одному з 21 випадку.[1]

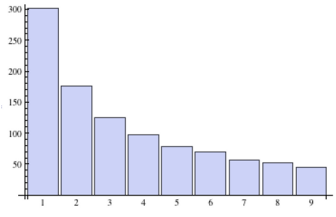

Рис. 1. Розподіл перших цифр відповідно до закону Бенфорда

Алгоритм виявлення ботів соцмереж за допомогою закона Бенфорда.

1. Необхідно зробити вибірку акаунтів осіб, серед яких буде проводитись дослідження. Це можуть бути випадкові користувачі, або члени певної спільноти, яку ми хочемо проаналізувати, тощо.
2. Потрібно вибрати параметри для дослідження, до яких відноситься параметр, за яким ми формуємо вибірку для кожного акаунта, змінна яка буде характеризувати об'єкт вибірки. Вибіркою може бути вибірка друзів, підписників або осіб, на які підписаний об'єкт досліджень. Змінною може бути кількість друзів, підписників, постів тощо. Множина значень цієї змінної для вибірки і буде множиною, що досліджується
3. Перевіряємо чи розподіл перших цифр досліджуваної множини задовольняє закону Бенфорда. Акаунти реальних осіб проходять перевірку і показують розподіл близький до розподілу на рисунку 1. Акаунти, які не задовольняють закону, підозрювані в тому, що вони боти. Їх необхідно додатково перевірити модераторам і заблокувати, якщо вони не підтвердять свою особу. [2]

Література. **1.** Adrien Jamain – Benford's Law, http://wwwf.imperial.ac.uk/~nadams/classificationgroup/Benfords-Law.pdf. **2.** Jennifer Golbeck Benford's Law Applies To Online Social Networks, http://arxiv.org/pdf/1504.04387v1.pdf

Цурін О.П., Слухай Я.О.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Web-сайти викладачів Університетів у світі

В сучасному суспільстві вища освіта розглядається як головний, провідний фактор соціального та економічного прогресу. Викладач є основною рушійною силою освіти. Можна виділити наступні напрямки його діяльності: викладацька, методична, наукова, виховна, організаційна. В залежності від університету та викладацького колективу кожна з них може бути проявленою у більшій чи меншій мірі. В час активного використання Інформаційних Технологій види діяльності залишаються, але більшість з них значно змінюється. Значне місце при цьому займають персональні сайти (блоги) викладачів. У них дві основні функції: інформаційна та комунікаційна. Розглянемо шляхи створення сайтів.

1. Сайт на власному хості. Цей сайт може буде унікальним та повністю відобразить особистість викладача. Але такий спосіб ведення персонального сайту вимагає спеціальних знань з html та веб-дизайну і часто йому треба володіти і веб-програмуванням. При цьому викладач стає повноцінним адміністратором сайту. Він має слідкувати, щоб код, який він використовує на сайті, був безпечний та вчасно оновлювати бібліотеки, які він використовує. Такий підхід вимагає певного професіоналізму, значної витрати часу та ретельної роботи при оновленні інформації, написанні новин, чи додаванні розділів на сайті. Оскільки часто викладач не є професіоналом в даній області, його сайт стає часто не достатньо захищений від вірусів, що може призвести до зараження та закриття сайту. Такий метод побудови сайтів використовується і у нас, і за кордоном. Практика ведення кожним викладачем персонального сайту на хості не є поширеною і використовується лише деякими окремими викладачами, яких є лише декілька осіб на цілий університет, що підтверджується відношенням кількості таких сайтів до загальної кількості на всіх проаналізованих ресурсах. В НТУУ "КПІ" є показовий приклад використання власного хоста http://oim.asu.kpi.ua.

2. Використання соціальних мереж. Викладачі можуть спробувати використовувати соціальні мережі, створивши там власну спільноту. Це досить зручне та просте рішення, яке можна втілити за кілька хвилин, не володіючи спеціальними навиками. Також такі спільноти зручні, тому що студент буде постійно переглядати новини, та одразу дізнаватись про зміни, так як новини викладачів будуть в одному місці з новинами соц. мережі. З боку захищеності такої спільноти використання соціальної мережі являється зручним. Вона завжди піклується про стабільність своєї роботи. При великій кількості відвідувачів соціальна мережа робить все, щоб не допустити зараження своїх серверів та користувачів вірусами і вводить певні обмеження. У більшості сучасних соціальних мереж неможливо завантажувати та пересилати файли програм, а у випадку програмування на javascript та php неможливо пересилати навіть програмний код, тому що він є інтерпретованим.

Соціальна мережа має дуже зручні засоби для спілкування та швидкого донесення інформації. Більшість студентів має встановлені на смартфон додатки від соціальних мереж і миттєво отримують повідомлення. Для тих, хто не користується смартфонами великі соціальні мережі відсилають безкоштовні Content Management System (CMC) повідомлення. В той же час викладачі не зможуть публікувати всі свої роботи та обмінюватись всією необхідною інформацією через обмеження соціальної мережі, тому що соціальна мережа відхилятиме будь-які файли, що вона вважатиме підозрілими.

Викладацькі сторінки у соціальних мережах погано індексуються пошуковими мережами. Використання соціальних мереж не є повноцінною заміною системи блоґів, а є скоріше її доповненням чи тимчасовим рішенням на період розробки системи блоґів. Більшість проаналізованих систем блоґів мають сторінки із посиланнями на профілі соціальних мереж.

3. Використання хмарних інструментальних засобів(на прикладі Google Sites). Google Sites надає всі можливості для створення персонального сайту. Вони надають 100ГБ місця для створення архіву, що дає можливість викласти не тільки всі свої публікації, а можна вносити

кожну лекцію, записану на відео, для студентів, які були відсутні на лекції чи займаються дистанційно [1]. В інтерфейсі Google Sites викладач зможе легко орієнтуватись. По суті йому не доведеться створювати персональний блоґ заново, тут вже все створено та підтримується, користувачеві залишається лише зареєструватись та наповнювати свій блоґ, не турбуючись про безпеку даних. Але ця система не досконала. Так, зручність та надійність у використанні Google Sites можлива лише після виділення університету аккаунту. Для його виділення необхідно пройти тривалі бюрократичні процедури, для того, щоб підтвердити в офісі Google той факт, що університет є дійсно університетом.

Користувач стає повністю залежним від цього функціоналу і буде змушений постійно дотримуватись його правил та умов користування сервісом Google. Більше того, система працюватиме рівно стільки, скільки існуватиме дана платформа. Google не гарантує, що він не закриє доступ до сервісу в найближчий час. Так, наприклад, раніше сервіси Google Apps були доступні усім бажаючим, але з меншою кількістю дискового простору. Нещодавно Google закрив нові реєстрації, а отже, відтепер невідомо - зробить він подальшу підтримку створених раніше аккаунтів платною чи ні.

Приклад: http://blog.stanford.edu/ - Google Sites

4. Створення викладацьких блоґів на базі широковживаної CMS. Використання CMS на власному сервері дозволяє отримати повний контроль над проектом. Сучасні широковживані CMS мають потужний функціонал для публікації контенту: зображень, документів, програм тощо. Використання такої CMS значно підвищує надійність проекту, тому що CMS тестується мільйонами користувачів та кваліфікованих тестерів. Якщо знаходиться проблема в процесі роботи, вона оперативно виправляється та публікується оновлення для CMS, щоб адміністратор сайту міг прийняти рішення щодо її критичності.

Приклади успішної реалізації:
http://blogs.bu.edu – WordPress – Blogs for the Boston University Community
http://blogs.kent.ac.uk/ - University of Kent – Wordpress - Великобританія
http://cad.edu.kpi.ua – кафедра СП ІПСА НТУУ “КПІ” – Wordpress - Україна

5. Створення викладацьких блоґів на базі спеціалізованих CMS. Якщо розглянути спеціалізовану CMS, то можна вважати, що це найбільш вдалий варіант. Одразу отримується повний контроль над проектом: можна керувати наповненням, правами доступу, системою контролю версій, управляти потоком документів. CMS дозволяє пошуковим системам швидко проводити індексацію файлів, таким чином полегшувати їх пошук. За рахунок того, що дана CMS спеціалізована, в ній є засоби для доопрацювання та розширення системи в разі потреби. Цей програмний продукт повинен надавати можливості спілкування на сайті. Як і для всіх спеціалізованих CMS, проблеми наступні: вартість такої системи і стає актуальною відповідність спеціалізованої CMS потребам Університету.

Подібні системи використовуються у провідних закордонних університетах. Вдалі приклади реалізації:

- http://www.ox.ac.uk/blogs.html - University of Oxford - спеціалізована CMS
- http://staffblogs.le.ac.uk/ - University of Leicester Staff Blogs – спеціалізована CMS на базі Wordpress
- http://blogs.brad.ac.uk/ - University of BrandFord – модифікована Wordpress

Література. **1.** https://support.google.com/sites/answer/96770?hl=ru

Шевченко Я.Т.
Институт прикладного системного анализа НТУУ "КПИ", Киев, Украина

Нагрузочное тестирование web-приложений

В настоящее время большее число комерческих и государственных компаний используют web-приложения как для взаимодествия с конечными пользователями, так и для внутренних потребностей. Практически каждая компания и практически каждый пользователь зависит от надежности web-приложений, функциональные возможности которых непрерывано растут. Возрастает сложность web-приложений, но вместе с тем и нагрузка на них.

Проблемы с производительностью приводят к отказу клиентов от использования приложений. Web-приложение должно быть доступно и выполнять свои задачи в любой момент времени, когда оно понадобится клиенту. Низкую производительность приложения замечает большинство пользователей, что негативно влияет на такой показатель удобства использования как "мера пользовательской реакции". В конечном итоге пользователи предпочитают более производительное программное обеспечение.

Низкая производительность уменьшает количество полезных операций (действия, направленные на удовлетворение потребности пользователя или бизнес-потребности владельца приложения) в единицу времени, что приносит убытки владельцу приложения. Низкая производительность может стать причиной выхода приложения из строя, возникновения проблем с безопасностью и иных нежелательных последствий и т.п.

Таким образом, тестирование производительности позволяет обнаружить и устранить множество потенциальных проблем, которые приложение может начать испытывать на стадии эксплуатации.В связи с этим актуальной является задача проведения качественного нагрузочного тестирования, которое обеспечит стабильность работы приложения.

Нагрузочное тестирование или Тестирование производительности - это автоматизированное тестирование, имитирующее работу определенного количества бизнес пользователей на какомлибо общем (разделяемом ими) ресурсе[1].

Основными целями нагрузочного тестирования являются[2]:

- Оценка производительности и работоспособности приложения на этапе разработки и передачи в эксплуатацию
- Оценка производительности и работоспособности приложения на этапе выпуска новых релизов, патч-сетов
- Оптимизация производительности приложения, включая настройки серверов и оптимизацию кода
- Подбор соответствующей для данного приложения аппаратной (программной платформы) и конфигурации сервера

Существующие системы нагрузочного тестирования либо требуют глубоких знаний и опыта для их использования(отсутствие GUI, сложная терминология), либо очень затратны в аппаратном обеспечении или позволяют тестировать только отдельные web-сервисы или загрузку отдельных документов(страниц), что не позволяет судить о возможностях системы по отношению к реальным пользователям.[3]

Предлагается система нагрузочного тестрования web-приложений на основе открытого программного продукта Apache Jmeter, позволяющая оценивать производительность web-приложений, находить «узкие места» в приложении и собирать необходимую информацию для выработки рекомендации по улучшению производительности.

Литература. **1.** Автоматизация тестирования производительности: основные положения и области применения[Електронный ресурс]. – Режим доступа: http://svyatoslav.biz/technologies/performance_testing **2.** Нагрузочное тестирование Електронный ресурс]. – Режим доступа: http://www.intuit.ru/studies/courses/611/467/lecture/20768 **3.** Обзор программ нагрузочного тестирования веб-серверов[Електронный ресурс]. – Режим доступа: https://xakep.ru/2008/04/21/43327/

Юдін Є.М.
Физико-технический институт НТУУ "КПИ", Киев, Украина

Застосування методу навчання з підкріпленням для керування вивільненням опертативної пам'яті

Задача керування вивільненням пам'яті, що також називається задачею збирання сміття, полягає у своєчасному вивільнені оперативної пам'яті від об'єктів, які не будуть використовуватися програмою у подальшому. Навчання з підкріпленням – це розділ штучного інтелекту, що дозволяє реалізувати заздалегідь не визначені алгоритми, використовуючи формалізм взаємодії агенту і середовища. На відміну від підходу навчання з вчителем, таке навчання реалізується не через прямі вказівки щодо дій, а за допомогою системи нагород і штрафів [1].

Постановка проблеми. Комп'ютерні програми використовують заздалегідь невідому кількість обчислювальних ресурсів, і одним з способів запобігти розробки прибиральника сміття для кожної програми окремо є створення універсального прибиральника, що може виробляти алгоритм на основі власного досвіду. Наприклад, у сучасній віртуальній машині HotSpot/JRockit мови Java, проблему вибору стратегії прибиральника вирішено завдяки наявності декількох різних алгоритмів, жоден з яких не використовує навчання з підкріпленням [2]. Таким чином, мета дослідження – запропонувати алгоритм керування прибиранням сміття, що використовує навчання з підкріпленням і що зменшить частоту викликів прибиральника та дозволить адаптуватися до поведінки обслуговуваних прикладних програм.

Методіка дослідження. Основними математичними поняттями навчання з підкріпленням є стан, дія, стратегія, функція користі та функція користі пари стан-дія [3].У формалізмі навчання з підкріпленням система, яка називається агентом, взаємодіє з середовищем, виконуючи деякі дії і отримуючи деяку винагороду. У загальному випадку процес навчання можна описати як на рис.1 [4] Стратегія – це відображення $\pi : S \times A \to [0;1]$, тобто ймовірність виконати агентом певну дію a, коли середовище знаходиться у певному стані s. Результатом повинна бути максимізація певної сумарної винагороди

Опис моделі та алгоритму. Створений алгоритм включає в себе сутності, що взаємодіють: оперативна пам'ять, що є середою навчання, мутатор та прибиральник сміття, що є агентом навчання. У моделі пам'ять – це масив байтів довжиною L. У якості вектору стану обрано двокомпонентний ветор, перша компонента s_1 якого – кількість зайнятих байтів, а друга s_2 – збільшення зайнятої пам'яті з минулого кроку. Мутатор – програмна сутність, що виконує запити на створення об'єктів у пам'яті як деякий пуассонівський потік з дискретним часом та інтенсивністю λ. Кожен об'єкт O_i має випадкові розмір l_i та час життя t_i з рівномірним розподілом. Для агенту нього визначено дві дії: виконання прибирання a_0 та очікування a_1. На кожному такті прибиральник за функцією користі стану пам'яті $Q = Q(a_i, s_j)$ визначає ймовірність збирання і після виконання дії отримує штраф. Ймовірність виконання дії визначається однозначно за формулою $\pi(s_t, a_1) = sigm(Q(a_1, s_t)) = \frac{1}{1-exp(-Q(a_1,s_t))}$. Більшу ймовірність мають дії з більшою користю, проте для інших дій ймовірність ненульова. У якості сигналу від середовища обрано саме штрафи, тобто від'ємнозначні числа: $r_{(1)} = -1$ за збирання та $r_2 = -3$ за нестачу пам'яті і виконання екстреного збирання. Прибиральник оновлює функцію корисності за принципом Actor-Critic. Спершу формується TD-похибка $\delta_t = r_{t+1} + \gamma\, V(s_{t+1}) - V(s_t)$. Далі поновлюється функція користі стану $V'(s_{t+1}) = V(s_t) + \alpha\delta_t$ та функція користі пари стан-дія $Q'(a_1, s_t) = Q(a_1, s_t) + \beta\delta_t$, що і використовується для визначення стратегії. Відповідно на кожному такті відбувається генерування та вкионання запиту мутатора, дії та навчання прибиральника.

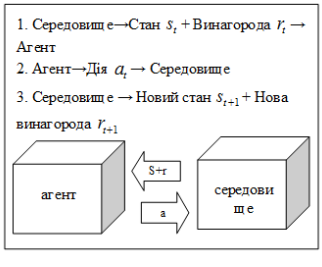

Рис. 1. Ілюстрація до моделі навчання з підкріпленням.

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Моделювання та аналіз отриманих результатів. У якості основного параметру ефективності взято співвідношення кількості виконаних збирань до кількості тактів на певний момент. Було перевірено, що для великої кількості тактів середній час, витрачений на збирання, збігається до деякої оптимальної величини. Було виявлено і продемонстровано на рис.2, що у загальному випадку процес навчання виконується у п'ять етапів: після ініціалізації функції користі відбувається початкове навчання і параметр спадає; далі після початкового встановлення функції користі параметр швидко росте через велику долю неправильних дій; на третьому етапі формується первинна функція користі і параметр спадає; потім відбувається досить тривалий період спроб та помилок із налаштуванням функції користі; наприкінці відбувається тонке налаштування функції і параметр збігається до деякого значення.

Також для порівняння було реалізовано класичний алгоритм, що полягає у прибиранні у випадку помилки виділенням пам'яті. Доцільно було дослідити кількість екстрених прибирань, оскільки така ситуація не завжди є рівноцінною звичайному прибиранню – воно може займати більше часу або потенційно викликати помилки. У такому випадку частота збирань, до якої збігається наш алгоритм, більша за частоту простого алгоритму у приблизно 1.5 рази, адже така частота мінімальна з усіх можливих. Проте доля екстрених прибирань для алгоритму з навчанням швидко падає до 1% від всіх збирань, або близько 1.5% від результату класичного алгоритму(рис.3) Якщо екстрене збирання вважати у $w = \frac{r_3}{r_1} = 3$ рази менш бажаним за звичайне, то алгоритм з навчанням виграє приблизно у 2.5 рази.

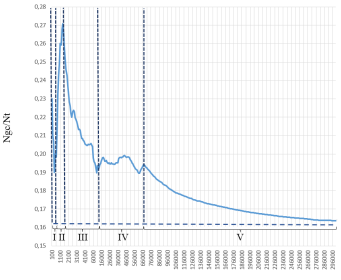

Рис. 2. Залежність частоти викликів прибиральника від часу у загальному випадку

Було досліджено поведінку алгоритмів у випадку зміни поведінки мутатора. На рис.4 зображено графік реакції на подвоєння інтенсивності виникнення об'єктів. Для класичного алгоритму зростають лише кількості екстрених викликів. Алгоритм з навчанням без інформації від користувача реагує на зміни і починає перерахунок функцій цінності. У результаті частота викликів для нього збігається до величини у приблизно 1.15 меншої ніж для класичного алгоритму.

Висновки. Програма демонструє працездатність алгоритму – для великої кількості тактів частота викликів прибиральника збігається до деякої оптимальної з точки зору моделі навчання величини. Було підтверджено адаптивність алгоритму до змін поведінки мутатора. Запропонований алгоритм дозволив майже повністю уникнути екстрених викликів прибиральника (частота екстрених викликів складає приблизно $1,5\%$ від частоти для класичного алгоритма), проте для даної моделі керування пам'яттю він не виявився кращим за простий алгоритм, що є за теорією оптимальним. У рамках ситуації, коли екстрене збирання у $w = 3$ менш бажане за звичайне, алгоритм з навчанням дозволив знизити частоту викликів приблизно у 1.5 рази.

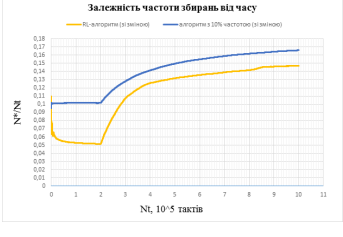

Рис. 3. Графік частоти від часу при зміні інтенсивності

Література. **1.** R. S. Sutton, A. G. Barto Reinforcement Learning: An Introduction — MIT Press, Cambridge, MA, 1998 A Bradford Book **2.** Java Documentation Java Platform, Standard Edition HotSpot Virtual Machine Garbage Collection Tuning Guide **3.** D.Andre Programmable Reinforcement Learning Agents. — Stanford University, 1994 **4.** E.Andreasson Reinforcement Learning for a Dynamic JavaTM Virtual Machine — Royal Institute of Technology Stockholm, 2002

Яковлева Г.О.
Інститут прикладного системного аналізу НТУУ "КПІ", Київ, Україна

Автоматизація документообігу в роботі Compliance менеджера

Робота Compliance менеджера полягає у контролі за дотриманням правил, законів, норм та принципів роботи компанії та ведення бізнесу. Саме поняття "Compliance" з певного часу закріпилося у бізнес-лексиконі як обов'язкова складова підприємницької діяльності та управління підприємством.

Впровадження системи електронного документообігу дозволяє уникнути багатьох помилок, що виникають при виконанні різних операцій з документами, а також значно прискорити дані процеси, що веде до економії робочого часу співробітників, а отже, до підвищення загальної ефективності роботи підприємства, що надає йому певних конкурентних переваг на ринку.

Метою виконання даного проекту є створення масштабованої, доступної й надійної СЕД для налагодження системи документообігу в IT компанії, що спростить роботу Compliance менеджера з документами, підвищить продуктивність праці співробітників за рахунок скорочення часу створення, обробки та пошуку документів.

Документообіг складається з потоків різних типів документів, що поділяються на 3 основні групи: вхідні, вихідні та внутрішні. На даний час документообіг здійснюється шляхом передачі документів за допомогою електронної пошти та передачі документів безпосередньо із рук у руки.

У зв'язку з цим виникає необхідність автоматизувати документообіг. Створювана система документообігу повинна виконувати такі задачі: упорядкувати і автоматизувати внутрішній організаційно-розпорядничий та договірний документообіг; оптимізувати бізнес-процеси, пов'язані з документообігом; створити електронний архів файлів будь-яких форматів; забезпечити неможливість втрати документів; забезпечити контроль версій; організувати спільну роботу над документами; забезпечити необхідну безпеку і розмежування доступу [1].

Проектована СЕД повинна відповідати трьом принципам: масштабованості, доступності та зручності [2]. У разі успішного розгортання системи можна прогнозувати швидке збільшення кількості користувачів та об'єму даних. Таким чином, система має бути добре масштабована, щоб мати можливість за мінімальних витрат швидко та багатократно збільшувати свою продуктивність, і разом з цим вона має бути зручна у використанні та максимально економна.

Проект матиме трирівневу структуру. Перший рівень буде відповідати за фізичне розміщення даних репозиторію на сервері. Другий рівень – сервер управління контентом, котрий складається з безпосередньо репозиторію контенту і набору сервісів, через які користувач може отримувати доступ до контенту. Третій рівень - клієнти, підключені до сервера управління контентом через різні протоколи та інтерфейси API.

Для розробки даного проекту обрано Alfresco Community Edition 4.2 - вільне середовище управління контентом підприємства ECM (Enterprise Content Management), вихідні тексти якої поширюються в рамках ліцензії LGPL [3].

В результаті проведених досліджень, планується розробити СЕД з перспективою впровадження її в роботу Compliance менеджера в IT компанії. В процесі розробки прототипу автоматизованої системи планується налаштувати інтерфейс користувача, автоматизувати бізнес-процес узгодження документів, занести у програму організаційну структуру, налаштувати основні функції програми для роботи з документами. Впровадження автоматизованої системи дозволить значно знизити трудові і вартісні витрати, і тим самим забезпечити високу окупність проекту та досить швидке повернення капітальних вкладень.

Література. **1.** Матвієнко О. В. Основи організації електронного документообігу / О. В. Матвієнко, М. Н. Цивін // Навчальний посібник. – К.: Центр учбової літератури, 2008. – С. 112. **2.** Галузинський Г. П. Перспективні технологічні засоби оброблення інформації / Г.П. Галузинський // Навчально-методичний посібник: Навчальне видання. – К.: КНЕУ, 2002. – С. 280. **3.** Alfresco Documentation: For Developers, http://docs.alfresco.com/community/index.jsp.

Authors · Авторы · Автори

Bakhrushin Vladimir E., 22
Belevets Marina S., 208
Bidyuk Petro I., 45, 133, 139, 208, 215
Bilushchak Yurii I., 23
Boldak Andriy O., 210
Bondarenko Alexey S., 128
Chaplya Yevhen Ya., 25
Chereda Hryhorii O., 27
Chernukha Olha Yu., 23, 25
Cojocaru Svetlana, 29, 37
Davydok Anastasiia Ye., 25
Fridman Ron, 126
Gaindric Constantin, 29
Gera Bohdan V., 23
Goncharuk Volodymyr Ye., 23
Ilchenko Kseniia O., 30
Ivanchuk Mariya A., 31
Ivanov Serhii S., 44
Ivchenko Dmytro A., 128
Kachko Nikita A., 212
Kazachenko Olga D., 213
Khurtsilava Konstantin V., 214
Kopychko Sergii M., 130
Korshunova Iryna A., 130
Koval Oleksandr V., 32
Kyrysha Bogdan A., 213
Litvinov Valeriy A., 214
Maistrenko Svetlana Ya., 214
Makukha Mykhailo P., 33
Malinovskiy Artyom D., 131
Malyk Igor V., 31
Mazur Vladislav G., 128
Nikolaiev Sergii S., 27, 131
Oparina Elena L., 33
Ovi Nafas Aghaei Agh Ghamish, 137
Pabyrivskyi Viktor V., 25
Paliichuk Liliia S., 34
Pankratova Nataliya D., 13, 35
Parkhomchuk Dmytro M., 36
Petic Mircea, 37
Piddubna Yuliya I., 215
Popcova Olga, 29
Ptuha Yulia A., 39
Pyshnograiev Ivan O., 30
Pyvovar Vitaliy V., 35
Sagan Vitaliy Yu., 126
Savastiyanov Volodymyr V., 41
Savchenko Illia O., 43
Schuka Roman V., 44
Secrieru Iulian, 29
Shcherban Mariia M., 45
Slashcheva Tetyana G., 39, 126
Slyota Maxim R., 39
Sokolov Pavlo Ye., 133
Tarasiuk Taras S., 39
Tkachenko Dmytro A., 135
Tripolsky Oleksandr V., 47
Tymoshenko Yuriy O., 36, 126
Volkova Kateryna A., 32
Yaremenko Vadym S., 217
Yefremov Kostiantyn V., 210
Zaychenko Yurii P., 137
Zyma Dmytro S., 139
Акимов Вадим Сергійович, 140
Акінфієва Анастасія Олегівна, 218
Амосов Олексій Віталійович, 141
Артюхов Віталій Григорович, 219
Архангельский Кирилл Евгеньевич, 221
Арчвадзе Натела Нодариевна, 222
Бабенко Олександр Анатолійович, 184
Базалий Оксана Юриевна, 142
Байас Марсія Марісоль, 238
Баклан Валерия Сергеевна, 143
Барсученко Владимир Витальевич, 144
Безносик Олександр Юрійович, 170
Белошицкая Оксана Константиновна, 48
Бидюк Петр Иванович, 142
Бідюк Петро Іванович, 146, 150, 234, 279

17-th International conference on System Analysis and Information Technology SAIT 2015
Institute for Applied System Analysis, National Technical University of Ukraine "KPI", Kyiv, Ukraine, June 22–25, 2015
ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Білова Марія Олексіївна, 74
Білорус Дмитро Сергійович, 226
Боровська Таїса Миколаївна, 180
Бояринова Юлия Евгеньевна, 50
Британов Олексій Анатолійович, 219
Булах Богдан Вікторович, 185, 227
Булах Олексій Романович, 173
Валькман Юрий Роландович, 15, 114
Васильев Владимир Иванович, 51
Вергун Олександр Сергійович, 186
Виклюк Ярослав Ігорович, 147
Винокуров Дмитро Сергійович, 53
Вишталь Дмитрий Михайлович, 51
Віжевський Петро Володимирович, 54
Візнюк Олена Василівна, 56
Вініченко Андрій Миколайович, 148
Галатенко Дмитро Володимирович, 228
Гіоргізова-Гай Вікторія Шалвівна, 184
Гоблик Анна Вікторівна, 57
Горбань Ірина Миколаївна, 59
Горбань Наталія Володимирівна, 60
Горелова Галина Викторовна, 61
Гоцко Уляна Ігорівна, 149
Грайворонська Анастасія Миколаївна, 62
Громова Юлія Андріївна, 64
Даниляк Оксана Сергіївна, 150
Дембовский Олег Юрьевич, 152
Джунковський Валерій Олександрович, 229
Джунковський Юрій Олександрович, 229
Дидковская Марина Витальевна, 284
Діденко Дмитро Георгійович, 231, 232
Дідик Світлана Анатолівна, 234
Довгаль Костянтин Ігорович, 66
Донченко Володимир Степанович, 68
Драган Алексей Владимирович, 236
Дрозденко Александр Николаевич, 249
Дубовой Володимир Михайлович, 238, 240
Дуда Марина Ернестівна, 238
Душин Олександр Сергійович, 242
Ємельянова Дар'я Ігорівна, 76
Забелин Станислав Игоревич, 243
Заец Роман Васильевич, 69
Зайченко Юрій Петрович, 172
Згуровский Михаил Захарович, 18
Зражевська Наталія Григорівна, 71
Ігнатченко Лука Олександрович, 187
Калиновский Яков Александрович, 50
Капшук Олег Олексійович, 188, 190
Карпенко Дмитро Ігорович, 154
Кирюша Богдан Анатолиевич, 192
Киселев Геннадий Дмитриевич, 193
Киселева Анна Геннадиевна, 193
Кислий Роман Володимирович, 244
Кірік Олена Євстафіївна, 72
Клунько Дмитрий Александрович, 155
Ковальчук-Химюк Людмила Олександрівна, 245, 268
Ковтун Олексій Олексійович, 194
Козуля Марія Михайлівна, 117
Козуля Тетяна Володимирівна, 74, 76
Кондіус Костянтин Юрійович, 248
Кондратенко Наталія Романівна, 157
Кондратова Людмила Павловна, 95
Копичко Сергій Миколайович, 158
Копычко Сергей Николаевич, 249
Коровай Ольга Володимирівна, 158
Королевський Герман Дмитрович, 250
Котенко Дмитро Андрійович, 251
Кравець Ірина Олександрівна, 159
Крамар Олександр Володимирович, 185
Круш Ігор Володимирович, 252
Куліш Златослава Ігорівна, 254
Кулык Роман Юрьевич, 249
Курдуп Іван Олександрович, 121
Курмак Игорь Сергеевич, 78
Кущ Сергей Николаевич, 273
Лавренюк Алла Миколаївна, 195
Лавренюк Сергій Іванович, 195
Лазарович Олег Мирославович, 79
Ланде Дмитро Володимирович, 62, 91
Лашко Олена Вікторівна, 227
Леко Тарас Богданович, 147
Лисецкий Юрий Михайлович, 80
Ліщук Катерина Ігорівна, 160
Лопатин Алексей Константинович, 82
Лугова Ольга Юріївна, 190
Любашенко Наталья Дмитриевна, 51
Ляхевич Сергій Володимирович, 161
Ляшенко Максим Максимович, 288
Мазурик Алексей Юрьевич, 255
Майкут Сергій Олексійович, 269
Малишев Андрій Ігорович, 257, 258
Маматов Владимир Игоревич, 259
Маркова Наталія Миколаївна, 84
Маслянко Павло Павлович, 84, 86
Медведський Андрій Миколайович, 260

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Мельничук Микола Михайлович, 262
Милявский Юрий Леонидович, 105
Мирошникова Ірина Юріївна, 263
Михалько Віталій Геннадійович, 264
Морозов Константин Вячеславович, 106
Надточий Александра Викторовна, 88
Назарага Інна Михайлівна, 68
Найдьонов Іван Михайлович, 89
Нарижна Ніна Сергіївна, 266
Науменко Тетяна Олександрівна, 267
Негода Катерина Володимирівна, 268
Недашковская Надежда Ивановна, 90
Олефір Олександр Степанович, 108, 162
Оріщенко Анастасія Олександрівна, 86
Павленко Анна Ігорівна, 163
Пазинич Ілля Сергійович, 91
Паничева Дарина Олеговна, 273
Панкратов Владимир Андреевич, 93
Панкратова Наталья Дмитриевна, 94
Папуч Тамара Юріївна, 165
Перевертайло Володимир Володимирович, 269
Петренко Анатолій Іванович, 196
Петренко Олексій Олексійович, 198
Печурин Николай Капитонович, 95
Печурин Сергей Николаевич, 95
Пикалов Александр Сергеевич, 199
Пилипенко Інна Віталіївна, 240
Пинчук Владислав Александрович, 94
Письменний Ігор Олександрович, 167
Поворознюк Анатолій Іванович, 168
Погорілий Сергій Дем'янович, 97
Попеляев Денис Павлович, 271
Попова Антоніна Вікторівна, 232
Попова Марина Валеріївна, 170
Потебня Артем Віталійович, 97
Прогонов Дмитрий Александрович, 273
Путренко Віктор Валентинович, 103
Пховелишвили Мераб Гайозович, 222
Рева Костянтин Володимирович, 275
Ревуцька Ірина Володимирівна, 201
Родіонов Андрій Миколайович, 275
Розпутний Михайло Васильович, 171
Романенко Виктор Демидович, 105
Романкевич Алексей Михайлович, 106
Романкевич Виталий Алексеевич, 106
Романко Алексей Ростиславович, 292
Романов Валерий Владимирович, 203
Романчук Катерина Геннадіївна, 107
Ротенберг Александр Валерьевич, 278
Рубець Микола Григорович, 279
Савін Євген Вадимови, 108
Савчук Артем Юрійович, 281
Савчук Олена Борисівна, 162
Сарапулов Александр Сергеевич, 282
Саржинский Василий Валериевич, 284
Семироз Катерина Володимирівна, 285
Сергеєв-Горчинський Олексій Олександрович, 140
Сидорук Інна Анатоліївна, 172
Сирота Сергій Вікторович, 242
Сливка Тарас Петрович, 286
Слухай Яна Олександрівна, 295
Смага Микита Вадимович, 287
Смаковський Денис Сергійович, 288
Снитюк Виталий Евгеньевич, 80
Снігур Ольга Олексіївна, 157
Станіна Ольга Дмитрівна, 116
Стефанишин Дмитро Володимирович, 107, 110, 111
Стефанишина-Гаврилюк Юлія Дмитрівна, 111
Стефанович Тетяна Олександрівна, 112
Стець Роман Сергійович, 240
Стецюк Роман Валерьевич, 114
Сутула Олександр Віталійович, 289
Тавров Данило Юрійович, 173, 245, 290
Тарасова Ольга Валеріївна, 68
Тимошенко Юрий Александрович, 292
Тимошенко Юрій Олександрович, 161
Ткаченко Владимир Александрович, 203
Тригубишин Ірина Любомирівна, 294
Туренко Алина Сергеевна, 50
Ус Світлана Альбертовна, 116
Устимчик Марина Володимирівна, 146
Филатова Анна Евгеньевна, 175
Фіногенов Олексій Дмитрович, 177
Харченко Костянтин Васильович, 205
Ходневич Ярослав Васильович, 110
Хоменко Ольга Володимирівна, 59
Цурін Олег Пилипович, 295
Чегренець Олександр Сергійович, 179
Чернацький Артем Юрійович, 269
Черненко Ольга Борисовна, 82
Чертов Олег Романович, 54, 66
Шаронова Наталья Валеріївна, 117

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf

Шевченко Ярослав Тарасович, 297
Шелест Антон Вікторович, 119
Шецирули Лия Дмитриевна, 222
Шишко Наталия Валериевна, 94
Штрик Руслана Валеріївна, 177
Шульган Інна Василівна, 180
Шумада Катерина Олегівна, 120
Щербовських Сергій Володимирович, 112
Щодро Олексій Євгенович, 110
Юдін Єгор Михайлович, 298
Юрченко Ігор Валерійович, 122
Яковлева Алла Петрівна, 121
Яковлева Ганна Олександрівна, 300
Ясинський Володимир Кирилович, 122

ISBN 978-966-2748-69-7, revision 1.0 (2015.05.11), available at http://sait.kpi.ua/books/sait2015.ebook.pdf