



LITHUANIAN COMPUTER SOCIETY
VILNIUS UNIVERSITY
INSTITUTE OF MATHEMATICS AND INFORMATICS
LITHUANIAN ACADEMY OF SCIENCES



8th International Workshop on
**DATA ANALYSIS METHODS FOR
SOFTWARE SYSTEMS**

Druskininkai, Lithuania, Hotel "Europa Royale"
<http://www.mii.lt/DAMSS>

December 1–3, 2016

Co-Chairmen:

Dr. Saulius Maskeliūnas (Lithuanian Computer Society)
Prof. Gintautas Dzemyda (Vilnius University, Lithuanian Academy of Sciences)

Programme Committee:

Prof. Janis Grundspenkis (Latvia)
Prof. Hele-Maja Haav (Estonia)
Prof. Albertas Čaplinskas (Lithuania)
Prof. Yuriy Kharin (Belarus)

Organizing Committee:

Dr. Jolita Bernatavičienė
Dr. Olga Kurasova
Aidas Žandaris

Contacts:

Dr. Jolita Bernatavičienė
jolita.bernatavicienne@mii.vu.lt
Dr. Olga Kurasova
olga.kurasova@mii.vu.lt
Tel. (8~5) 2109 312

Supported by:Main sponsors

Research Council of Lithuania
www.lmt.lt

Insoft
www.insoft.lt

BAIP
www.baip.lt

NRD CS
www.nrdcs.lt

Sponsors

Ernst & Young Baltic
www.ey.com

NetCode
www.netcode.lt

Baltic Amadeus
www.baltic-amadeus.lt

Computer Science Center
www.inscience.lt

DOI: <http://dx.doi.org/10.15388/DAMSS.2016>

ISBN 978-9986-680-61-1

© Vilnius University

Goals of the Event

1. Introduce computer science and IT business community with research undertaken at Lithuanian and foreign universities in the fields of software engineering and data mining (the number of scientists presenting their projects at the conference grows every year).
2. Foster relations between business entities and research community. Business representatives will have an opportunity to introduce their problems that need scientific solutions.
3. Facilitate scientific networking that will lead to joint research projects.
4. Initiate connections with foreign research institutions and scientists.

Topics

- Data Mining
- Software Engineering
- Visualization Methods of Multidimensional Data
- Medical Informatics
- Ontological Engineering
- Business Rules

Publications

We will recommend the authors of talks to submit their papers for publication in the special issue of *Baltic Journal of Modern Computing*, www.bjmc.lu.lv

Schedule of events

Thursday, December 1

11⁰⁰–13⁰⁰ – Registration/Info Desk Open

13⁰⁰–13¹⁵ – Opening. Welcome and Introductory Remarks

Gintautas Dzemyda “Data Analysis Methods for Software Systems” (Vilnius University, Institute of Mathematics and Informatics)

13¹⁵–14¹⁵ – Plenary Session

Anatoly Zhigljavsky “Some Geometrical and Analytical Features of Problems Involving Big Data and High Dimensions” (Cardiff University, UK)

Ignacy Kaliszewski “The Strength and Beauty of Optimization: the Case of Radiotherapy Planning in the Service of the Society” (Systems Research Institute of the Polish Academy of Sciences, Poland)

14¹⁵–15⁰⁰ – Coffee, Poster Session, and Discussions

1. *J. Zubova, O. Kurasova* “Multi-Level Method for Big Data Visualization”
2. *J. Mandravickaitė, T. Krilavičius* “Hybrid Approach for an Automatic Identification of Multi-Word Expressions for Latvian and Lithuanian”
3. *R. Užupytė, T. Krilavičius, R. Navickas, A. Šatrauskienė* “An Exploratory Analysis of the Relation Between Metabolic Syndrome Factors and MicroRNA Data”

4. *V. Mickevičius, T. Krilavičius, V. Morkevičius* “Visualization and Automatic Thematic Classification of Voting in the Lithuanian Parliament”
5. *V. Čyras, F. Lachmayer, K. Lapin* “Different Views to Law and Representation of Meanings”
6. *J. Venskus, M. Kurmis, P. Treigys* “Modified SOM for Abnormal Marine Traffic Detection”
7. *J. Tichonov, O. Kurasova, E. Filatovas* “Classification-Based Storage of JPEG Images”
8. *M. Sabaliauskas* “An Investigation of Surface Deformation Using Singular Value Decomposition”
9. *A. Valatavičius, S. Gudas* “Modeling Environment to Maintain Interoperability of Enterprise Applications”
10. *G. Tamulevičius, J. Kaukėnas, A. Rybakovas* “Evaluation of Speaker’s Individual Qualities Using High Order Autoregressive Model”
11. *Š. Germanas* “Two-Stage Variant Calling Method for Next-Generation Sequencing Experiments”
12. *M. J. Bilinskas, G. Dzemyda* “The Registration of Computed Tomography Scan Slices Using the Model of the Ribs-Bounded Contour”
13. *T. Petkus, K. Straukaitė, O. Kurasova* “Estimation of Teachers’ Demand in Municipalities”
14. *A. Rapečka, J. L. Jedzinskas, G. Dzemyda* “Detailed Analysis of the Dreams Meanings Dataset”
15. *M. Vaitonis, S. Masteika* “Computerized High Frequency Trading of Nanoseconds in Futures Market”
16. *M. Liutvinavičius, V. Sakalauskas, D. Krikščiūnienė* “Anomaly Detection in Financial Markets Using Investors Sentiment Indicator”
17. *L. Sakalauskas, D. Kavaliauskas* “Simulated Annealing Algorithm Optimization Method Convergence Study”

15⁰⁰–16⁰⁰ – Plenary Session

Enrique Herrera-Viedma “Intelligent Decision Making and Consensus” (University of Granada, Spain)

Inga Timofejeva, Roza Joffe, Alfonsas Vainoras, Minvydas Ragulskis “Identification of Clusters in a Trial Group of Persons Based on Their Cardiac Intervals and Magnetic Field Data” (Kaunas University of Technology)

16⁰⁰–17³⁰ – Coffee, Poster Session, and Discussions

1. **S. Konovalenko** “Methods and Means of the Customs Data Mining”
2. **A. Jakaitienė, D. Stumbrienė, R. Želvys** “Principal Component versus Data Envelopment Analysis in Construction a Composite Indicator for Education Monitoring”
3. **O. Pozdniakova, D. Mažeika** “Systematic Literature Review of the Cloud-Ready Software Architecture”
4. **J. J. Moreno, G. Ortega, E. Filatovas, J. A. Martínez, E. M. Garzón** “Evaluation of an Efficient NSGA-II Version on Heterogeneous Low-Power Platforms”
5. **G. Ortega, L. O. López, I. García, A. Plaza, E.M. Garzón** “Accelerating the LRASR Anomaly Detection Method on Heterogeneous Platforms”
6. **I. Drulytė, T. Ruzgas, R. Raišutis, S. Valiukevičienė, G. Linkevičiūtė** “Automatic Method for Assessment and Integration of Measurements Performed by Ultrasonic and Spectrophotometric Technologies in Order to Estimate the Lesion Parameters of the Human Tissue”
7. **R. Kriauzienė, A. Bugajev, R. Čiegis, T. Leonavičienė, J. Žilinskas, G. Jankevičiūtė** “Optimization of Efficient Absorbing Boundary Conditions for Schrödinger Equation”

8. **E. Filatovas**, *O. Kurasova*, *T. Petkus*, *A. Igumenov*, *S. Ivanikovas* “Efficient Utilization of Multicore Processors in PC Clusters”
9. **T. Sidekerskienė** “Time Series Modelling Using a Novel Method of Decomposition”
10. **N. Pozniak** “Application of Fractional Euclidean Distance Matrices to Extrapolation of Scattered Data”
11. **A. Gimbutas**, *A. Žilinskas* “Remarks on a Multi-Criteria Simplicial Optimization with an Estimate of Lipschitz Constant”
12. *O. Vasilecas*, **T. Vysockis**, *T. Rusinaitė* “Goal-Oriented Dynamic Business Process Simulation: Thoughts on Implementation”
13. **L. Dovydaitis**, *V. Rudžionis* “Speaker Recognition Using Deep Neural Networks”
14. **G. Korvel** “Analysis and Synthesis of Lithuanian Consonants”
15. **R. Savukynas**, *V. Marcinkevičius*, *A. Čaplinskas* “A Review of Infrastructures of Internet of Things for Smart Environments”
16. **J. Vaičiulytė** “Recurrent Estimation of Homogeneous Hidden Markov Model Parameters”
17. *A. Čaplinskas*, **J. Miliauskaitė** “On Algorithm Notion”

18⁰⁰–19³⁰ – Welcome Party

Friday, December 2

9⁰⁰–10⁰⁰ – Plenary Session

Florin Gheorghe Filip “Modern Information and Communication Technologies and Their Impact on Decision Support Systems” (Romanian Academy, Romania)

Gracia Ester Martín Garzón “Energy-Aware Scientific Computing” (University of Almeria, Spain)

10⁰⁰–11⁰⁰ – Parallel Plenary Sessions

Section 1

1. **Y. Kharin** “Discrete-Valued Time Series: Models and Statistical Inferences” (Belarusian State University)
2. **L. Sakalauskas** “Distance Matrices Geometry and Data Mining”
3. **J. Borzovs, L. Niedrite, D. Solodovnikova, A. Klavins** “Factors to Predict Attrition among First Year Computer Science Students” (University of Latvia)

Section 2

1. **I. Markievicz, M. Tamošiūnaitė, D. Vitkutė-Adžgauskienė, T. Krilavičius** “Text Mining for Robotic Action Ontology Engineering”
2. **D. Krikščiūnienė, V. Sakalauskas** “Prudent Monitoring of Quality and Cost Containment in Health Care Systems”
3. **V. Krinickij, L. Bukauskas, A. Brilingaitė** “Table-Top Exercise Gamification with Dynamic Scenario for Cyber Security Qualification Assessment”

11⁰⁰–11²⁰ – Coffee

11²⁰–12²⁰ – Parallel Plenary Sessions

Section 3

1. **J. Žilinskas**, *R. Paulavičius* “Global Optimization Using DIRECT Algorithm and Modifications”
2. *M. Viola*, *M. Sangiovanni*, *G. Toraldo*, **M. R. Guarracino** “Semi-supervised Learning with Generalized Eigenvalues” (ICAR-CNR, Italy)
3. **R. Paulavičius**, *C. S. Adjiman* “Global Optimization of Nonconvex Bilevel Problems: Implementation and Computational Study of the Branch-and-Sandwich Algorithm”

Section 4

1. **A. Laurinavičius** “Comprehensive Immunohistochemistry: Digital, Analytical and Integrated”
2. **J. Gordevičius**, *K. Koncevičius*, *P. Gibas*, *A. Kriščiūnas*, *T. Grigalis* “Cell-Free Blood Circulating DNA Methylation in Cancer”
3. **E. Vaičiukynas**, *A. Verikas*, *A. Gelžinis*, *M. Bačauskienė*, *K. Vaškevičius*, *V. Uloza*, *E. Padervinskis*, *J. Čičelienė* “Fusing Various Audio Feature Sets for Detection of Parkinson’s Disease from Sustained Voice and Speech”

Section 5

1. **D. Passey** “Learning, Data Analysis and Software Systems – Qualitative and Quantitative Dilemmas?” (Lancaster University, UK)
2. **M. Visnovitz** “Aspects of Choosing Textual Programming Languages for High School Education” (Eötvös Loránd University, Hungary)
3. **E. Sutinen** “Digital Theology: A Computer Scientist's View” (University of Turku, Finland)

12³⁰–14⁰⁰ – Lunch

14⁰⁰–15³⁰ – Plenary Session

Panos Papadopoulos “Commercial Analytics to Improve Business Performance through Micro Customer Segmentation, Consumer Behaviour Analysis, and Advanced Analytical Methods Application” (Ernst & Young, Greece)

James M. Calvin “The Bayesian Approach to Global Optimization” (New Jersey Institute of Technology, United States)

Alexander V. Tuzikov, Ivan A. Kashyn, Yuri V. Kornoushenko, Alexander M. Andrianov “Computer Screening and Modelling for Anti HIV-1 Drug Development” (National Academy of Sciences of Belarus, Institute of Bioorganic Chemistry, United Institute of Informatics Problems)

15³⁰–17³⁰– Coffee, Poster Session, and Discussions

1. ***S. Minkevičius, E. Greičius*** “An Application of Recurrent Method for the Analysis of Computer Network”
2. ***K. Paulauskienė, O. Kurasova*** “Massive Data Visualization via Selecting a Data Subset”
3. ***K. Zabkiewicz*** “Classification of Data Streams Using Nearest Neighbour Classifier and Normalized Compression Distance”
4. ***V. Jakštys, V. Marcinkevičius, P. Treigys*** “An Application of Radial Distortion Model for Chromatic Aberration Correction”
5. ***E. Monstavičius, K. Stasiukynaitė, V. Veikutis*** “The Effect of Body Vibration on the Physical Function of Humans”
6. ***R. Baronas, L. Petkevičius*** “Micro-reactor with an Outer Layer: Experimental Investigation and Modeling”

7. *I. Veitaitė, A. Lopata* “Knowledge-Based UML Dynamic Models Generation Method Usage in IS Lifecycle Design Stage”
8. *K. Lapin* “User Needs and Quality in Use: an Overview of User Satisfaction Models”
9. *V. Kaminskas, E. Ščiglinskas* “Control of Human Excitement as Reactions to a Dynamic Virtual 3D Face”
10. *M. Morkūnas, P. Treigys, A. Laurinavičius* “An Overview of Methods to Spatially Map Intra-tumor Genetic Heterogeneity in Whole Slide Pathology Images”
11. *R. Petrolis, A. Pilypas, G. Jurkevičiūtė, S. Lukoševičius, K. Petrikonis, A. Kriščiukaitis* “New Diagnostic Possibilities Using Multivariate Analysis of Multimodal Magnetic Resonance Images”
12. *E. Šabanovič, D. Matuzevičius, A. Serackis, V. V. Borutinskaitė, D. Navakauskas, R. Navakauskienė* “Investigation of Background Extraction Techniques to Novelty Estimation in Biomedical Imaging”
13. *K. Morkūnaitė, V. Veikutis, E. Sakavičiūtė* “IVUS-VH Relation to the Extent and Composition of Atherosclerotic Plaque and its Rupture Risk Prognosis”
14. *G. Gimbutienė, A. Žilinskas* “Investigation of the Effect of the Assumed Statistical Model in Global Optimization”
15. *T. Savickas, O. Vasilecas* “An Approach to Simulate Business Processes Using Probabilistic Process Models”
16. *V. Gurčinas, A. Čenys, N. Goranin* “Authorship Attribution Techniques for Short Instant Message Analysis”
17. *J. Jucevičius, P. Treigys, R. Briedienė, I. Naruševičiūtė, J. Bernatavičienė* “Comparative Analysis of Abnormal Prostate Region Detection Algorithms in mpMRI Scans”

18. **P. Čepulionis**, *R. Raišutis*, *S. Valiukevičienė*, *G. Linkevičiūtė* “Detection of Melanoma Skin Cancer by Classification of Deep Neural Network”
19. **G. Stabingis**, *J. Bernatavičienė*, *P. Treigys*, *A. Paunksnis*, *R. Vaičiaitienė*, *V. Medvedev* “Automated Blood Vessels Diameter Measurement in Eye Fundus Images”
20. **L. Paliulionienė**, *V. Paliulionis* “Computer-aided Analysis and Evaluation of Chess Helpmate Problems”

19⁰⁰–21⁰⁰ – Dinner

Saturday, December 3

9³⁰–11³⁰ – General Discussion

11³⁰–12⁰⁰ – Coffee, Poster Session, and Discussions

12¹⁵–14⁰⁰ – General Discussion, Concluding Remarks

14⁰⁰–14³⁰ – Closing

Micro-Reactor with an Outer Layer: Experimental Investigation and Modelling

R. Baronas, L. Petkevičius

Faculty of Mathematics and Informatics
Vilnius University
Linass.mif.vu@gmail.com

This paper presents a one-dimensional-in-space mathematical model of a micro-reactor model in a polar coordinates with an outer Nerst layer. The micro-reactor action was modelled by reaction-diffusion equations with a nonlinear term representing the Michaelis-Menten kinetics of an enzymatic reaction. The conditions at which the 1-D model can be applied to simulate the micro-reactor response accurately were investigated numerically. The numerical simulation at transition conditions was carried out using the finite difference technique.

The mathematical model and the numerical solution were validated by experimental data.

The Registration of Computed Tomography Scan Slices Using the Model of the Ribs-Bounded Contour

M. J. Bilinskas, G. Dzemyda

Institute of Mathematics and Informatics
Vilnius University
Mykolas.Bilinskas@mii.vu.lt

In this work, a method for registering the transversal plane images from computer tomography scans is developed. The registration is used when seeking for similar slices in different scans of the same patient. This method is based on the approximation of ribs-bounded contour by a mathematical model. The ribs-bounded contour is determined and the problem of approximation is solved by finding out the optimal parameters of the mathematical model using the least-squares method. The model allows us to speed up the registration of images. The experiments show that the new method improves the accuracy of registration as compared with other ones, but it is limited to the slices where the ribs are visible.

Factors to Predict Attrition Among First Year Computer Science Students

J. Borzovs, L. Niedrite, D. Solodovnikova, A. Klavins

University of Latvia

Darja.Solodovnikova@lu.lv

The stable trend to lose from one-third to half of students in the first study year of computing studies motivated us to explore, which methods are used to determine in advance such applicants, who have no chance to overcome the first study year. Initially, a research about the factors influencing the attrition in Faculty of Computing at the University of Latvia was conducted. The results of our study indicated that none of the studied factors is determinant to identify those students, who are going to abandon their studies, with great precision. The research revealed that the trend of non-beginning studies might indicate the wrong choice of the study field and possible lack of understanding of what is programming by enrolled students (applicants as well as pupils). During the research, some promising programming aptitude tests were discovered for more detailed study. The study concluded that the most promising psychological and problem solving self-test summary should be offered to the prospective students. University of Latvia conducted a research based on a “Cambridge Personality Questionnaire and Behaviour Scale” to find out the correlation between systemizing quotient and empathy quotient and the computing skills of the prospective students. Testing was carried out for secondary school computing students. Further study is conducted by testing the 1st study year students.

The Bayesian Approach to Global Optimization

J. M. Calvin

New Jersey Institute of Technology

Calvin@njit.edu

Many problems in engineering and science require finding parameter values that minimize a cost function.

If no information is available about the number and location of minimizers (in particular if the function is not known to be convex) then the problem is a global optimization problem.

For such problems, the information available to the optimizer, such as function values and derivatives at a finite number of parameters values, is

insufficient to know the optimizing values and one must settle for an approximation.

A popular approach to approximating solutions of global optimization problems is to adopt a probability model for the unknown cost function.

The probability model motivates optimization algorithms, and also allows for analyzing the average approximation error of algorithms.

In this talk, we describe some of the algorithms that have been proposed based on probability models, and some average-case complexity results.

On Algorithm Notion

A. Čaplinskas, J. Miliauskaitė

Institute of Mathematics and Informatics

Vilnius University

Jolanta.Miliauskaite@mii.vu.lt

The presentation analyses the modern notion of algorithm with the aim to highlight compositional properties of modern algorithms. In the 1940s Turing, Church, Hilbert, Kleene and Gödel formalise the intuitive notion of algorithm. For long time their ideas remained rigid and fixed, but with the advent of modern computing technologies the new ideas emerged about the notion of algorithm and, especially, about the nature of computation. It seems today, that algorithm and computation are far broader and complex phenomena than that modelled by Turing machine. Despite the fact that there still not exists generally accepted understanding of computing, these phenomena are intensive investigated in many domains, including computer sciences, information systems engineering, AI, mind philosophy, mathematical logic and even physic. Our work is an attempt to synthesise to some extent these different points of view, ideas, and opinions. The main conclusion is that the modern computing theory does not neglect the Turing computability theory, which still remains very important special case that deals mostly with the traditional sequential algorithms.

Detection of Melanoma Skin Cancer by Classification of Deep Neural Network

P. Čepulionis¹, R. Raišutis¹, S. Valiukevičienė², G. Linkevičiūtė²

¹ Prof. K. Baršauskas Ultrasound Research Institute
Kaunas University of Technology

² Department of Skin and Venereal Diseases
Lithuanian University of Health Sciences
Paulius.Cepulionis@ktu.lt

Melanoma is a potentially fatal disease. Each year there are more 130,000 cases of melanoma diagnosed and 8.5 % of patients diagnosed with this form of cancer did not survive over the past 5 years. Skin cancer segmentation is an important task in medical image processing. Early diagnosis of skin cancer plays an important role in improving treatment possibilities and increases the survival rate of the patients.

There are number of existing review papers, focusing on traditional methods for skin cancer image segmentation. Different than others, we focus on the recent trend of deep learning methods in this field. Deep learning methods can enable efficient processing and objective evaluation of the large amounts of image data.

This study explores the significance and impact on the application of the deep learning techniques to the task of classification of skin cancer images, in particular utilising convolutional neural network (CNN), aiming at providing supplementary information for the early diagnosis of melanoma. More than a hundred features were extracted from images to categorize melanoma images.

Experimental results demonstrated that the proposed method performs well for image classification and may be potentially used in clinical computer-aided diagnosis of skin cancer.

Deep learning methods can assist in the diagnosis of melanoma. Due to its ability to provide a unique insight into the skin structure, Deep learning neural networks is becoming a preferred tool for the diagnosis of pigmented skin lesions and early melanoma detection. In particular it can increase the performance of practising clinicians in the early diagnosis of a deadly disease.

Different Views to Law and Representation of Meanings

V. Čyras¹, F. Lachmayer², K. Lapin¹

¹ Vilnius University

² University of Innsbruck

Vytautas.Cyras@mif.vu.lt

We address the “meaning of meaning” problem in the domain of law. This theme was addressed by Ogden and Richards in 1923 from the standpoint of semiotics. Their semantic triangle (thought, thing, symbol) is nowadays extended to a tetrahedron. An actor, a system viewer, and representer is added. Different views of the system of law emerge because different stakeholders view the law from different standpoints. Understanding the meaning of law is important in compliant software engineering.

We follow the 8 views/4 methods/4 syntheses model proposed by Erich Schweighofer. His model describes the eight different representations of a legal system and four computer-supported methods of analysis, which lead to a synthesis, a consolidated and structured analysis of a legal domain, either a commentary, an electronic legal handbook, or a dynamic electronic legal commentary, or a representation for citizens, or a case-based synthesis. The eight views (or representations of law) are text corpus, metadata view, citation network view, user view, logical view, ontological view, visualization view, and argumentation view. The four methods are an interpretation (search, reading and understanding), documentation (search and processing), structural analysis (conceptual and logical), and fact analysis.

We distinguish between two kinds of meaning of a legal act, the institutional meaning and the content. Representing the content of a norm (or a legal act) solely in rules leads legal expert systems to a failure. Implementing the institutional meaning is necessary. Successful systems, such as tax administration systems or citizen recommendation systems, model situations in which the actors encounter legal meanings.

Within the visualization view, we propose a structural legal visualization (SLV) approach. SLV is about the generation of diagrams which facilitate comprehension of legal meanings. SLV differs from information visualization. SLV relates to a scenario-centred graphical narrative rather than information display. Different pathways through the informational space are concerned.

Speaker Recognition Using Deep Neural Networks

L. Dovydaitis, V. Rudžionis

Kaunas Faculty of Humanities

Vilnius University

Laurynas.Dovydaitis@khf.vu.lt

Speaker recognition task is a challenge that is mainly addressed by extracting robust features from speaker voice utterance sample and afterward creating a speaker model using different techniques. Later in this process, the conclusion is made by comparing given sample models in database and model generated from the new sample. Previously Hidden Markov models were used to create these models. In recent years a breakthrough was achieved using deep neural networks and studies have shown that there is a moderate improvement in recognition accuracy while using this technique. In this work, we present results on different neural network architectures and recognition results based on their accuracy. Features for model creation are extracted using Mel-frequency cepstral coefficients. We conclude that by using Long Short-term memory deep neural network architecture we achieve best results on our sample database.

The Assessment and Integration of Automatic Method of Ultrasonic and Spectrophotometric Technologies Measurements in Order to Estimate the Parameters of Lesion of the Human Tissue

I. Drulytė¹, T. Ruzgas², R. Raišutis¹, S. Valiukevičienė³,
G. Linkevičiūtė³

¹ Prof. K. Baršauskas Ultrasound Research Institute
Kaunas University of Technology

² Faculty of Mathematics and Natural Sciences
Kaunas University of Technology

³ Department of Skin and Venereal Diseases
Lithuanian University of Health Sciences

Tomas.Ruzgas@ktu.lt

Nowadays the growth of skin tumours is more and more widely spreading due to too much misuse of ultraviolet rays, also it depends on changes in faster lifestyle and genetics. Having more accurate estimation of the

lesion parameters of human tissue structure impacts the faster diagnostic, more reliable prediction of the illness, more saved lives and also the lower cost of treatment.

In medicine, data mining methods and algorithms have ontological complexity, various standards of medical data and changing quality of data. All these features with the privacy of patient data policies leads to more effective and practical implemented improvement of medical knowledge. Nowadays medicine is able to monitor not only diagnostics or treatment methods, but also the changes of health and concepts of diseases. More attention is given to responsible attitude to patient, where the automatic data mining methods are very important.

The methodology of ultrasonic and spectrophotometric measurements of post-processing and features identification leads to more accurate classification of skin lesions. The automatic statistical post-processing method is intended to estimate the parameters of malignant tumours and to increase the accuracy of measurements. The accuracy of classification of combining two different technologies in order to identify the malignant tumours and benign nevus is more than 80 percent. Furthermore, this method compared to classification of ultrasonic images only is approximately two times more accurate. The created software leads to a faster diagnostics and optimization of time consuming analysis procedures of medical images. The proposed prognostic models for classification of skin tumours could be used as a decision making tool in the field of dermatology.

Efficient Utilization of Multicore Processors in PC Clusters

E. Filatovas¹, O. Kurasova², T. Petkus², A. Igumenov¹,
S. Ivanikovas¹

¹ Vilnius Gediminas Technical University

² Lithuanian University of Educational Sciences

Ernestas.Filatovas@vgtu.lt

Many real-world computationally expensive problems can be solved by using PC clusters available in universities without employing resources of high-performance computing centres. In this research, we address the problem of effective utilization of multicore processors in PC clusters. There is a range of the problems that can be parallelized by master-slave strategy, where processes do not need intensive data exchange. In such

problems, quality of the obtained results mainly depends on the number of the processed tasks (e.g., the number of function evaluations). We have simulated a general problem of this type, and it has been parallelized by varying the task sizes and number of processes dedicated to each CPU core. The investigation shows that when executing multiple processes in each CPU core allows solving parallelized problems faster compared to the case, where the number of processes is equal to the number of CPU cores.

Modern Information and Communication Technologies and Their Impact on Decision Support Systems

F. G. Filip

Romanian Academy, Romania
ffilip@acad.ro

In recent years, important progress has been noticed in information and communication technologies (I&CT). The new technologies have had seriously influenced the industry business models and human skills and knowledge. Collaboration engineering, a new methodology, got traction in the new context. The paper aims at reviewing such developments with a particular emphasis on the collaborative decision-making activities in control and management (C&M) settings. The attributes of the current day organization, which is characterized by an ever increased degree of intra- and inter-enterprise collaboration, are reviewed first and the transition of the C&M schemes from genuine multilevel hierarchical structures to ever more cooperation schemes are described. Several advanced key technologies, such as BI&A (business intelligence and analytics), web technology, social networks, mobile and cloud computing that enable collaboration are reviewed and their impact on collaborative management and control activities are discussed.

Two-Stage Variant Calling Method for Next-Generation Sequencing Experiments

Š. Germanas

Vilnius University

Sarunas.Germanas@mf.vu.lt

Next-generation sequencing (NGS) is often used to identify genetic variants. This sequencing technique suffers from large sequencing errors and demands sophisticated mathematical methods to control this problem. Therefore probability of variant detection depends on the variant caller. There are several variant calling procedures which assume constant mutation and sequencing error probabilities across genetic positions of target region. These assumptions may lead to smaller sensitivity and specificity of a variant caller. Here we propose a novel variant calling approach. It consists of two stages – the first is clusterization of subregions of target region regarding the estimated mutation and sequencing error rates and the second is testing of existence of an allele frequency changepoint in that subregion. We modify known clusterization techniques in order to apply them to genetic data and use non-parametric maximum likelihood approach to test for an existence of allele frequency changepoint. We apply our method to the genetic data provided by 1000 genomes project.

Energy-Aware Scientific Computing

E. M. Garzón

University of Almeria, Spain

gmartin@ual.es

Nowadays an intensive effort is being developed to design approaches for improving energy/power efficiency of computational devices and platforms. Energy costs represent a relevant share of the total costs of High-Performance Computing (HPC) systems. They include several kinds of processing units, such as CPU cores and GPU, whose energy consumption depends on the kind of processing which is being performed. The tutorial revises the main resources to measure power/energy on architectures and the approaches to improve the energy efficiency of HPC systems. Then, we focus on approaches based on resources selection, which optimize energy efficiency on multicore and/or multi-GPU architectures. Moreover, a tool is analyzed to automatically find the

optimal resources on heterogeneous platforms when iterative algorithms are executed. The approach allows automatically adapting the resources selection to the combination platform-resources/problem-size. This way the energy efficiency is optimized without previous knowledge about the HPC system.

Remarks on a Multi-Criteria Simplicial Optimization with an Estimate of Lipschitz Constant

A. Gimbutas, A. Žilinskas

Institute of Mathematics and Informatics

Vilnius University

Albertasgim@gmail.com

Multi-criteria optimization has been applied in many fields of science where optimal decisions need to be chosen when a trade-off between two or more conflicting objectives has to be made. In this paper, we provide some remarks on multi-criteria multi-variate deterministic algorithms for black-box Lipschitz functions with an unknown Lipschitz constant. In particular, simplicial optimization using an estimate of a Lipschitz constant is analyzed and comments upon the different approaches to one-step worst- case optimal decision making strategies for multi-criteria optimization problems are provided.

Investigation of the Effect of the Assumed Statistical Model in Global Optimization

G. Gimbutienė, A. Žilinskas

Institute of Mathematics and Informatics

Vilnius University

Grazina.Gimbutiene@mii.vu.lt

Several different statistical models are considered as an assumed model of the objective function in global optimization, organized according to two basic conceptual algorithms: the P-algorithm and the maximum expected improvement algorithm. The results of an experimental investigation are presented to show the performance of the considered algorithms in the situations when the algorithm is constructed with either a correct or false assumption about the statistical model of the objective function being minimized.

Cell-Free Blood Circulating DNA Methylation in Cancer

J. Gordevičius^{1,2}, K. Koncėvičius¹, P. Gibas^{1,2}, A. Kriščiūnas¹,
T. Grigalis¹

¹ Institute of Mathematics and Informatics
Vilnius University

² Institute of Biotechnology,
Vilnius University

Juozas.Gordevicius@mii.vu.lt

Colorectal and prostate cancers are the most prevalent forms of malignancy worldwide. Detection and treatment at a pre-metastatic stage is the most effective way to reduce mortality from these cancers. Use of a simple and effective biomarker for early diagnosis can prevent more deaths than any pharmacologic approach to advanced disease.

Here we present three studies investigating DNA modification (methylation) profiles in the cell-free blood circulating DNA (cirDNA). First, we investigated 200 cases and 200 controls of colorectal cancer. Second, we investigated 50 case and 50 controls of prostate cancer. Third, we investigated metastatic prostate cancer individuals treated with abiraterone acetate. In all the cases we built classifiers that can predict the disease or predict response to treatment. These findings suggest that cirDNA modification studies can lead to the identification of non-invasive diagnostic and prognostic biomarkers. However, more sensitive technology is necessary to improve prediction accuracy.

Authorship Attribution Techniques for Short Instant Message Analysis

V. Gurčinas, A. Čėnys, N. Goranin

Vilnius Gediminas Technical University

Vitalijus.Gurcinas@vgtu.lt

Authorship attribution is a complicated task on the internet since anonymity is one of the key internet features. Cybercriminals know how to comprehend the benefits of anonymity as well. They use modern techniques to destroy traces and evidence. Anonymity is one of the reasons for the high rise of cybercrimes. Cybercrime investigations face new problems related to authorship attribution. Investigators have to find

an indirect or direct linkage to criminals by using stylometric text analysis. Traditional linguistic or stylometric text analysis methods are adapted for long texts but are completely unsuitable for short texts used in instant message communication platforms that are dominant in internet communication. Other problems are related to constantly changing forms of language, used abbreviations, emoticons, automatic corrections that hide individual user features. Also, authorship attribution problem becomes very difficult to resolve due to specific language features (e.g. nonnative speaker). The complex analysis methods should be used for solving this problem. In our research, we highlight the main problems of short text authorship attribution, verification, and profiling using stylometric and contextual analysis. The possibilities of short text user identification using popular stylometric tools are evaluated. Test was performed with initial big feature set and later with selected significant and proposed language specific feature set. Application of selected significant and proposed language specific features has allowed to increase the reliability of authorship attribution by 15-25 percent compared to the use of standard feature set. The results have shown that due to today's instant message conversation characteristics that combine formal and informal language, only language independent feature sets can be used for authorship attribution and feature extraction for accurate authorship attribution. Also, the results have shown, that it is insufficient to use typical research methods or empirical data and essential features should be extracted and prepared each time for authorship attribution due to their nature and rapid changes in time.

Principal Component Versus Data Envelopment Analysis in Construction a Composite Indicator for Education Monitoring

A. Jakaitienė, D. Stumbrienė, R. Želvys

Institute of Mathematics and Informatics

Vilnius University

Dovile.Stumbriene@mii.vu.lt

We can monitor various indicators of education system over time and rank periods of each indicator, but we can't rank the countries or the periods of all indicators. It is not possible to rank if we do not aggregate the indicators. Composite indicators are very useful for this purpose. This research focuses on the construction of a composite indicator for the

education monitoring. We went through following five stages for construction of composite indices: data treatment, data normalization, weighting, aggregation and comparing the indices. At the first stage, we used single imputation for missing data. Moreover, all indicators were treated as the profit type - "the larger the better". At the second stage, we standardized data by subtracting the mean of the data and dividing by the standard deviation. At the data weighting stage, we used two different methodologies in order to compare how the different approaches affect the results.

The first one is principal components analysis, which groups individual indicators which are collinear to form a composite indicator that captures as much as possible of the information common to individual indicators. The second one is the application of data envelopment analysis known as the "benefit of the doubt" approach, which selects weights so as to maximize the composite indicator for each country. Composite indicator for education monitoring was calculated for Lithuania, Latvia, Estonia, United Kingdom, Finland and Germany over time using data from EUROSTAT and OECD databases. The index was constructed following structural CIPO framework, which describes relationships between Input, Process and Output in education within a certain Context. This model includes comprehensive information of education system.

An Application of Radial Distortion Model for Chromatic Aberration Correction

V. Jakštys, V. Marcinkevičius, P. Treigys

Institute of Mathematics and Informatics
Vilnius University
Vytautas.Jakstys@mii.vu.lt

This study focuses on the lateral chromatic aberration correction in images captured with Optomed SmartScope M5 camera. Lateral chromatic aberration is observed when different wavelengths are focused on the same focal plane but placed adjacent one to another. Without achromatic lenses, the camera images of eye fundus have blurred vasculature; unsharp boundary of optic nerve disc; artificial color appearance. For photo camera system without chromatic lenses, it is necessary to apply image processing algorithms for lateral chromatic aberration effect correction. Previous investigation of calibration images taken by the camera showed that lateral chromatic aberration could be

equated to configuration of radial distortion, which reveals these features: from the aberration centre it is symmetric in all directions, also the size of distortion is always increasing in proportion (according to a certain function) from the centre to image borders. The most popular second-order radial distortion model used to reduce effects of chromatic aberration. This model compresses red channel and expands blue channel relative to the green channel using the following parameters: chromatic aberration center, the relationship between the magnitude of the distortion and the distance to the center. In this study, second-order radial distortion model compared with affine and polynomial models for lateral chromatic aberration effect correction.

Comparative Analysis of Abnormal Prostate Region Detection Algorithms in mpMRI Scans

J. Jucevičius¹, P. Treigys¹, R. Briedienė², I. Naruševičiūtė²,
J. Bernatavičienė¹

¹ Institute of Mathematics and Informatics
Vilnius University

² National Cancer Institute
Vilnius University

Justinas.Jucevicius@mii.vu.lt

Prostate cancer not only being one of the most frequent tumours worldwide, but also one of the most morbid tumours draws attention to improving the ways it can be diagnosed. Currently, the main method to detect prostate cancer is biopsy. However, it gives correct results only if performed on the right location, thus 20-30% of all tests give false results. This usually leads to repeated biopsy which in turn increases the risk of possible side effects such as temporary erectile dysfunction and urinary problems.

Multi parametric magnetic resonance imaging (mpMRI) can be used to determine the location to perform biopsy on, reducing both the number of samples needed to detect prostate cancer and the chance of needing repeated biopsy test. Abnormal prostate areas must be found in least three different scans of mpMRI to detect prostate cancer: T2W, DWI and ADC. Otherwise such areas are only treated as suspicious and need further tests. Usually the detection of these areas is performed manually, but is often ambiguous. This causes the need of software to aid locating abnormal prostate areas. This work is dedicated to analysing currently used

algorithms for locating abnormal prostate regions in three different mpMRI scans: T2W, DWI and ADC, their effectiveness and arising problems. Further study will be conducted to determine possible ways of improving abnormal prostate region localization algorithms.

The Strength and Beauty of Optimization: The Case of Radiotherapy Planning in the Service of the Society

I. Kaliszewski

Systems Research Institute
Polish Academy of Sciences
Ignacy.Kaliszewski@ibspan.waw.pl

Radiotherapy planning is a key stage in a chain of actions preceding any cancer radiation therapy and it gives rise to large-scale multiobjective problems. The battle here is for manageable problem sizes, solution accuracy and planners' agility in selecting clinically acceptable solutions (irradiation plans). As cancer diseases put nowadays a heavy toll on societies worldwide, the outcome of this battle is of importance to be measured on societal scales.

In the presentation, we shall make an attempt to list issues which should be addressed from the large-scale multiobjective perspective to improve the process of radiotherapy planning. It is expected that the tangible gains of the efforts will be twofold: better radiation therapy treatment effects, but also a more effective usage of equipment, personnel, and resources of radiation oncology departments. We also present our personal experience from a cooperation with an oncology clinic aimed at radiotherapy planning enhancements.

Control of Human Excitement as Reactions to a Dynamic Virtual 3D Face

V. Kaminskas, E. Ščiġlinskas

Department of Systems Analysis
Vytautas Magnus University
Vytautas.Kaminskas@vdu.lt

This paper introduces how generalized minimum variance control principles are applied to the control of human emotion – excitement

signal. We use changing distance-between-eyes in a virtual 3D face as a stimulus - control signal. Human responses to the stimuli were observed using EEG based signal - excitement. We have investigated predictive input-output structure models for exploring dependencies between virtual 3D face features and human reaction to them. A generalized minimum variance control law is synthesized by minimizing quality control criterion in an admissible domain. Admissible domain is composed of control signal boundaries. Sufficiently good control quality of excitement signal (maintained signal level is at average about to 80% higher) is demonstrated by modelling results. We can decrease variations of the control signal using a limited signal variation speed or changing weights coefficient in a generalized minimum variance control.

Discrete-Valued Time Series: Models and Statistical Inferences

Y. Kharin

Research Institute for Applied Problems of Mathematics and Informatics
Belarusian State University
Kharin@bsu.by

Development of mathematical models and statistical inferences based on these models for discrete-valued time series (random processes with discrete time and with discrete state space) is a new intensively growing branch in modern Data Analysis. Results of this branch are widely used in genetics, economics, medicine, ecology, etc.

We present results of our research for three new models of discrete time series (DTS):

- Q – conditional autoregressive DTS of order s ;
- Parsimonious models of high-order Markov chain;
- Multivariate Markov chain with partial connections.

For these models we present the following results:

- theoretical properties of the models;
- algorithms for statistical estimation of model parameters;
- statistical tests for the hypotheses on the values of model parameters;
- statistical forecasting statistics for the future values of the observed

DTS.

Theoretical results are illustrated on simulated and real statistical DTS.

Methods and Means of the Customs Data Mining

S. Konovalenko

University of Customs and Finance, Ukraine

Customslab@ieee.org

The work is dedicated to research and development methods and means of the processing information of customs control from fiscal authorities of Ukraine. As is not uncommon to provide economic agents false information about the characteristics of goods for understatement (overstatement) of the customs value, transportation of smuggling, the use of high-quality business intelligence will improve the fight violation of customs legislation. Currently, there are many methods and tools are used to develop decision support systems, classification, recognition, but from this mass want to highlight methods of Data Mining, a model that use the Rough Set theory. This approach (appeared in the early 80s) developed and used for extracting knowledge from databases. So, on the basis of the approximations creating a classifier that uses the of production model (if {condition} then {class}). Inductively generated rules form the basis of knowledge that solves the problem of generalization and next classification of the input vector. Using rough sets theory allows simply teacher training (supervised learning) and create an understandable classifier, but customs data – is primarily a multidimensional vector. The capacity of training sample must also be large enough to display an adequate state of thinks and can have continuous values of conditional attributes, which greatly complicates the formation of equivalence classes (because of their increasing excessiveness). Numerical, continuous attribute values proposed to use the method of forming an orderly grouping of minimax set intervals corresponding classes decisive attribute. To do this, follow these steps:

Step 1. Form for each attribute of the class ordered pair {min, max};

Step 2. Hold the unification of all ordered pairs {min, max} for each attribute, thus forming an ordered set of intersections of these unities. This will be our minimax set.

Analysis and Synthesis of Lithuanian Consonants

G. Korvel

Institute of Mathematics and Informatics

Vilnius University

Grazina.Korvel@mii.vu.lt

In order to solve the problem of Lithuanian speech synthesis, it is necessary to develop mathematical models for speech phonemes. The models of all the vowels and semivowels have been created. Now the main attention is given to the consonants. Lithuanian speech has 34 consonant phonemes. These phonemes consist of hard and soft spoken consonants. This paper provides an overview of Lithuanian speech phonemes. Voiceless stop consonant modelling framework is proposed. The proposed framework decomposes signals of complex form into the sums of simpler basic components (formants). The second order quasipolynomial is chosen as the formant model in time domain.

Optimization of Efficient Absorbing Boundary Conditions for Schrödinger Equation

R. Kriauzienė^{1,2}, A. Bugajev¹, R. Čiegis¹, T. Leonavičienė¹,
J. Žilinskas², G. Jankevičiūtė¹

¹ Vilnius Gediminas Technical University

² Institute of Mathematics and Informatics

Vilnius University

Kriauziene@gmail.com

We consider the linear one-dimensional Schrödinger equation. The aim of this research is to optimize the coefficients of special boundary conditions for Schrödinger equation.

In practice to solve Schrödinger equation often artificial boundary conditions are used, because the standard boundary conditions corrupt the modeling results. There exist different approaches solving this problem.

We will focus on the boundary conditions that approximate the boundary accurately enough but still with some amount of computational error. These conditions are derived from exact boundary conditions, however, they are much more efficient in terms of computational complexity. Exact boundary conditions use non-local operators, this leads to significant increase in computational costs. Moreover, the computer memory grows

fast when the modeling time increase – it becomes impossible to perform long-term modeling due to computer memory limits.

J. Szeftel introduced the reflection coefficient and used it as an objective functional in minimization problem, which is solved by the simplex method.

In our work, we analyze the optimality of the proposed technique by performing alternative fitting of parameters for two different examples. The question is how to choose these coefficients in order to get as small errors as possible. So the objective functions for optimization are related to errors due to boundary approximation.

Acknowledgment: This research was funded by a grant (No. MIP-074/2015) from the Research Council of Lithuania.

Table-Top Exercise Gamification with Dynamic Scenario for Cyber Security Qualification Assessment

V. Krinickij, L. Bukauskas, A. Brilingaitė

The Department of Computer Science II
Faculty of Mathematics and informatics
Vilnius University
`Linus.Bukauskas@mif.vu.lt`

Table-top exercises help professionals to learn and analyse different scenarios of human behaviour and get prepared for real-life situations. In particular, it is important to use table-top exercises with gamification capabilities for agile adaptation of learners competencies. Paper-based table-top exercises are one of the most common when simulating a small number of cyber events and situations. Agile control of such exercise is limited in time and is prone to human error. Evaluation of learners competencies is obscured and most of the time over-simplified.

The complexity of the table-top exercise is also limited to the number of learners and the complexity of their collaboration network, verbal and non-verbal communication skills. Most of the implemented paper-based table-top exercises complexities proportional to the number of participants. More learners participate - much more complex decision path is worked out - fewer scenarios is exercised through, and insufficient stress situations are tested.

This research paper describes a multi-dimensional evaluation of competencies based on a large number of collaborators balancing the

stress level of the learners. The prototype of a table-top exercise implements an online service that acts as a broker for stress and complexity balance as well as is providing a specialised environment for the table-top scenario play. The implemented prototype puts an effort on the test of general cyber and organisational skills. The prototype is introducing a number of specialised roles users can choose from as well as the difficulty of a scenario.

Acknowledgment:

Authors would like to express great thanks to a whole team involved for The First National Cyber Security training "Cyber Shield 2016" that was done in collaboration with National Cyber Security Center of Lithuania. The authors are thankful for the high-performance computing resources provided by the Information Technology Research Center of Vilnius University.

Prudent Monitoring of Quality and Cost Containment in Health Care Systems

D. Krikščiūnienė, V. Sakalauskas

Kaunas Faculty of Humanities

Vilnius University

Virgilijus.Sakalauskas@khf.vu.lt

The emerging costs for healthcare are becoming a major problem of the society (EU wide 1,400 billion/year). While the rising trend in costs is continuous and significant, the efficiency still can be significantly improved by developing innovative care models. The research aims to enable breakthrough in the field of cost-containment in healthcare while also maintaining the quality of health care.

We propose the model for efficient long term health care by using prudent monitoring concept. This implies designing cost containment structure, evaluation metrics and budget constraints of specific areas of health care. In order to achieve sustainability of health care and reliable long term forecasting we propose risk evaluation of relevant conditions and scenarios. Prudent monitoring outputs are measured as a coherence percentage of possible cost reduction and correlated quality level including indication of risk. The results of the research, monitoring and visualization of evaluation parameters will be tested in the framework of the COST action "European Network for cost containment and improved quality of health care" http://www.cost.eu/COST_Actions/ca/CA15222.

User Needs and Quality in Use: an Overview of User Satisfaction Models

K. Lapin

Faculty of Mathematics and Informatics
Vilnius University

`Kristina.Lapin@mif.vu.lt`

User satisfaction with software products is studied from two perspectives: product manufacturer's and those who buy and use their products. The first perspective is covered by software engineering, the second is investigated in human factors as part of ergonomics. Human-computer interaction relates to both, as a multidisciplinary subject integrates knowledge of ergonomics, software engineering, psychology and sociology to the development of pleasurable software products.

User needs are underlined in ergonomics and were primarily described by Maslow in a hierarchy of human needs in 1970. It states that the human reaches a state of complete satisfaction only temporarily because once people have fulfilled the needs lower down the hierarchy, they will want to fulfil the needs higher up. Patrick Jordan applied Maslow's hierarchy to human factors and proposed three-level user needs a model which consists of functionality, usability, and pleasure. Functionality covers the basic needs, any lack causes dissatisfaction. Satisfaction is included as an attribute of usability in ergonomic standard ISO 9241-11. Satisfaction is also a part of many other models, such as Norman usability principles. They create a common language and support communication between business stakeholders and interaction designer during the analysis of business goals, user objectives, and usage context. It is important to deliver gained product assumptions in a consistent way to the software development team.

In software engineering, satisfaction with a product is considered as an attribute of quality in use, defined in ISO/IEC 25010 SQuaRe Quality model. Both perspectives interpret differently common terms. For example, user satisfaction in human factors' usability context is limited to the avoidance of discomfort whereas in software engineering it includes also the notion of comfort and pleasure.

Our goal is to investigate the models which include the notion of user satisfaction and identify similarities and differences in their contexts.

Comprehensive Immunohistochemistry: Digital, Analytical and Integrated

A. Laurinavičius

National Centre of Pathology

Vilnius University

Arvydas.Laurinavicius@vpc.lt

Immunohistochemistry (IHC) is widely used in pathology as a diagnostic and, increasingly, as a prognostic and predictive tool. While “conventional IHC” is based on visual evaluation and provides rather limited “readouts”, digital microscopy technologies open new perspectives for digital IHC enabling retrieval of high-capacity data by image analysis. Better accuracy and reproducibility of the IHC measurements is expected. Even more importantly, objective evaluation of the spatial heterogeneity of biomarker expression and tissue microenvironment as well as development of combined/integrated biomarkers can be achieved. A concept of “comprehensive IHC” (Laurinavicius et al., 2016) is presented as a multistep process starting with standardization of image analysis and sampling aspects, validation and calibration of IHC biomarker quantification and developing new methodologies to measure spatial and combinatorial aspects of biomarker expression with subsequent clinical validation.

Laurinavicius A, Plancoulaine B, Herlin P, Laurinaviciene A: Comprehensive Immunohistochemistry: Digital, Analytical and Integrated. *Pathobiology* 2016, 83(2-3):156-163.

Anomaly Detection in Financial Markets Using Investors Sentiment Indicator

M. Liutvinavičius, V. Sakalauskas, D. Krikščiūnienė

Kaunas Faculty of Humanities

Vilnius University

Marius.Liutvinavicius@khf.vu.lt

Emotions based investment decisions in most cases lead to losses. However, investors’ sentiment indicators can be used to eliminate this irrational behaviour and increase the investment efficiency. They reveal the hidden patterns of investors’ behaviour and enable to better predict financial bubbles and markets crashes. In this research we focused on prediction of anomalous situations in financial markets that usually cause

losses. We investigated if investors' sentiment indicator can solve this problem.

We adapted such indicator for long-term periodical investment case. In accordance to sentiment indicator the portfolio assets were sold and the profit was taken when strong greed level was observed. The amount of contribution was also calculated in accordance to greed and fear indicator to prevent of buying overpriced assets.

We simulated the investment process to three exchange traded funds: SPDR S&P 500, SPDR Euro STOXX and SPDR Gold Shares. Six different strategies were investigated. The specific period of time was selected with negative dynamics of asset process. This let to imitate the situations when most investors suffer losses due to irrational behaviour. The results confirmed that indicator of investors mood can successfully predict the corrections of asset prices and prevent from losses.

Hybrid Approach for an Automatic Identification of Multi-Word Expressions for Latvian and Lithuanian

J. Mandravickaitė^{1,2}, T. Krilavičius^{1,3}

¹ Baltic Institute of Advanced Technology

² Vilnius University

³ Vytautas Magnus University

Justina@bpti.lt

A Multi-Word Expression (MWE) is a sequence of ≥ 2 words, which functions as a single unit at linguistic analysis, e.g. syntactical, morphological, etc. Identification of MWEs is one of the most challenging problems in NLP. Many techniques are used for this problem, however, not all of them can be transferred to Lithuanian and Latvian due to rich morphology.

In this stage, we use raw corpus (LT and LV, 9 mln. words for each language) and a combination of lexical association measures (LAMS) and supervised machine learning (ML), and look for bi-gram MWEs. EuroVoc, a Multilingual Thesaurus of the European Union is used to evaluate MWE candidates. The candidate MWE bi-grams were extracted from raw text and 5 LAMs (Maximum Likelihood Estimation, Dice, Pointwise Mutual Information, Student's t score and Log-likelihood) were calculated. Reference lists based on EuroVoc were used for evaluation. Then Naïve Bayes, OneR (rule-based classifier) and Random

Forest were applied. SMOTE and Resample filters were used due to the sparseness. Precision, Recall, and F-measure were used to evaluate the results.

LAMs and ML algorithms were combined in 3 ways: without any filter, with SMOTE, and with Resample. 10-fold cross-validation was used.

The best results for, both Latvian and Lithuanian were achieved with Random Forest+Resample (LV: P = 92.4%, R = 52.2% and F = 66.7 and LT: P = 95.1%, R = 77.8% and F = 85.6%). Our future plans include experiments for extraction of different types of MWEs and a greater diversity of MWEs.

Text Mining for Robotic Action Ontology Engineering

I. Markievicz¹, M. Tamošiūnaitė^{1,2}, D. Vitkutė-Adžgauskienė¹,
T. Krilavičius¹

¹ Faculty of Informatics
Vytautas Magnus University

² Bernstein Center for Computational Neuroscience
University of Göttingen, Göttingen, Germany
Irena.Markievicz@gmail.com

The development of the action ontology (ACAT) from domain specific texts allows to discover previously unknown dependencies between robotic actions and their environment objects. This study explains the conceptual model of the ontology actions and environment objects and relations between them. Two main ACAT ontology classes determine the hierarchical structure of action and object hyponymy/hypernymy, troponymy: „action“ and „object“. Each action and object synset contains a subset of synonymous entities. All synsets from the ontology are described by the semantic roles, used in action execution by robots: main action, main object, primary object and secondary object.

Study also explores various text mining methods for action ontology learning: collocation extraction, frequency lists, bag-of-words, word space model and Hearst's hyponymy recognition patterns. Verbnet thematic roles and frames are used to identify text syntactic and semantic structure – in this way recognized new text patterns allow to define dependencies between ontology synsets. Robotic action classes are identified by text classification with SVM machine learning method,

where action categories are treated as classes, and appropriate verb context – as classification instances.

The action ontology completeness and utility is evaluated empirically, by running as additional source of knowledge base in instruction processing system. This study introduces the preliminary testing results of the ACAT ontology usage in instruction processing to sequence of robotic execution tasks (chosen rotor assembly and biotechnology laboratory scenarios). While the explicit knowledge is parsed directly from the instructions, reasoning on queried ACAT ontology data allows to cover instruction tacit knowledge. It helps to execute human readable instructions with polysemous information, omissions, too general or non-robotic actions and not relevant texts.

Visualization and Automatic Thematic Classification of Voting in the Lithuanian Parliament

V. Mickevičius^{1,2}, T. Krilavičius^{1,2}, V. Morkevičius³

¹ Baltic Institute of Advanced Technology

² Vytautas Magnus University

³ Institute of Public Policy and Administration

Kaunas University of Technology

Vytautas.Mickevicius@gmail.com

Data science and statistical analysis are becoming increasingly involved in political analysis and monitoring of the activities and behavior of the political actors, as more political data becomes available. Parliamentary roll call voting is one of the most interesting and information-rich areas in this case. Statistical analysis and visualization of the results of parliamentary votes is a well-researched topic nowadays, no exception in Lithuania. However, much less attention has yet been paid to the thematic categorization of the votes when conducting similar studies. Statistical analysis, performed on thematically predefined groups of datasets (voting results, according to their thematic category), could yield more informative and interpretable results, in comparison to the similar statistical analysis when only one dataset of voting results is used.

This research is focused on dimensionality reduction and visualization of multi-dimensional voting data, as well as automatic text classification of vote titles building the dataset for more comprehensive statistical analysis of voting in separate thematic categories. The main issues regarding the

multi-dimensional data analyzed are selecting the most suitable coding for the categorical variables (outcomes of votings), and selecting the proper dimensionality reduction methods (Multidimensional Scaling is examined in depth). Main challenges in the automatic text classification part include testing several popular text classification algorithms (Support Vector Machines, Artificial Neural Networks, Naïve Bayes and Logistic Regression), as well as feature selection (correlation analysis, filtering etc) which plays an important role in the research since the text documents classified are rather specific – they are short and the dictionary is small (ratio of the number of documents to the number of unique terms is approximately 4 to 1).

The results of both parts (dimensionality reduction and text classification) are combined, thus allowing examination of the results of the Lithuanian Parliament voting in different thematic categories.

An Application of Recurrent Method for the Analysis of Computer Network

S. Minkevičius¹, E. Greičius²

¹ Institute of Mathematics and Informatics
Vilnius University

² Faculty of Mathematics and Informatics
Vilnius University

Saulius.Minkevicius@mii.vu.lt

Probabilistic models and queueing networks have long been used to study the performance and reliability of computer systems and of distributed information systems. One can apply the theory of queueing networks to obtain probability characteristics of technical systems (for example, the idle function of computer networks). The idle function of computer networks shows which part of time computer network is not busy (idle). We present an example of the application of the recurrent method for the analysis of computer network. An open queueing network model in light traffic has been developed. The probability limit theorem for the idle time process of customers in light traffic in open queueing networks has been presented. Finally, we present an application of the theorem - an idle time model from computer network practice.

The Effect of Body Vibration on the Physical Function of Humans

E. Monstavičius, K. Stasiukynaitė, V. Veikutis

Lithuanian University of Health Sciences

Stasiukynaitė@gmail.com

A lot of people has contact with vibration beginning from household appliances to medical interventions and those vibrations effects our bodies. This review aim is to figure out how much is known about vibration different parameters on human body.

Humans starts to feel vibration from 10 Hz and it is proven that 180 Hz vibration starts to destroy neurons. Most tools work in 50-200 Hz range and does no significant harm. Most controlled studies are done on whole body vibration (WBV) which gives as parameters in range of 15-50 Hz, 1-6 mm amplitudes. But not all studies agree on benefit and harm ratio of this procedure. Some studies testing healthy or osteoporotic elderly women state that this exercise increases bone mass and coordination, other works state creating weak spots on skeleton. WBV can relieve chronic obstructive pulmonary disease symptoms or neuropathies caused by diabetes mellitus or other neurodegenerative diseases.

There are few studies that look into effects on adjacent tissues apart those that are affected directly - making burr holes in skull causes sensorineural hearing loss. Differences in the hearing thresholds at higher frequencies (4-6 kHz) were significant before and after craniotomy. Using an ultrasonic bone curette that produces 500 - 8,000 Hz skull vibration levels can be safely applied without the risk of harmful effects on the inner ear. But they do not test sensorineural hearing thresholds on different frequencies which leads to more questions than answers.

Concluding, there are a lot of research done on WBV, but there are little known about selective body parts vibration effects with different parameters. Also there are practically none research on vibration effects to different tissues and mechanism that brings all in these studies seen outcomes.

Evaluation of an Efficient NSGA-II Version on Heterogeneous Low-Power Platforms

J.J. Moreno¹, G. Ortega¹, E. Filatovas², J.A. Martínez¹,
E.M. Garzón¹

¹ TIC-146 Supercomputación-Algoritmos, Dpt. of Informatics
University of Almería, Spain

² Institute of Mathematics and Informatics
Vilnius University
Ernest.Filatov@gmail.com

The application of Evolutionary Multi-objective Optimization (EMO) algorithms in real-time systems receives considerable interest. In this context, the energy efficiency of computational systems is of paramount relevance. Application of embedded systems based on heterogeneous (CPU and GPU) platforms is consistently increasing. In this research, we address the NSGA-II algorithm that is the most widely-used representative of EMO algorithms with a dominance-based population selection strategy. In the algorithms of the NSGA family, the selection is based on Non-Dominated Sorting (NDS). NDS procedure is the most computationally expensive phase of the EMO algorithms when the computational cost of the evaluation of the objective functions is not high. In this work, we propose a new GPU implementation of NDS which takes advantage of the NVIDIA Jetson TK1 platform. The NVIDIA Jetson boards are considered a prototype of low-power High-Performance Computing systems. An experimental comparison between our GPU version of the NSGA-II and the best known recent GPU implementation of NSGA-II has been carried out in terms of performance and using several population sizes of EMO problems. The results have shown that the new GPU version achieves better performance when several fronts are computed, and/or there is a high number of objectives. As consequence, we also propose an adaptive version that switches between both GPU versions of NSGA-II, improving the performance in all test cases. Additionally, we show that the Jetson TK1 embedded platform is adequate to accelerate EMO algorithms for large numbers of objectives and populations, achieving respectable performance and high acceleration factors.

IVUS-VH Relation to the Extent and Composition of Atherosclerotic Plaque and Its Rupture Risk Prognosis

K. Morkūnaitė, V. Veikutis, E. Sakavičiūtė

Lithuanian University of Health Sciences

E.Sakaviciute@gmail.com

Background and aim. One of the most serious and unpredictable coronary artery disease complication is atherosclerotic plaque rupture. Ordinary visualization methods are not sufficient to detect early signs of plaque remodeling into the high-risk unstable structure. The aim was to determine standard angiography and intravascular ultrasound virtual histology (IVUS-VH) possibilities to identify the high-risk plaques.

Methods. We sorted 50 of 146 plaques analyzed with angiography digital assessment tool and IVUS-VH. The plaque differences were calculated and compared to the histological structure evaluated with IVUS-VH.

Results. Fibrous tissue (FT) formed 57,89% of the plaques (average 2,6 mm²); 19,60% were necrotic (average 0,75 mm²), and (8,58%) were calcified (average 0,3 mm²). Plaques with greater necrotic mass were significantly smaller in angiography compared to IVUS. The differences between stenosis measured by angiography and IVUS were calculated and results were grouped according to percentile: group A - <25 percentile - >18,62% (stenosis significantly higher with IVUS-VH - 15 plaques, 30%); group B - 25-75 percentile (difference interval -18,62% to 12,64% - 23 plaques, 46%); group C - >75 percentile <12,64% (stenosis significantly lower with IVUS-VH - 12 plaques, 24%).

In group A necrosis made up 1,40 ± 1,05 mm², group B - 0,87 ± 0,52 mm², group C - 0,62 ± 0,45 mm² (p=0,020). Observations of FT were similar: group A - 3,38 ± 3,20 mm²; group B - 2,90 ± 2,6 mm², and group C - 2,04 ± 1,65 mm² (p=0,082). Relationship between groups and necrosis (r= -0,40, p=0,004), FT (r= -0,29, p=0,039) components was revealed to be moderate negative by correlation analysis.

Conclusions. Coronary angiography alone is not a good predictor of future events. Using of IVUS technologies is necessary to assess clinical outcomes risk better. Despite advantages IVUS-VH remains costly and not always technically adaptive, so it enforces pursue new methods to identify high-risk atherosclerotic plaques.

An Overview of Methods to Spatially Map Intra-tumor Genetic Heterogeneity in Whole Slide Pathology Images

M. Morkūnas^{1,2}, P. Treigys¹, A. Laurinavičius²

¹ Institute of Mathematics and Informatics
Vilnius University

² National Center of Pathology
Affiliate of Vilnius University Hospital Santariškių klinikos
Mindaugas.Morkunas@mii.vu.lt

The majority of cancer patients have their diagnosis confirmed while performing obtained tissue pathology. There are developed grading systems that classify tumor cell abnormalities and predict tumor developments at a tissue level. Recently introduced the whole-slide scanning systems improved those grading systems by enabling precise quantification of parameters such as lengths, surface areas, object counts, perimeter, distance to nearest neighbor and other features that can be measured while examining tissue specimen image. Advances in image processing research area utilizing unsupervised machine learning techniques enables possibility to make the discovery of other fundamental prognostic data captured by the medical imaging techniques. Features that may not be visually discriminated and interpreted by a pathologist offer the opportunity for even better quantitative disease development modeling.

Tumor heterogeneity can be described as the coexistence of different tumor cell populations within tumor tissue. There are at least three layers of tumor tissue heterogeneity - genetic, meaning that populations of tumor cells harboring different mutation sets exist in the tumor, functional - meaning that genetic variability results in cell populations with altered abilities (growth rate, metabolic activity, proliferation), and phenotypic - meaning that genetically and therefore functionally dissimilar tumor cell populations within the same investigative sample are likely to form morphologically distinguishable structures (differing in cell size, shape, and organization). Tumor heterogeneity can be directly detected and measured as intra-tumor mutation frequency using genomic approaches or by visual specimen image inspection.

Several computational methods based on different statistical models have been developed to extract the composition of the tumor tissues (as the number and proportions of the subpopulations of tumor cells) from genomic data. In this work we discuss the recent advance of digital

pathology methods that allow histological measurement of tumor tissue heterogeneity, and elaborate on possibility to integrate genomics data with digital image analysis.

Accelerating the LRASR Anomaly Detection Method on Heterogeneous Platforms

G. Ortega¹, L.O. López¹, I. García², A. Plaza³ and E.M. Garzón¹

¹ TIC-146 Supercomputación-Algoritmos, Dpt. of Informatics
University of Almería, Spain

² Dpt. of Computer Architecture, University of Málaga, Spain

³ Hyperspectral Computing Laboratory
University of Extremadura, Cáceres, Spain
Gol315@ual.es

Anomaly detection can be defined as target detection without any prior information about the target signature.

Anomaly detection has been researched within diverse areas and application domains, for instance, in hyperspectral images (HSIs) processing. HSIs collect information of a target scene across the electromagnetic spectrum, from ultraviolet to long-infrared. This enormous quantity of information facilitates the identification of objects. In HSIs, pixels that have a significantly different spectral signature from their neighbouring background clutter pixels are defined as spectral anomalies.

In the last decade, the number of algorithms for anomaly detection in HSIs has considerably increased. One of the state-of-the-art methods for anomaly detection in hyperspectral images is called LRASR and it is based on the low-rank and sparse representation of the data. LRASR consists of the separation of the background and the anomalies in the observed data. Since each pixel in the background can be approximately represented by a background dictionary and the representation coefficients of all pixels form a low-rank matrix, a low-rank representation is used to model the background part. Experimental results have indicated that LRASR achieves very promising anomaly detection performance; however, LRASR takes more computation time than other methods.

This work shows several strategies to accelerate the LRASR anomaly detection method for HSIs, with a focus on the most time-expensive sections of the algorithm (mainly the Singular Value Decomposition

operation, SVD). To achieve this acceleration, several libraries have been used to exploit the power of GPU and CPU-GPU heterogeneous architectures. Different experiments have been studied to observe the impact of the libraries over the test images and the SVD operation in their processing. These experiments have shown that heterogeneous libraries for computing the SVD present notable accelerations when working with the largest test images.

Computer-Aided Analysis and Evaluation of Chess Helpmate Problems

L. Paliulionienė, V. Paliulionis

Institute of Mathematics and Informatics

Vilnius University

Laima.Paliulioniene@mii.vu.lt

Chess was one of the first domains where artificial intelligence methods and information technologies were applied. Chess-playing programs ground their success on computer performance, large database of games, pattern-recognition and search algorithms. Chess problem solving engines (where the goal is to give a checkmate in pre-determined number of moves) is another rather popular type of chess-related software. Much less investigated topic is using computer methods in chess composing. In chess composition tournaments, chess problems created by composers are rated and awarded for their originality, aesthetics, and conformity to requirements.

Though there are some few works on computer-aided analysis and aesthetic evaluating of classical chess problems (direct-mate problems and endgame studies), computer-aided analysis of helpmates is almost not investigated so far. Helpmate is a class of chess problems in which black help white to checkmate the black king. While directmates are zero-sum games, helpmates are non-zero-sum games and can be viewed as not competitive but collaborative game. In our work, we analyze computational peculiarities of helpmates and propose an outline of a model that could be used for computer-aided analysis, composing and evaluating quality of chess helpmates.

Commercial Analytics to Improve Business Performance Through Micro Customer Segmentation, Consumer Behaviour Analysis and Advanced Analytical Methods Application

P. Papadopoulos

Ernst & Young Greece

Panos.Papadopoulos@gr.ey.com

Market dynamics are pushing companies across different industries to re-evaluate their Pricing (Budget) and Marketing Strategies in order to improve performance. Revenue and Profit growth objectives are now tighten with tougher margins, cost reduction and the necessity of higher value creation. Analytical services propose innovative solutions to its recipients and offer not only financial numbers growth but also strong customer relationships through enhanced customer engagement. Revenue Management and advanced analytics solutions are now in the front line for creating actionable value for our clients. The reason behind this is that analytics brings customer centricity and better decision-making to the business. So companies shift data resources to analytics for self-service Analytics-On-Demand, streamlined decision engine, prioritize delivery over data. The most important targeted (and specific) service that analytics can offer is the use of BI reporting tools that will help management keep track of essential metrics, the statistical analysis that will reveal causal relationships, the implementation of efficient and effective data management across the whole data lifecycle and above all how to embed a more holistic analytics strategy within each organization. Commercial analytics refer to a range of solutions that is mainly applied in the consumer packaged goods and Retail industry, but its techniques and learnings are applicable to many other industries, such as pharmaceutical, automotive, telco's even banking. Analytics leverages past data to evaluate the underlying relationship between data input and output. The goal is to understand what drives customer and consumer behavior, what affects shopping patterns, and most importantly quantify the impact on sales, revenue and profit of various pricing, marketing but also competitive activities. During the presentation we will exhibit real world case of how applications of analytics helped our clients to get actionable insights, design and implement new strategies. In order to effectively do this and drive better decisions we must first ask the right business questions and then seek answers in the data.

Learning, Data Analysis and Software Systems – Qualitative and Quantitative Dilemmas?

D. Passey

Department of Educational Research
Lancaster University, UK
D.Passey@lancaster.ac.uk

Educational software is often now run through linked or integrated digital management systems that collect increasing quantities and forms of background data. Frequently as a minimum, when users log on, where they log on from when they select pages to view or read, how often they stay on a page when they click on links, who else they contact about their learning or work, and when they log off, are all collected. Arguments are made that these forms of data can be used to interpret features of users that will inform better learning. Some previous research studies have focused on learning features of the individual, using background data in qualitative ways (sometimes displaying outcomes through forms of imagery). Other previous studies have focused on much wider sets of data, gathered from across (sometimes very large) numbers of users; interpretations of those data are sometimes stated to say something about an individual's learning from a statistical or quantitative perspective. This paper explores these different paradigms and perspectives, and through related dimensions concerned with making choices when analysing learning from data in software systems, argues that dilemmas in choice of methodological approach are not the only or necessarily the key dilemmas that researchers face if their research findings are to be of value to the field of education and learning. Different stakeholders – policy makers, educational advisers, head teachers or principals, teachers, parents, students, educational software developers – all need specific forms of data output if they are individually to be most effectively supported in terms of enhancing learning or teaching. I will use a number of studies and their outcomes to illustrate a current gap in our research concern, and consider future implications and dilemmas we face in this field.

Global Optimization of Nonconvex Bilevel Problems: Implementation and Computational Study of the Branch-and-Sandwich Algorithm

R. Paulavičius^{1,2}, C. S. Adjiman²

¹ Institute of Mathematics and Informatics
Vilnius University

² Centre for Process Systems Engineering
Department of Chemical Engineering
Imperial College London, UK
`Remigijus.Paulavicius@mii.vu.lt`

Mathematical programming, and optimization in general have become essential tools in process systems engineering, and have found applications in synthesis, design, control, operations and planning. Many of these optimization problems are bilevel problems in nature and can thus be framed as two-person, hierarchical optimization problems having a second optimization problem as part of the constraints. This includes, for instance, optimization problems in design or parameter estimation in which phase equilibrium is present in the constraints, or problems in which the optimum economic design is sought while minimizing energy consumption. Special cases of bilevel problems, such as problems in which the inner problem is linear, have been studied extensively and many algorithms have been proposed in the literature. However, the general nonconvex form is a very challenging problem. Recently, a new deterministic global optimization algorithm, named Branch-and-Sandwich (B&S), was proposed for optimistic bilevel programming problems that satisfy a regularity condition in the inner problem.

In our current work, we describe our implementation of the B&S algorithm written in C++ within the MINOTAUR framework discussing, wherever appropriate, the use of suitable data structures and associated algorithms. We present detailed computational results of the B&S implementation solving extended set of bilevel benchmark problems, including problems relevant to chemical engineering problems. Alternative choices in the way each step of the B&S algorithm is tackled, as well as different strategies for branching, node selection and for the node lists management are explored in this work.

Massive Data Visualization via Selecting a Data Subset

K. Paulauskienė, O. Kurasova

Institute of Mathematics and Informatics

Vilnius University

Kotryna.Paulauskiene@mii.vu.lt

Every day we face with massive databases and millions of records which come from everywhere. Data visualization is an important approach which helps to understand the data better. The aim of this research is to visualize data points with less overlapping and still keeping the image informatively. We have proposed a visualization strategy which consists of two main steps: the first one is selecting a subset of the given data and then calculating the projection; the second one is the mapping of the projection of the subset. Multidimensional scaling algorithm is usually infeasible to deal with millions of points or requires much computational time. Our approach reduces the computational time during the dimensionality reduction. Random sampling or prior clustering are the most popular ways for selecting the subset. Thus, we propose a simple but efficient selection of the subset. The proposed approach for data subset selection prevents the rejection of outliers from the subset which may be significant for the given data. The efficiency of proposed subset selection is confirmed by a comprehensive set of comparisons. In the end, we show how to visualize the projection of subset without overlapping as well.

Estimation of Teachers' Demand in Municipalities

T. Petkus, K. Straukaitė, O. Kurasova

Lithuanian University of Educational Sciences

Tomas.Petkus@leu.lt

A change of population directly influences the numbers of pupils as well as the numbers of teachers. Recently, there have been seen a total Lithuanian population decline, however, the birth rate although slightly but is growing. In this research, the numbers of pupils are estimated by municipalities, taking into account the birth rate and the changes in the number of the pupils moving from a lower to a higher class. The numbers of the teachers of various subjects in municipalities are also examined. These numbers are determined by the numbers of pupils, teachers' age and the fact that in recent years, the enrolment in teacher training

programs is significantly reduced. The study shows that, in some parts of Lithuanian regions, there have been noticed a need of the teachers of exact subjects, and within a few years there will be a huge problem.

New Diagnostic Possibilities Using Multivariate Analysis of Multimodal Magnetic Resonance Images

R. Petrolis, A. Pilypas, G. Jurkevičiūtė, S. Lukoševičius,
K. Petrikonis, A. Kriščiukaitis

Lithuanian University of Health Sciences
Algimantas.Krisciukaitis@lsmuni.lt

Approximately two-thirds of population suffers from low back pain at some point during their lives, most commonly between the ages of 30 and 50 years. Signal intensity changes of vertebral endplates and subchondral bone are often observed in Magnetic Resonance Imaging (MRI) in the patients with spinal degenerative diseases. Roos with co-authors first reported bone marrow signal changes in the vertebral bodies in MRI images in 1987. In 1988, Modic with co-authors described these changes and proposed their classification. The medical term “Modic changes” (MC) is now used to spinal degenerative diseases becoming one of parameters describing spinal degenerative diseases together with related morphological changes in spinal structures (bone marrow and endplate lesions). Different MC types are reflected as certain changes of signal intensity in different MRI modalities (T1, T2 and their variations). We demonstrate how multivariate analysis applied to the set of adjusted and superimposed multimodal MRI images could be used to construct optimal representation of spinal structures enhancing contrast of certain type of MC in regard to healthy tissue or other types of degenerative changes. The method based on Principal Component Analysis enhances contrast level of lesion area and allows automatic estimation of lesion volume. Method based on Independent Component Analysis could be used to visualize certain pathogenic mechanisms related to whole process of MC.

Application of Fractional Euclidean Distance Matrices to Extrapolation of Scattered Data

N. Pozniak

Institute of Mathematics and Informatics
Vilnius University
Natalija.Pozniak@mii.vu.lt

This presentation is about the application of fractional distance matrices to construct the efficient extrapolator of scattered data by means of a matrix of degrees of the distances between pairs of scattered points. Theoretical properties of the extrapolator developed are sound to study of fractional Euclidean distance matrices: positive definiteness, non-singularity, etc. The resulting model can be presented as a random Gaussian field, defined only by the mean and variance parameters that are efficiently evaluated by the maximal likelihood method. The study and comparison of the developed kriging interpolator with the Shepard interpolator are performed by computer simulation.

Systematic Literature Review of the Cloud-Ready Software Architecture

O. Pozdniakova, D. Mažeika

Faculty of Mathematics and Informatics
Vilnius Gediminas Technical University
Olesia.Pozdniakova@vgtu.lt

Many companies are adopting cloud for hosting applications delivered as a service over the Internet. An application service delivered in such way is referred as Software as a Service (SaaS). The SaaS providers get benefits from getting cloud resources provisioned on-demand and pay-as-you-go billing models. These capabilities enable faster application development and deployment with lower upfront investment into infrastructure.

To benefit the most from the cloud, software architecture must be designed with the consideration that software will run on the cloud. However, a legacy software often is developed by using monolithic architecture approach and might not get all advantages provided by a cloud computing. A growing monolithic application gets less flexible in development, has longer provisioning time, slows down the speed of innovation and lowers economy of scale.

As result of such problems, not only commercial companies, but also the various academic researchers aim to design software as a “cloud-native” application. This kind of software has specific non-functional requirements that define scalability, reliability, fault tolerance and other cloud-specific requirements.

This systematic literature review is based on the case studies, published articles and other literature related to the cloud-native or cloud-ready applications. It generalizes the common traits of non-functional requirements defined in analyzed resources. Also, it walks through the architectural styles of the applications running on a cloud and presents recently developed architectures that are used for large-scale software services delivered on a cloud.

Detailed Analysis of the Dreams Meanings Dataset

A. Rapečka, J. L. Jedzinskas, G. Dzemyda

Institute of Mathematics and Informatics

Vilnius University

`Aurimas.Rapecka@mi.vu.lt`

There are a lot of methods to recommend products or services. Each method has its advantages and disadvantages. For example, widely distributed universal methods demonstrate good results in most data sets, but, because of basic operating principles (calculating the correlation coefficient between the target and every other user in a dataset), these methods require large computing resources. With limited computation resources or large datasets, application of these methods is not successful. Sapnai.net Dreams Meanings Dataset is an example of large datasets. Dataset consists of 27000000 search records by 1470000 users in 4200 dreams meanings. This dataset is available for scientific researches and is published online at www.sapnai.net/db.

The aim of this work is to present a detailed analysis of this dataset. Several challenges were solved in this analysis: determination similar groups of users, relations between dreams by users, relations between week or month days and dreams popularity etc. Also, some strange results were detected (for example, two users randomly picked same dream in same second). Looks like some users have trackers, it can be viruses or other tracking bots. Some more insights of this case were presented.

An Investigation of Surface Deformation Using Singular Value Decomposition

M. Sabaliauskas

Institute of Mathematics and Informatics

Vilnius University

Martynas.Sabaliauskas@mii.vu.lt

Minimizing of surface deformation is of prime importance for preprocessing of 3D models in computer graphics such as texture mapping, animation creation or various surface parameterizations. Commonly used 3D mesh model is a 3-connected graph obtained from point cloud. To calculate a deformation of that 3D model we have to estimate a deformation of each corresponding triangle. This can be done by calculating singular value decomposition of Jacobian transformation matrix between these triangles. This work presents experiments that were made by deforming segments of 3D models using ABF++, LSCM and ARAP parameterization methods. Finally, surface distortions were estimated by calculating general angular and areal deformations. According to mean deformation results in comparison with ABF++ and LSCM method results ARAP method gives the best minimized conformal and authalic distortions therefore each triangle rigidity is best approached.

Distance Matrices Geometry and Data Mining

L. Sakalauskas

Institute of Mathematics and Informatics

Vilnius University

Leonidas.Sakalauskas@mii.vu.lt

Euclidean distance matrices (EDM) have been receiving increased attention because of many their applications, such as molecular conformation in bioinformatics, dimensionality reduction in machine learning and statistics, the problem of wireless sensor network localization, etc. Statistical approach for modeling the response function by random process or random field has been proposed in Zilinskas (1985), Mockus (1989). Fundamental properties of Euclidean geometry, obtained by Schoenberg (1935) have surprisingly late date (Gover, 1985). Note, square EDM are the mostly studied in literature, although Euclidean geometry is dealing very often with matrices, which entries are fractional degrees of distance square. This lecture presents the basics of fractional

distance matrices theory: nonsingularity of EDM, sufficient&necessary conditions for matrix to be EDM, characterization of models with EDM, random Gaussian fields, defined by EDM, as well as applications to scattered data interpolation, clustering, multidimensional scaling, etc. Properties of EDM are explored introducing the kernel matrix.

Simulated Annealing Algorithm Optimization Method Convergence Study

L. Sakalauskas, D. Kavaliauskas

Institute of Mathematics and Informatics

Vilnius University

Donatas.worshipper@gmail.com

The proposed simulated annealing algorithm based on the principle of stable alpha. The algorithm is designed to solve two processes scheduling task. The essence of the task is to put the work of processors so that their performance would be the shortest time The Objective belongs to P versus NP group of tasks. The suggest method is compered with a simulated anealing algorithm without any corrections results. Survey results also suggest trying heuristic algorithms convergence of research methodology.

An Approach to Simulate Business Processes Using Probabilistic Process Models

T. Savickas, O. Vasilecas

Information Systems Department

Faculty of Fundamental Sciences

Vilnius Gediminas Technical University

TSavickas@gmail.com

Business processes are the core of any organization therefore their analysis and improvement is vital. Analysis of real process data as opposed of manual modelling can provide useful insight and process mining techniques facilitate this. Static analysis of process data limits the usefulness of the current approaches. Business process simulation allows dynamic analysis. Integration of process mining techniques into business process simulation creates an opportunity to perform analysis of what-if scenarios. This paper presents an approach how to simulate business processes using probabilistic process models discovered from process

data existing in information systems. This allows the generation of simulation models from process data and see live how the processes behave and analyze how changes in behavior impact the process. The presented approach is based on previous research on probabilistic business model discovery which is integrated with business process simulation tool to facilitate the simulation. The provides initial results on applicability of the approach and further research directions.

A Review of Infrastructures of Internet of Things for Smart Enviroments

R. Savukynas, V. Marcinkevičius, A. Čaplinskas

Institute of Mathematics and Informatics

Vilnius University

Raimundas.Savukynas@mii.vu.lt

Smart environmental objects integration methodologies and technology research and new development are closely linked to each person and whole human society adequate quality of life. Usually seeks to provide maximum comfort, when the separate individual and general society feel full well of the existing technical and technological development conditions. Humanity needs of information technology are getting more ambitious challenges, so the world wide web development opens up great opportunities to create smart environments. Smart environments are usually identified with the smart home solutions and technologies, but smart becomes and other environmental, including smart institutions, hospitals, cars, stores, airports and even smart cities. Smart environments are indispensable for people with disabilities or the elderly, because significantly improves the quality of life, ensure proper health and protects against various disasters. Smart environments and benefits of internet of things include security, comfort, optimal use of energy resources and many other important services, which improves the quality of life in these environments. Smart environments are not limited to the possibility of the use of the services, which are designed for people, but allows you to customize the ideas and technologies in agriculture, industry and even in virtual reality, where virtual program agents communicate with each other and make decisions. Infrastructures are designed for building systems of internet of things and ensure interoperability of internet of things. In this work provides a review of existing infrastructures of smart environments and internet of things.

Time Series Modelling Using a Novel Method of Decomposition

T. Sidekerskienė

Department of Applied Mathematics
Faculty of Mathematics and Natural Sciences
Kaunas University of Technology
Tatjana.Sidekerskiene@ktu.lt

We present the results of time series modeling using a combination of several methods: 1) a proposed method of time series decomposition into a sum of oscillatory intrinsic mode functions (IMFs) based on the non-negative factorization of Hankel matrix of time series; 2) random cointegration of intrinsic modes; 3) selection of best fitting random cointegration vector using a nature-inspired optimization method; 4) modeling of cointegrated time series using ARX (autoregressive with exogenous inputs) model. The results of modeling using a historical data (daily highest price) of S&P 500 stocks from 2009 are presented and compared against ARX models. The results are evaluated using a variety of metrics (RMSE, MAE, MAPE, Pearson correlation, Nash-Suttcliffe efficiency coefficient, Index of Agreement and others) as well as illustrated graphically using Taylor diagrams and Target diagrams. The proposed modeling method can be used for variety applications in the data mining domain (time series denoising, prediction, etc.).

Automated Blood Vessels Diameter Measurement in Eye Fundus Images

G. Stabingis^{1,2}, J. Bernatavičienė¹, P. Treigys¹, A. Paunksnis³,
R. Vaičaitienė⁴, V. Medvedev¹

¹ Institute of Mathematics and Informatics, Vilnius University

² Klaipėda University

³ Telemedicine Research Center

⁴ Gen. J. Žemaitis Lithuanian Military Academy

Giedrius.Stabingis@mii.vu.lt

Arterial hypertension, diabetic retinopathy, aging macula degeneration, and glaucoma are diseases, the condition of which is reflected by the changes in eye fundus anatomical structures. Timely evaluation of changes enables not only to evaluate the degree of compensation but also

to prevent the progression of the disorder. Some of above mentioned disease signs can be evaluated by measuring the vein and artery ratio before and after physical load. Even more, review of the literature reveals that the ratio can be coherent not only to the physical but to the mental state of the investigative as well. However, the problem is that arteries and veins are hardly differentiated by the application of automated classification algorithms also is burden by the fact that ratio changes in eye fundus images occur at the micro level. To overcome this issue the most of the image analysis algorithms are applied to high quality images captured by desktop fundus cameras, however, those are very expensive and can hardly be used under field conditions when the impartial evaluation of the investigative physical or mental state is needed. To that extent the mobile eye fundus cameras can be used.

The authors present the method for automated measurement of blood vessel diameter change at a specified location on both: left and right eye fundus images before and after the physical load. Then the comparison of the obtained results with those made by the expert is discussed. For this investigation Optomed OY digital mobile eye fundus camera Smartscope M5 PRO was used.

Digital Theology: A Computer Scientist's View

E. Sutinen

Professor Erkki Sutinen
Interaction Design Group
Department of Information Technology
University of Turku, Finland
Erkki.Sutinen@cs.joensuu.fi

Brief seminar on Digital Theology. It will be a good opportunity to hear about one of the more unusual interdisciplinary research areas involving advanced computing and digital technologies. Professor Sutinen's talk will discuss the use and development of digital tools such as human language technologies and analytics for solving theological problems and inter-faith dialogue to serve as instruments of peace.

Investigation of Background Extraction Techniques to Novelty Estimation in Biomedical Imaging

E. Šabanovič¹, D. Matuzevičius¹, A. Serackis¹,
V. V. Borutinskaitė², D. Navakas¹, R. Navakauskienė²

¹ Department of Electronic Systems
Vilnius Gediminas Technical University
² Department of Molecular Cell Biology
Institute of Biochemistry, Vilnius University
Arturas.Serackis@vgtu.lt

In our investigation, we have evaluated the influence of two different background subtraction techniques to the results of quantitative analysis of same proteins emphasized in gel using different staining techniques. Using 2DE gel image background extraction the variation of the estimated protein quantity increases more than 50% for 12%–33% of protein spots and more than 100% for 4.3%–12% protein spots.

Evaluation of Speaker's Individual Qualities Using High Order Autoregressive Model

G. Tamulevičius¹, J. Kaukėnas¹, A. Rybakovas²

¹ Institute of Mathematics and Informatics
Vilnius University
² Hospital Santariskiu Klinikos
Vilnius University
Gintautas.Tamulevicius@mii.vu.lt

Various models can be applied for speech signal description: sinusoidal, fractal, autoregressive-moving average, autoregressive. Autoregressive (AR) model employs statistical properties of the speech signal and describes the signal as the weighted sum of past signal values with added white noise. The AR model and its derivative forms (generally known as Linear predictive coding analysis) were widely used for speech analysis and synthesis: speaker verification, speech recognition, coding, speech enhancement and restoration, voice activity detection.

Most of researchers (Lithuanian researchers among them also) have been exploiting 10-30th order AR models considering them as sufficient to characterize spectral properties of the speech signal and. Our experimental results revealed the need of higher order AR models for

adequate modeling of Lithuanian phonemes – the proper model order value varied up to 100th.

The proposed AR model order and parameter estimation technique enabled us to employ high order (up to 200th) AR models for speech signal modeling. Analysis of various persons natural, pathological, singing and emotional speech recordings has shown a high power of frequency discrimination of the AR model. High order AR model describes individually fine properties of speaker's vocal system: twofold pitch, variability of the pitch, presence of high frequencies and others. Capability of such detailed spectral analysis and discrimination between speakers' qualities provides potentials for following speech analysis applications: evaluation of vocal fold functional status for clinical purposes, speaker verification and identification, speech recognition, evaluation of speaker's emotional state.

Classification-Based Storage of JPEG Images

J. Tichonov¹, O. Kurasova¹, E. Filatovas²

¹ Institute of Mathematics and Informatics
Vilnius University

² Faculty of Fundamental Sciences
Vilnius Gediminas Technical University
Jevgenij.Tichonov@mii.vu.lt

Currently, various devices that take high-quality images are widely used, and the produced images are usually compressed by lossy algorithms in order to save storage space. In this investigation, we address the problem of image classification before compression. The goal of our research is to classify images before using JPEG algorithm to set the proper value of Quality Factor (QF) for each image.

An image compression approach has been proposed, where images are assigned to one of three classes before their compression by a JPEG algorithm. The quality of images after compression is estimated by a structural similarity (SSIM) index method and Peak Signal-to-Noise Ratio (PSNR). The experimental investigation has shown that the proposed approach allows saving storage space compared to a conventional JPEG algorithm. Moreover, it has been determined which features describing the original images have to be used in the image classification into three classes with high, middle and low quality after compressing them by the JPEG algorithm.

Identification of Clusters in a Trial Group of Persons Based on Their Cardiac Intervals and Magnetic Field Data

I. Timofejeva¹, R. Joffe², A. Vainoras², M. Ragulskis¹

¹ Department of Mathematical Modelling
Kaunas University of Technology

² Human Health Investigation Institute
Lithuanian University of Health Sciences
Minvydas.Ragulskis@ktu.lt

Data on heart interbeat intervals for a group of 20 persons wearing cardiac monitors for two weeks are fused together with data generated by a local magnetic field magnetometer. The algorithm for the identification of slow dynamics for every person is constructed in order to provide meaningful interrelationships between the persons and the magnetic field. Near-optimal chaotic attractor embedding techniques are used to identify specific patterns in interpersonal dynamics. The revealed clusters are validated by data from psychological questionnaires.

Computer Screening and Modelling for Anti HIV-1 Drug Development

A. V. Tuzikov², I. A. Kashyn^{1,2}, Y. V. Kornoushenko¹,
A. M. Andrianov¹

¹ Institute of Bioorganic Chemistry, National Academy of Sciences of Belarus, Minsk, Belarus

² United Institute of Informatics Problems, National Academy of Sciences of Belarus, Minsk, Belarus

Tuzikov@newman.bas-net.by

Developing new drugs against HIV is an area of intensive research. From more than 30 drugs approved today for use in clinical medicine for the treatment of HIV-1, only two could prevent the penetration of the virus into the target cell. The rest can act on the later stages of the virus replication cycle when it has already penetrated into the cell. About 10 years ago there were discovered antibodies that can neutralize various strains of HIV-1. The antibodies VRC01 and 10E8, for example, can neutralize more than 90% of viral isolates. The report is devoted to the presentation of computer technology that allows to find small chemical

compounds that can interact with HIV-1 envelope proteins similarly to broad neutralizing antibodies (3074, VRC01, 10E8). The technology contains database to search for finding potential chemical compounds, molecular docking and dynamics for evaluation of their inhibitory activity. It takes a lot of computational resources to implement the above simulation stages. The choice of the antibodies was not accidental. The antibody VRC01 blocks the first stage of HIV entry into a target cell, i.e. the virus binding to the primary CD4 receptor of the target cells. The antibody 3074 blocks the second step – the binding to co-receptors CCR5 or CXCR4. Finally, antibody 10E8 targets virus envelope protein gp41, which is responsible for the fusion of the virus membrane with the host cell membrane. We propose 18 compounds that can be considered as promising basic structures for the rational design of novel, potent, and broad-spectrum anti-HIV-1 drugs.

An Exploratory Analysis of the Relation Between Metabolic Syndrome Factors and MicroRNA Data

R. Užupytė^{1,2}, T. Krilavičius^{1,3}, R. Navickas², A. Šatrauskienė²

¹ Baltic Institute of Advanced Technology

² Vilnius University

³ Vytautas Magnus University

Ruta.Uzupyte@bpti.lt

The metabolic syndrome (MetS) is associated with increased risk for cardiovascular disease. The detection and treatment of the underlying factors of the metabolic syndrome have a significant influence on the reduction of the cardiovascular disease. In this study, we analyze relations between MetS components and RNA molecules (microRNAs) regulating gene expression at the posttranscriptional level, in order to determine the predictive value of different microRNAs for subjects with metabolic syndrome. We apply correlation and linear regression to analyze the relationship between microRNAs and selected arterial markers. Logistic regression models were used to explore the statistical relationship between microRNAs and categorical variables. Results show that statistically significant linear relationship exists between arterial markers and several microRNAs, however, the observed relationship is very weak (<0.25). Since cardiovascular diseases are usually multifactorial diseases, caused by various mechanisms, it is more likely, that the combination of microRNAs will have stronger predictonal or diagnostic power.

Moreover, it is possible that more valuable results can be obtained by analyzing relations between microRNAs and binary variable determining the absence/existence of metabolic syndrome. Hence, we plan to use canonical correlation analysis to investigate linear combinations of microRNAs which have a maximum correlation with arterial markers.

Fusing Various Audio Feature Sets for Detection of Parkinson's Disease from Sustained Voice and Speech Recordings

E. Vaičiukynas^{1,2}, A. Verikas^{1,3}, A. Gelžinis¹, M. Bačauskienė¹, K. Vaškevičius¹, V. Uloza⁴, E. Padervinskis⁴, J. Čičelienė⁵

¹ Department of Electrical Power Systems
Kaunas University of Technology

² Department of Information Systems
Kaunas University of Technology

³ Centre for Applied Intelligent Systems Research
Halmstad University, Sweden

⁴ Department of Otolaryngology
Lithuanian University of Health Sciences

⁵ Department of Neurology
Lithuanian University of Health Sciences
Evaldas.Vaiciukynas@ktu.lt

This study explores the suitability of voice and speech recordings for the task of Parkinson's disease detection. Voice modality corresponds to a sustained phonation /a/ and speech modality to a short sentence in the Lithuanian language. Recordings were done through two channels simultaneously, namely, basic and phone microphones. Diverse information from the audio signal is extracted by 22 well-known audio feature sets. Random forest (RF) is used as a learner, both for individual feature sets and for decision-level fusion. Detection is evaluated by the out-of-bag equal error rate (EER) and the cost of log-likelihood-ratio (Cllr) measures, conforming to the leave-one-subject-out validation scheme. Essential descriptors were found as the best individual feature set for the basic channel, achieving EER of 16.3% for voice and 13.3% for speech. The best feature set for the phone channel was Tsanas, achieving EER of 29.5% for voice, and jAudio, achieving EER of 26.7% for speech. Fusion of feature sets and modalities improved the detection performance and resulted in EER of 10.6% for the basic and EER of 23.9% for the

phone channel. Variable importance in fusion revealed speech modality as more important than voice. Non-linear mapping of an RF proximity matrix into 2D space by t-distributed stochastic neighbour embedding (t-SNE) enables visualization of multivariate data for medical decision support.

Recurrent Estimation of Homogeneous Hidden Markov Model Parameters

J. Vaičiulytė

Institute of Mathematics and Informatics
Vilnius University
Jurate.Vaiciulyte@mii.vu.lt

Hidden Markov models are very important for analysis of signals and systems due to the ability to learn model parameters from an observation sequence. We analyze a homogeneous Hidden Markov Model determined by its state transition probabilities, its initial state distribution, observation sequence, and observation probability density. The observation sequence is represented by a single Gaussian distribution. Baum-Welch and EM algorithms were adapted to estimate Hidden Markov Model parameters in the recurrent way. The recurrent algorithm has been studied by computer simulation. Application to speech recognition is discussed as well.

Computerized High Frequency Trading of Nanoseconds in Futures Market

M. Vaitonis, S. Masteika

Kaunas Faculty of Humanities
Vilniaus University
Mantas.Vaitonis@khf.stud.vu.lt

Pair trading is a popular strategy where a profit arises from pricing inefficiencies between stocks. The idea is simple: find two stocks that move together and take long/short positions when they diverge abnormally, hoping that the prices will converge in the future. Recent computational expansion in financial modelling and the ever increasing demand for order execution speed is moving market making and price discovery strategies into high frequency trading (HFT) or the milliseconds

realm, where HFT already generates nearly 2/3 of the overall trading volume. During the last few years high frequency trading in milliseconds or nanoseconds has drawn attention of not only to financial players but and to researchers and engineers. The main objective of this research is to check two different statistical arbitrage strategies using nanosecond data of five different future contracts from NYMEX from 2015-08-01 till 2015-08-31. High frequency trading algorithms trade in high speed, thus it is important to receive and send information as fast as possible. Currently nanosecond information is the fastest data provided to traders. These five future contracts are BZ (Brent last day financial futures), CL (crude oil futures), HO (NY harbor ULSD futures), NG (Henry Hub natural gas futures) and RB (RBOB gasoline futures). One strategy used in this paper was first implemented by M.S Perlini, the other one by J.F. Caldeira and G.V. Moura. Strategies were back tested applying MatLab software of technical computing. Together with the strategies a pair selecting algorithm was presented. All three strategies were modified in order to be able to work with high frequency data. At the end of the research strategies were measured accordingly to the profit they did generate. The applied strategies with pair selecting algorithm can be interesting to financial engineers, market microstructure developers or market participants implementing low latency trading strategies.

Modeling Environment to Maintain Interoperability of Enterprise Applications

A. Valatavičius, S. Gudas

Institute of Mathematics and Informatics
Vilnius University

Valatavicius.Andrius@gmail.com

Integration and interoperability of applications in a dynamic enterprise environment is a complex issue. In practice integration projects are expensive and mostly tend to fail. Most researchers are focused on application integration in a limited or a very specifically defined environment and do not cover dynamic business behavior. In practice the integration solutions need constant updates and fixes to keep on working. This paper presents the internal modeling paradigm based approach to dynamic business environment modeling aimed to ensure the interoperability of applications. Our approach is based on the adapted principals of the internal model control theory. The enterprise architecture

models are explored for autonomic or semi-autonomic integration solutions. ArchiMate enterprise modeling language is used to model and represent such integration scenarios in the dynamic enterprise environment. The self-aware integration component is introduced which is designed to monitor the state of other enterprise applications and ensure the interoperability. The ArchiMate motivation layer and business layer have been modified by introducing feedback loops that are necessary and required in the dynamic business environment. Application layer meta-model is modified to support modeling of applications interoperability. Finally, the application layer autonomic integration solution model is presented, which includes models from the dynamic business environment. These models are modified using principle of internal model-based control system. This modification of application integration process ensures interoperability in a dynamic environment. Further work will be focused to develop detailed structures and to compare with agent-based integration solutions in a simulated dynamic business environment.

Goal-Oriented Dynamic Business Process Simulation: Thoughts on Implementation

O. Vasilecas, T. Vysockis, T. Rusinaitė

Information Systems Department
Faculty of Fundamental Sciences
Vilnius Gediminas Technical University
Tadas.Vysockis@vgtu.lt

Traditional business process (BP) simulation approaches consider BP as a pre-defined sequence of activities, limiting possibilities to adapt a process model to changes occurring in the business environment. To be able to use dynamic BP in simulation we need an approach, which enables us to use goal-oriented BP in simulation. Possibility to define goal is one of the key element in dynamic BP simulation, which allows reaching planned BP goal without using predefined sequence of activities in simulation model. Goal of the system can be described as purposed system state. In addition, to be able to measure if goal is reached it should be defined as a set of variables and set of functions dependent on these variables. In this paper, we present concept of a goal, goal-oriented approach and goal-oriented BP modelling. In addition, we presented the general way to describe simulation state and by doing that — describe the simulation goal. The article provides explanations of how simulation state

is generally described and how activity selection process should choose most suitable activity based on historical data and evaluated activity benefit for reaching the goal.

Knowledge-Based UML Dynamic Models Generation Method Usage in IS Lifecycle Design Stage

I. Veitaitė, A. Lopata

Kaunas Faculty of Humanities

Vilnius University

`Ilona.Veitaite@khf.vu.lt`

Business and IT alignment have kept as a significant management concern for over two decades because it remains an essential goal. Superior strategic alignment between business and IT strategy should lead to superior performance in comparison with lower stages of strategic alignment. Strategic business and IT alignment affect business performance and IT effectiveness.

In today's organizations exists a big gap between business and IT fields. Information technology strategy planning is a multistage process and information system development depends on the implementation of each stage of IS lifecycle. Combination of main business goals and IS development goals do not have direct relation. To ensure business and IT alignment it is important to create communication between these two sides.

Traditional IS engineering stages from user requirements specification to project development are empirically based, and traditional CASE tools implement empirical UML based IS project development process. One of the many computerized IS engineering methods' disadvantages is the fact that IS design models are generated only partially, since in the design stage, the designer forms design models based on personal experience, rather than applying the principles of knowledge generation which are stored in the enterprise model. Solution to these problems is integration of enterprise model, as the core subject knowledge repository. IS engineering process is ensured with knowledge-based subsystem as CASE tool component with enterprise meta-model and enterprise model inside. The method of UML models generation from enterprise model implements a knowledge-based design stage in the IS development cycle.

Enterprise meta-model is a formal structure which ensures more qualified project development process and knowledge base data collection. Enterprise model and enterprise meta-model make UML project models generation process more effective and qualified and ensure lower number of mistakes in the final IS development stage.

The goal of this work is to extend possibilities for generating UML dynamic models in an understandable way for business participants, via developing transformations from enterprise model to UML dynamic models. Knowledge-based generation method combines main principles of knowledge-based methods, ISO standards, and MOF architecture. This method gives the possibility to create advanced software development approaches. Within the research, the existing enterprise meta-model was supplemented with new elements and relations. In order to implement transformation for UML dynamic model generation from enterprise model, new models transformation algorithms were created.

The main practical significance of the research is the step towards making modelling process better understandable and usable for all participants of IS development life cycle. Application of the knowledge-based generation method provides the opportunity to use supplementary models validation methods that enterprise meta-model defines.

Research Action designated as COST Action CA15123: The European research network on types for programming and verification (EUTYPES).

Modified SOM for Abnormal Marine Traffic Detection

J. Venskus¹, M. Kurmis², P. Treigys¹

¹ Institute of Mathematics and Informatics
Vilnius University

² Klaipėda University

Julius.Venskus@mii.vu.lt

In recent years growing marine traffic around seaport surroundings rise traffic and security control problems and increase the workload for traffic control operators. Automated Identification System of vessel movement provides data that must be analyzed by the operators. However, increasing marine traffic increases the amount of data as well, thus for the traffic control operators decision support system has to be developed that is able to detect abnormal traffic patterns.

The paper presents a new self-learning adaptive algorithm for maritime traffic abnormal movement detection. The algorithm is based on the combination of self-organizing map and virtual pheromone method for traffic abnormality evaluation and the Mexican hat wavelet was used as neighborhood function. The proposed approach provides rapid self-learning and fast adaptation characteristics.

The accuracy of the proposed algorithm has been evaluated by two different approaches while introducing different learning factors. The dataset for verification of the proposed algorithm was provided by the Klaipeda seaport from marine traffic Automated Identification System.

Intelligent Decision Making and Consensus

E. Herrera-Viedma

University of Granada, Spain

Viedma@decsai.ugr.es

We analyze the fuzzy decision-making models as to develop intelligent decision-making processes in the real world. In particular, we focus on the soft consensus models by analyzing an overview of consensus models based on soft consensus measures, showing the pioneering and prominent papers, the main existing approaches and the new trends and challenges.

Semi-Supervised Learning with Generalized Eigenvalues

M. Viola¹, M. Sangiovanni², G. Toraldo³, M. R. Guarracino²

¹ University of Rome "La Sapienza", Italy

² National Research Council, Italy

³ University of Naples "Federico II", Italy

Mario.Guarracino@cnr.it

Supervised classification is a powerful data analysis technique. It uses prior knowledge on class membership (labels) of data samples. Since the cost associated to the labelling process can be both expensive and time-consuming, it is interesting to investigate semi-supervised algorithms that produce classification models taking advantage of relatively small sets of labelled data. In this talk we introduce LapReGEC, a novel technique that introduces a Laplacian regularization term in a generalized eigenvalue classifier. As a result we produce models that are both accurate and

parsimonious in terms of labelled data. We empirically prove that the obtained classifier well compares with other techniques, using as little as 5% of labelled points needed by a supervised method to compute models.

Aspects for Choosing Textual Programming Languages for High School Education

M. Visnovitz

Eötvös Loránd University, Budapest, Hungary
Vimtaai@inf.elte.hu

The implementation of algorithms is an integral part of teaching programming and data modeling in high school. For this, the teacher has to choose a programming language to use. Choosing the right language and development environment is of decisive importance for having an effective and useful teaching process. I collected a list of aspects of textual programming languages that we have to take into consideration when choosing. For languages commonly used in Hungarian high school education I examined how well they do in terms of these aspects. Based on the results I listed the benefits and drawbacks for different groups of these languages.

I also examined the latest changes in recent years and based on this I also examine the possibility of a new trend of using script languages for education. Among them, I deal with the educational aspects of the Python and JavaScript, also the typed version of JavaScript, TypeScript.

Some Geometrical and Analytical Features of Problems Involving Big Data and High Dimensions

A. Zhigljavsky

School of Mathematics, Cardiff University, UK
ZhigljavskyAA@cf.ac.uk

In this talk, I will concentrate on the following aspects of the science of big data science and large dimensions. First, I will outline some peculiar features of high-dimensional spaces. This will include a discussion on some counter-intuitive but very practically useful properties of the multivariate uniform and normal distributions. Second, I will discuss the so-called 'curse of dimensionality' in Monte-Carlo methods and global random search. I will try to demonstrate that it is impossible to guarantee

any acceptable accuracy for high-dimensional black-box global optimization problems in the absence of Lipschitz-type information about the objective function. Third, I will discuss the problem of finding the most influential predictors in linear regression with a very large number of explanatory variables. Finally, I will discuss some big-data practical problems; the main attention will be paid to the problem of adaptive targeting in online advertising and effective methods of feature selection.

Classification of Data Streams Using Nearest Neighbour Classifier and Normalized Compression Distance

K. Ząbkiewicz^{1, 2}

¹ Faculty of Economics and Informatics
University of Białystok, Poland

² Systems Research Institute, Polish Academy of Sciences
Kamil.Zabkiewicz@gmail.com

Development of computer technologies resulted in creating large collections of data. The medical tests, financial and weather forecasts, network activity etc. represent large data streams, often featureless or having a large number of them. Classification of this sort streams became a crucial problem.

In this work, we will present an approach for classifying streams based on nearest neighbour classifier and distance measure derived from data compression called normalised compression distance (NCD).

The main reason for choosing this approach is that the NCD is parameter free, feature-free, and alignment-free. There is no need to set various parameters, such as learning speed, the number of epochs, weights for features etc. We do not have to perform dimensionality reduction because all information is preserved using lossless data compression. Another reason for using this technique is the fact that modern data streams have repeating patterns that can be compressed. Moreover, if we use lossless compression, we are guaranteed that all compressed information can be restored in future. Using classical feature selection or extraction methods is not always the case, especially when the context of data can change, e.g. features important for classification at the moment may in future become less significant and vice versa. Our implemented approach accepts all kinds of data, it does not matter if it is numerical or categorical.

We will also present initial results on classifying data sets having a large number of features as well as data set from AAIA' 16 Data Mining Contest.

Multi-Level Method for Big Data Visualization

J. Zubova^{1,2}, O. Kurasova¹

¹ Institute of Mathematics and Informatics
Vilnius University

² Faculty of Electronics and Informatics
Vilniaus kolegija / University of Applied Sciences
Jelena.Zubova@mii.vu.lt

Big data analytics are widely used to examine large data sets to uncover hidden and useful information. An ability to access new data sources provides many opportunities, but it also raises new challenges related to capabilities to handle such enormous amounts of data. We need effective methods and powerful environments to complete large and complex tasks. This research focuses on big data visualization that is based on dimensionality reduction methods. Our goal is to find the most effective ways to analyse and visualize data of such type. In our approach, whole data mining process is divided into separate steps. In each step, individual dimensionality reduction and visualization method is applied according to data volume and type. The selection of methods is based on their speed and accuracy. Therefore, the comparison of dimensionality reduction methods is presented in this work. An idea of the proposed visualization approach is as follows. All the data are clustered and visualized on the surface of a sphere. Moreover, there is an ability to see the features of each data cluster. The further analysis is performed only for the selected cluster. At the initial stage, the accuracy is not so important, thus the fastest method can be used. In the following dimensionality reduction steps the more accurate but little bit slower methods are chosen. During each step, the selected cluster is divided into smaller subclusters. At the end, the data should be processed by the most accurate method. It would require too many resources at the beginning of dimensionality reduction, but at the end, the data set is reduced enough to be processed in the most accurate way.

Global Optimization Using DIRECT Algorithm and Modifications

J. Žilinskas, R. Paulavičius

Institute of Mathematics and Informatics

Vilnius University

`Julius.Zilinskas@mii.vu.lt`

The well known DIRECT algorithm for global optimization is based on dividing hyper-rectangles and evaluating objective function at the centers. In this talk we discuss various alternatives to improve performance of the algorithm including various subdivisions of subregions and different strategies for sampling of points where the objective function is evaluated. Experimental investigation is used to compare the modifications solving various test problems for global optimization.

8th International Workshop
DATA ANALYSIS METHODS FOR SOFTWARE SYSTEMS

Sudarytoja Jolita Bernatavičienė
Parengė spaudai ir išleido
Vilniaus universiteto Matematikos ir informatikos institutas,
Akademijos g. 4, Vilnius
Spausdino UAB „BMK leidykla“
J.Jasinskio g. 16, Vilnius

Main Sponsors



Sponsors

