



Lithuanian Computer Society
Vilnius University Institute of Mathematics and Informatics
Lithuanian Academy of Sciences



7th International Workshop

DATA ANALYSIS METHODS FOR SOFTWARE SYSTEMS

Druskininkai, Lithuania, Hotel „Europa Royale“

<http://www.mii.lt/DatAMSS>

December 3-5, 2015

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

BAIP
www.baip.lt

Algoritimų sistemos
www.algoritmusistemas.lt

Sponsors

NetCode
www.netcode.lt

Computer Science Center
www.inscience.lt

Insoft
www.insoft.lt

Baltic Amadeus
www.baltic-amadeus.lt

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 3

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

Yurij Kharin “Statistical Analysis of Time Series Based on Incomplete Discrete Data” (Belarusian State University)

Yaroslav D. Sergeyev “New Computational Technologies in Numerical Analysis” (University of Calabria, Italy)

14¹⁵–15⁰⁰ – Coffee, Poster Session and Discussions

1. *S. Gudus* “Normalization of Domain Modelling in Information Systems Development”
2. *L. Sakalauskas, A. Ušpurienė* “Activity Based Costing by Two Stage Stochastic Programming”
3. *R. Užupytė, T. Krilavičius, I. Žliobaitė, H. Simonavičius* “Respiratory Motion Compensation by Model-based Functional Target Tracking”
4. *A. Žilinskas, G. Gimbutienė* “On One-Step Worst-Case Optimal Trisection in Univariate Bi-Objective Lipschitz Optimization”
5. *M. Liutvinavičius, V. Sakalauskas, D. Krikščiūnienė* “Multicriteria Modelling of Financial Markets Efficiency”

6. **J. Zubova, O. Kurasova, V. Medvedev** “Visual Analytics for Big Data”
7. **L. Paliulionienė** “A Method of Comparing Business Rule Sets”
8. **M. Sabaliauskas, V. Marcinkevičius** “An Application of ARAP Method for Parametrization of Shoetrees”
9. **M. J. Bilinskas, G. Dzemyda, M. Trakymas** “Registration of Computed Tomography Scan Slices Using Mathematical Model for Ribs-bounded Contour Approximation”
10. **A. Rapečka, G. Dzemyda** “The Challenges for Recommendation Systems of a New Dreams Meanings Dataset”

15⁰⁰–16⁰⁰ – Plenary Session

Boštjan Brumen “Measuring and Comparing Performances of Classification Algorithms” (University of Maribor, Slovenia)

Vineta Arnicāne, Guntis Arnicāns, Juris Borzovs “A Draft Approach to Quality Evaluation of Terminology Databases” (University of Latvia)

16⁰⁰–17⁰⁰ – Coffee, Poster Session and Discussions

1. **V. Tiešis, A. Lančinskas, V. Marcinkevičius** “Presymptomatic Genetic Risk Assessment Methods”
2. **T. Leonavičienė, P. Baltrėnas** “Effects of the Parameters on the Air Flow in a Four-Channel Cyclone”
3. **R. Baronas, L. Litvinas, A. Žilinskas** “Multicriteria Optimisation of Biosensor Utilizing Synergistic Substrates Conversion”
4. **V. Jakštys, V. Marcinkevičius, P. Treigys** “Experimental Investigation of Chromatic Aberration Elimination in Digital Images of Eye Fundus Camera”

5. **G. Jankevičiūtė, R. Čiegis, M. Meilūnas, O. Suboč, A. Ilgevičius** “On Economic - Technological Optimization of High-Power Electric Cables”
6. **L. Dovydatis, V. Rudžionis** “GMM-UBM Enhancement with DNNs for Automated Speaker Recognition Systems”
7. **T. Pranckevičius, V. Marcinkevičius** “Classification and Visualization Algorithms on Cloud Computing: Issues and Advantages”
8. **J. Mandravickaitė, T. Krilavičius** “Visualization of Differences in Language Usage of Parliamentary Position and Opposition in the Lithuanian Parliament 2008-2012”
9. **A. Gimbutas** “Multicriteria Lipschitz Optimization Algorithm Using Local Lipschitz Constant Estimate”
10. **R. Karbauskaitė, G. Dzemyda** “Relation between the Number of Latent Variables and Degrees of Freedom of the Object Motion in the Image”
11. **A. Čaplinskas, J. Miliauskaitė** “Analysis of Non-functional Properties of an Ensemble of Collaborating Algorithms”

18⁰⁰–19³⁰ – Welcome Party

Friday, December 4

9⁰⁰–10⁰⁰ – Plenary Session

Ioan Dzitac “From Fuzzy Logic to Soft Computing: New Paradigms in Artificial Intelligence” (Agora University of Oradea, Romania)

Albertas Čaplinskas “Theoretical Research and Theory Building in Informatics” (Institute of Mathematics and Informatics, Vilnius university)

10⁰⁰–11⁰⁰ – Parallel Plenary Sessions

Section 1

1. **A. Kaliszewska, M. Syga** “Computerized Similarity Detection in Ceramic Assemblages Based on Morphological Properties”
2. **R. Čiegis, V. Starikovičius, M. Ragulskis, R. Palivonaitė** “Parallel Boinc Computing for Visual Cryptography Algorithms”
3. **L. Pavlič, M. Vaupotič, M. Heričko** “An Evaluation Model for Web-Service-Based Information System Integration”

Section 2

1. **M. Pavlinek, M. Heričko, V. Podgorelec** “Determining the Number of Topics in a Semi-Supervised Setting”
2. **T. Skersys, P. Danėnas, R. Butleris** “Model-driven Development and Application of Model-to-Model Transformations in a CASE Tool”
3. **M. Heričko, Č. Gerlec** “Software Defect Prediction Using Structural Source Code Changes”

11⁰⁰–11²⁰ – Coffee

11²⁰–12²⁰ – Parallel Plenary Sessions

Section 3

1. **R. Bieńkowski, K. Leśniewski** “Computer-Aided Reconstruction of Archaeological Settlement Pattern”
2. **A. Kriščiukaitis, R. Petrolis, R. Ramonaitė, R. Pečiulis, D. Jančiauskas, A. Kiškis, L. Kupčinskas** “Quantitative Evaluation of Morphologic Changes in Histologic Samples by Means of Mathematical Morphology”
3. **A. Jakaitienė, D. Stumbrienė, R. Želvys** “Construction of Education Monitoring Index”

Section 4

1. **E. Filatovas, O. Kurasova, J. L. Redondo, J. Fernández** “Solving Multi-Objective Competitive Facility Location Problems Using Preference-Based Evolutionary Approach”
2. **D. Podkopaev, A. Marques, A. Mäkinen, B. Oliveira, J. Rasinmäki** “Optimal Management of Forest-to-Biofuel Supply Chain: a Case Study”
3. **J. Žilinskas** “On Branch and Bound for Non-convex Multi-objective Optimization”

12³⁰–14⁰⁰ – Lunch

14⁰⁰–15⁰⁰ – Plenary Session

Ignacy Kaliszewski “Multiple Criteria Decision Making decision making support in a quest for a universal toolbox” (Systems Research Institute of the Polish Academy of Sciences, Poland)

Janis Grundspenkis “Concept Maps as Knowledge Assessment Tool: pro and cons” (Riga Technical University, Latvia)

15⁰⁰–17⁰⁰– Coffee, Poster Session and Discussions

1. **S. Minkevičius, E. Greičius** “On the Inequality in Open Queueing Networks”
2. **I. Veitaitė, A. Lopata** “Knowledge Based UML Dynamic Models Generation Method”
3. **O. Kurasova, T. Petkus, E. Filatovas, A. Igumenov, S. Ivanikovas** “Energy Efficient Computing in PC Clusters on Demand”
4. **J. Jucevicius, P. Treigys, R. Briedienė, I. Naruševičiūtė** “Computer-Aided Localization of Prostate in mpMRI”
5. **M. Vaitonis, S. Masteika** “High Frequency Statistical Arbitrage Strategy Engineering and Algorithm for Pairs Trading Selection”
6. **T. Petkus, O. Kurasova** “Analysis and Forecast of Teachers’ Demand in Lithuania”
7. **V. Mickevičius, T. Krilavičius, V. Morkevičius** “Automatic Thematic Classification of the Titles of the Seimas Votes”
8. **T. Liogienė, G. Tamulevičius** “Multistage Speech Emotion Recognition for Lithuanian: Experimental Study”
9. **J. Tichonov, O. Kurasova, E. Filatovas** “Image Classification for Quality Prediction after JPEG Compression”
10. **G. Ciganaitė, A. Mackutė-Varoneckienė, T. Krilavičius** “Text Document Clustering Using Language-Independent Techniques”
11. **R. Baušys, M. Espenko** “Image Segmentation by Neutrosophic Sets”
12. **O. Myronycheva, V. Vigont, A. Topaj, V. Denisov** “Hybrid Approach for Model Renovation: New Life of Classical Chanter-Thornley Model of Mushroom Cultivation”

13. **V. Kaminskas, E. Ščiglinskas** “Modelling and Control of Human Emotion Signals as Reactions to a Dynamic Virtual 3D Face”
14. **G. Stabingis, J. Bernatavičienė, G. Dzemyda, D. Imbrasienė, A. Paunksnis, V. Slavinskytė, P. Treigys** “Automated Blood Vessel Detection and Characteristic Evaluation in Eye Fundus Images”
15. **S. Konovalenko** “Methods and Means for Generalization of Information in Specialized Information Systems”
16. **V. Gudžiūnas, E. Paulėkas, G. Urniežiūtė, A. Vaitkus, V. Veikutis** “The Validity of Eye Blink Rate by Using Quantitative Parameters of Attention Level”

19⁰⁰–21⁰⁰ – Dinner

Saturday, December 5

9³⁰–11³⁰ – General Discussion

11³⁰–12⁰⁰ – Coffee, Poster Session and Discussions

12¹⁵–14⁰⁰ – General Discussion, Concluding Remarks

14⁰⁰–14³⁰ – Closing

A Draft Approach to Quality Evaluation of Terminology Databases

V. Arnicāne, G. Arnicāns, J. Borzovs

Faculty of Computing

University of Latvia

Juris.Borzovs@lu.lv

We present an approach to the evaluation of the quality of terminology data that caters for the differing levels of quality required of various parameters in order to meet different goals. The key parameters associated with particular quality classes and their acceptable range of values are obtained through interview with expert users. The approach described is currently being applied in an experimental evaluation of the quality of software testing terminology database developed by ISTQB (International Software Testing Qualifications Board).

Multicriteria Optimisation of Biosensor Utilizing Synergistic Substrates Conversion

R. Baronas¹, L. Litvinas¹, A. Žilinskas²

¹ Faculty of Mathematics and Informatics

Vilnius University

² Institute of Mathematics and Informatics

Vilnius University

Linus.Litvinas@gmail.com

Biosensors are relatively cheap and reliable devices capable to resolve a large number of analytical problems in clinical diagnostic, environmental and industrial fields. The biosensor response is determined by the catalytical substrate conversion to a reaction product. In a case of amperometric biosensor the response is a result of oksidation-reduction reaction on surface of electrode. The concentration of the analyte is proportional to the measured current and can be determined by the calibration curve. In the simplest Michaelis-Menten scheme the substrates are directly converted to the products. In a more complex synergistic scheme, an enzyme catalyzes the parallel conversion of substrates into the products, with the concomitant cross reactions of the substrates and the products. A biosensor must meet user requirements for a range of the analyte concentrations, a level of the saturation current and an enzyme amount. The measurement range and saturation current should be high, while the enzyme amount

should be low as enzyme usually is expensive. The biosensor utilizing synergistic substrates conversion involves multiple changeable parameters: the thickness of enzyme and dialysis membranes, the concentrations of the enzyme and two reaction mediators. By varying these parameters multicriteria Pareto optimal solutions can be achieved. A digital model based on a mathematical model of the biosensor facilitate this. It has been widely used to fine-tune the analytical characteristics of the biosensors. An optimisation algorithm uses weighted sum scalarization techniques to transform a multicriteria problem into the single criteria optimisation problem. The optimised function was assumed to be convex. Hooke-Jeeves single criteria optimisation algorithm was used due to fast convergence like gradient descent algorithms, but it requires no gradient of the optimised function. The achieved Pareto optimal results can be used by experts in a design phase of a practical biosensor utilizing synergistic substrates conversion.

Image Segmentation by Neutrosophic Sets

R. Baušys, M. Espenko

Faculty of Mathematics and Informatics
Vilnius Gediminas Technical University
Romualdas.Bausys@vgtu.lt

Image segmentation is an important task for image processing, analysis and object recognition in computer vision. The neutrosophic sets allow us to model explicitly the uncertain phenomena of the information. By logic of the neutrosophic sets image pixel information is represented by degree of the truth (T), a degree of the indeterminacy (I) and a degree of the falsity (F). Applying this approach, we deal with image information in three ways, which are mutually independent.

Computer-Aided Reconstruction of Archaeological Settlement Pattern

R. Bienkowski, K. Leśniewski

Institute of Archaeology and Ethnology
Polish Academy of Sciences
Raf.Bienkowski@gmail.com

This paper presents a GIS-based method of reconstructing settlement patterns in the past. We propose a mathematical model of selecting the places that are most likely to be the settlement location. The model combines the use of GIS databases and mathematical multi-criteria decision aid tools. The multi-criteria approach has been chosen for the above problem since it adequately reflects its complexity. The criteria are based on a general a priori knowledge of the settlement structure of a given region in a given time period. The criteria are divided in two main groups: geographical and anthropogenic. A solution method of approximate settlement location is proposed and is tested on archaeological data. The output information on the approximate settlement location in the area might impact the choice of places for archaeological excavations.

For a case study we choose the island of Crete under the Venetian rule (1204-1669). In our opinion it is suitable for such an experiment as the settlement pattern is regular and there are clear criteria for location of new sites.

Acknowledgment

The result present on the paper were obtained during the internship at the Systems Research Institute of the Polish Academy of Science, financially supported by the project "Information technologies: research and their interdisciplinary applications"(PO KL 04.01.01-00-051/10-00).

Registration of Computed Tomography Scan Slices Using Mathematical Model for Ribs-bounded Contour Approximation

M. J. Bilinskas¹, G. Dzemyda¹, M. Trakymas²

¹Institute of Mathematics and Informatics
Vilnius University

²National Cancer Institute
Mykolas.Bilinskas@gmail.com

In this paper, a method for registering transversal plane images from computer tomography (CT) scan is presented. CT scans are 3D images – a collection of 2D images (slices), representing slices by transversal plane. The method consists of two steps: firstly, a mathematical model that describes the ribs-bounded contour is approximated, then this model is used to compare slices from different CT scans. The problem of approximation is solved by finding out the optimal parameters of the mathematical model using least-squares. The quasi-newton method is used for optimization. In this paper, we are restricting the slices to ones where ribs are visible, but this does not lessen the significance of our work, because many important internal organs are located here: liver, heart, stomach, pancreas, lung, etc. The research revealed two important problems: (1) how to fit two models; (2) how to consider the distribution of bone tissue pixels when registering two models. The first needs approximating y value in the model at given x. The latter one arises when different configurations of bone tissue pixels are approximated by similar models.

Measuring and Comparing Performances of Algorithms

B. Brumen

Faculty of electrical engineering and computer science
University of Maribor
Bostjan.Brumen@um.si

Computer science is largely an experimental science. Looking at an algorithm does not tell anymore how it will behave – at least without rigid formal specifications based on set theory and predicate logic. No longer can we easily relate input to the output. Repetitions are required, of the code, and more importantly, of the problems in samples, to demonstrate

the programs have expected results. At least statistically. Studying computer programs comprises of six components: agent (program), task, environment, protocol, data collection and analysis. The first three components are a part of behaviour observation; the last three are a part of empirical methods. Often though, one forgets about the complexity and rigidity of the experimental design and jumps to the conclusions about the algorithm's behaviour (superiority) without a rigid proof.

The tutorial will address the experimental design of measuring the performance of programs or algorithms. Bootstrapping and cross-validation method for data gathering are covered in detail and the differences are discussed.

The tutorial aims at informing the audience on how to design the experiment when measuring and reporting the performances of an algorithm.

Text Document Clustering Using Language-Independent Techniques

G. Ciganaitė¹, A. Mackutė-Varoneckienė¹, T. Krilavičius^{1,2}

¹ Vytautas Magnus University

² Baltic Institute of Advanced Technology

gciganait@gmail.com

Clustering is a technique for grouping objects by their similarity. Document clustering is used for topic extraction, filtering and fast information retrieval. However, due to the high dimensionality, clustering of documents is rather slow and computationally intensive. In case of highly inflective languages, such as Lithuanian, it becomes even problematic. We investigate language-independent document clustering for Lithuanian and Azeri languages. Bag-of-words (BOW) is used for documents representation. We propose four feature selection models based on the terms frequencies in the corpora. The best results have been achieved by the model where features which occur less than α_{min} times (or more than α_{max} times) in the whole corpora are eliminated from feature set as non-informative. The importance of features in defined feature subset is evaluated by the weights of term frequency-inverse document frequency (TF-IDF). Results show that it is enough to use only 2% – 5.6% of features of initial feature set to get the best clustering results. Hierarchical and flat clustering algorithms based on documents similarity were applied and precision of results was evaluated. Cosine distance was selected as the best distance measure. Many experiments with well-known Euclidean

distance were made, but this measure is inappropriate due to the sparsity of the feature matrix. Best clustering results were reached by using spherical k-means algorithm (F-score value approx. 0.8 for both languages).

Theoretical Research and Theory Building in Informatics

A. Čaplinskas

Institute of Mathematics and Informatics

Vilnius University

`Albertas.Caplinskas@mii.vu.lt`

Despite the fact that modern informatics as a scientific discipline emerged more than half century ago, it is still not enough mature. Most of the theoretical results still are rather at the exploratory level, many researchers still are not enough knowledgeable neither with the theoretical research methods commonly accepted in other more mature scientific disciplines nor with the modern experimentation techniques and threat evaluation approaches. The presentation discusses the role of theoretical research in informatics, relations between theoretical and empirical research, the concept of scientific theory, structure and other formal properties of scientific theories, classification of scientific theories, theory building approaches, relations between scientific theories and theoretical research methods, criteria to evaluate scientific theories, the concept of research programmes proposed by Imre Lakatos and its critic given by Paul Feysabend.

Analysis of Non-functional Properties of an Ensemble of Collaborating Algorithms

A. Čaplinskas, J. Miliauskaitė

Institute of Mathematics and Informatics

Vilnius University

`Jolanta.Miliauskaite@mii.vu.lt`

Compositional properties of computer programs, components and even software services are investigated already many years. Significantly less attention was paid to the investigation of compositional properties of algorithms, especially to investigation of non-functional properties of algorithms composition. However, this is a significant topic in many research

areas, including enterprise management systems, quality evaluation systems and even in optimization problems. The presentation introduces the concept of collaborating algorithms, investigates the most important patterns of collaboration and their modelling using UML. It analyses also algorithmic and other properties of such ensembles. Special attention is paid to the analysis of non-functional properties.

Parallel Boinc Computing for Visual Cryptography Algorithms

R. Čiegis, V. Starikovičius, M. Ragulskis, R. Palivonaitė

Department of Mathematical Modelling
Faculty of Mathematics and Informatics
Vilnius Gediminas Technical University
rc@vgtu.lt

The recent activities to construct exascale and ultrascale distributed computational systems are opening a possibility to apply parallel and distributed computing techniques for applied problems which previously were considered as not solvable with the standard computational resources. In this paper we consider one global optimization problem where a set of feasible solutions is discrete and very large. There is no possibility to apply some apriori estimation techniques to exclude an essential part of these elements from the computational analysis, e.g. applying branch and bound type methods. Thus a full search is required in order to solve such global optimization problems.

The considered problem describes visual cryptography algorithms. The main goal is to find optimal perfect gratings, which can guarantee high quality and security of the visual cryptography method.

The full search parallel algorithm is based on master-slave paradigm.

We present a library of C++ templates that allow the developer to implement parallel master-slave algorithms for his application without any parallel programming and knowledge of parallel programming API.

These templates automatically give parallel solvers tailored for clusters of computers using MPI API and distributed computing applications using BOINC API.

Results of some computational experiments are presented.

GMM-UBM Enhancement with DNNs for Automated Speaker Recognition Systems

L. Dovydaitis, V. Rudzionis

Kaunas Faculty of Humanities

Vilnius University

Laurynas.Dovydaitis@khf.vu.lt

In the past decade speaker recognition systems widely used Gaussian mixture model (GMM) and Universal background modelling (UBM) for speaker recognition. This method requires to collect fine-tuned features by expert to further form speaker model which is then processed with GMM-UBM and compared to test features for scoring to get the decision result. Increased computational power and most recent studies suggest use of deep neural networks (DNN) for unsupervised feature extraction and acoustic modelling. This allows high performance gains for many speech recognition tasks. DNN shows promising results and it could be expected to achieve better speaker recognition results comparing with other methods. We present our findings by summarizing current studies which suggest combination of traditional speaker recognition methods with DNNs and propose future work for increased accuracy and efficiency in automated speaker recognition systems.

From Fuzzy Logic to Soft Computing: New Paradigms in Artificial Intelligence

I. Dzitac^{1,2}

¹Agora University of Oradea

Romania, 410526 Oradea, Piata Tineretului, 8

Rector@univagora.ro

²Aurel Vlaicu University of Arad

Romania, 310330 Arad, Elena Dragoi, 2\

Ioan.Dzitac@uav.ro

In the near future Artificial Intelligence (AI) will surpass human intelligence in more and more domains. Indeed, based on soft computing, fuzzy control, bio-inspired computing, computational theory of perceptions and computation in natural language, Artificial Intelligent computers can write their own programs as they encounter situations and try different

ways to achieve a goal. Artificial Intelligent machines (net-centric automobiles, intelligent aircraft, intelligent home utilities, intelligent learn labs, entertainment devices, military defence arms, health applications), will be a commonplace. Humans have a remarkable capability to reason and make decisions in an environment of imprecision, uncertainty and partiality of knowledge, truth and class membership. It is this capability that is needed to achieve human-level machine intelligence. Achievement of human-level machine intelligence is beyond the reach of existing AI techniques and more of these are based on fuzzy sets theory and fuzzy logic. In this paper we will present a summary of AI problems and a survey of some new trends and new directions in Soft Computing and AI research.

Solving Multi-Objective Competitive Facility Location Problems Using Preference-Based Evolutionary Approach

E. Filatovas¹, O. Kurasova¹, J. L. Redondo², J. Fernández³

¹ Institute of Mathematics and Informatics
Vilnius University

² Department of Informatics
University of Almeria, Spain

³ Department of Statistics and Operations Research
University of Murcia, Spain

`Ernestas.Filatovas@mif.vu.lt`

Continuous competitive facility location and design problems with two and three conflicting objectives are considered in this research. Several evolutionary multi-objective optimization algorithms have been applied for obtaining a discrete set approximating the complete Pareto front. However, such evolutionary algorithms are computationally expensive and time consuming. Additionally, only a reasonable number of solutions should be provided to a Decision Maker (DM) so that he/she make decision an adequate decision avoiding the usually complex analysis of a large amount of information. Therefore, optimization algorithms based on DM's preferences with the aim to approximate only the part of the Pareto should be used.

In this research, we apply a modification of preference-based evolutionary approach R-NSGA-II for two and three objective facility location and design problems. The applied method enables to find a suitable solution

for the DM in a reasonable amount of time. The experimental comparison of the preference-based evolutionary approach with widely-used multi-objective evolutionary algorithms has shown the promising results.

Multicriteria Lipschitz Optimization Algorithm Using Local Lipschitz Constant Estimate

A. Gimbutas

Institute of Mathematics and Informatics
Vilnius University
Albertasgim@gmail.com

In this paper a new deterministic global optimization algorithm to determine Pareto Front of multicriteria black-box Lipschitz functions with unknown Lipschitz constants is proposed. At the beginning of the proposed algorithm feasible region is divided into simplexes. At each iteration of the algorithm a local Lipschitz constant estimate is found and lowest function value is estimated for each criteria in each simplex; Tolerance is calculated for each simplex to determine simplexes' potential; most promising simplices are selected and divided. Experiments were performed with two- and three- dimensional optimization problems using 400 bicriterial test problems generated with the GKLS test function generator.

Normalization of Domain Modelling in Information Systems Development

S. Gudas

Institute of Mathematics and Informatics
Vilnius University
Saulius.Gudas@khf.vu.lt

Concept of normalization (normalisation) is widely known in science and technology areas. Normalization has become traditional in database design theory and practice. The paper is focused on a knowledge-driven approach towards normalization of a set of business process models. Common knowledge content required to perform normalization are revealed by summarising normalization and functional dependency concepts in diverse subject domains.

Normalization procedure of enterprise model is based on the concept of management functional dependency (MFD). Management functional dependency (MFD) expresses the idea of clustering of subject domain into a set of self-managed complex activities - management transactions (F x P) between management functions F and enterprise processes P. To find a MFD in the empirical models of subject domain two frameworks are introduced: Detailed Value Chain Model (DVCM) and Elementary Management Cycle (EMC). Finally, definitions of enterprise model normal forms are presented.

The Validity of Eye Blink Rate by Using Quantitative Parameters of Attention Level

V. Gudžiūnas¹, E. Paulėkas¹, G. Urniežiūtė¹, A. Vaitkus²,
V. Veikutis³

¹ Academy of Medicine
Lithuanian University of Health Sciences

² Department of Neurology
Lithuanian University of Health Sciences

³ Institute of Cardiology
Lithuanian University of Health Sciences
Vincentas.Veikutis@lsmuni.lt

The aim of the study was to validate the eye blink rate by using quantitative parameters of attention level. We evaluated the eye blink rate in healthy adults (permission No.BEC-MF-35) free of disability, medical disease or neurological deficit. They had to do an increasing composition (matrix) of the numbers from 0 to 99, which were randomly spread over the table. During the task, EEG activity, attention level and an eye blink were recorded. The meaning of attention is presented in a scale from 0 to 100. The attention level was categorized into five levels which were: 1-19 (very poor attention), 20-39 (poor attention), 40-59 (neutral), 60-79 (good attention) and 80-100 (great attention).

We selected the intervals, where the average value of attention level was reduced and increased by 25 percent of average attention value. In these attention levels the eye blink rate per minute was evaluated. Also, we divided the task duration into three equal sections and evaluated the eye blink rate in each section without considering standardized attention level. These results were used as a control value and were analyzed by considering the smallest value of mean and standard deviation.

The average age of subjects (10 women and 4 men) was 23.4 ± 0.9 , average mean of attention – 56.6 ± 7.1 ; the duration of the task – 8.7 ± 2 minutes. After the analysis of the attention we divided three equal duration sections: in the first section the average mean of attention was 58.9 ± 7.9 , in the second – 57.3 ± 7.8 , and in the third section – 53.7 ± 8.8 . These results indicate that the given task (matrix) was homogenous.

The main part of time during the task was spent in higher attention levels. In good attention level (60–79) the average time in minutes was 2.7 ± 1 and in great attention level (80–100) – 1.1 ± 0.8 minutes.

A very poor attention level (1–19) was eliminated because of short duration, the average time in this level was only 0.2 ± 0.4 minutes.

General eye blink rate per minute was 4.5 ± 2.2 . Similar eye blink rate per minute remained in neutral attention level (40–59) – 3.9 ± 2.2 and in good attention level (60–79) – 4.6 ± 2.2 eye blinks per minute.

The smallest value of standard deviation was calculated in 40–59 and 60–79 attention levels, it means that in these attention levels the eye blink rate is standardized.

Our preliminary results indicate that the eye blink rate has standardized in neutral (40–59) and good (60–79) attention levels.

Concept Maps as Knowledge Assessment Tool: pro and cons

J. Grundspenkis

Department of Artificial Intelligence and Systems Engineering
Faculty of Computer Science and Information Technology
Riga Technical University, Latvia
`Janis.Grundspenkis@rtu.lv`

Concept maps (CM) nowadays are rather popular for learning and knowledge assessment/self-assessment purposes. Intelligent tutoring systems based on CMs have proved their usability for adaptive and student centred individualized learning. At the same time there are known only a few powerful CM based knowledge assessment tools, namely COMPASS and IKAS. Both provides assessment of the learners' knowledge level through various concept mapping tasks and supports the learning process generating the informative tutoring feedback after analysis and evaluation of learners' CMs. The advantages of CMs are the following: 1) CMs promote systems thinking because concept interrelatedness is an essential property of knowledge; 2) the top most quality of CMs is representation

of knowledge structure of each individual learner; 3) there are a wide variety of high-directed (fill-in-the-map) and low-directed (construct-the-map) tasks which may be defined at different levels of difficulty; 4) CMs as mathematical objects may be processed and analyzed using computers; 5) CMs promote a process oriented learning, regular knowledge assessment and student support – help and feedback. At the same time there are discussions how suitable are CMs for knowledge assessment purposes due to the fact that each learner may have his/her own knowledge structure, there exist a variety of scoring systems for CMs and no one is commonly accepted, as well as it is unclear may the final examination be based only on automatic evaluation of learners' CMs without the additional oral comments. The paper discusses the advantages and drawbacks of usage of CMs for knowledge assessment at the Faculty of Computer Science and Information Technology and lessons learnt on the bases of data analysis of students' solutions.

Software Defect Prediction Using Structural Source Code Changes

M. Heričko, Č. Gerlec

Faculty of Electrical Engineering and Computer Science
University of Maribor, Slovenia
Marjan.Hericko@um.si

Delivering high quality and reliable software products is one of the most challenging tasks in IS/IT organizations. Due to the complexity of today's software solutions it is difficult to develop products without any defects. In order to ensure the proper functioning and improve the quality of software, it is necessary to reduce and eliminate the number of defects.

There are various techniques to identify critical pieces of code. In this paper we will present a source code quality evaluation method that is based on past structural source code changes and the concept of code churn. Several machine-learning algorithms were used to predict files with defects and bugs. In order to test the usefulness of structural source code changes as a software quality indicator, we compared the efficiency of predicting bugs with structural source code changes and modified lines of code.

Construction of Education Monitoring Index

A. Jakaitienė^{1,2}, D. Stumbrienė¹, R. Želvys³

¹Institute of Mathematics and Informatics
Vilnius University

²Bioinformatics and Biostatistics Center
Department of Human and Medical Genetics
Faculty of Medicine, Vilnius University

³Department of Educology, Faculty of Philosophy
Vilnius University

Audrone.Jakaitiene@mii.vu.lt

Given education is one of milestones, which guarantees long-run well-being of a country, the assessment of the situation in education system status might be vital. Researchers calculated many indices those monitor diverse activity. However information about Education monitoring index is very variant and scarce, especially over time. A single index, which incorporates big data and gives notion about status of education system, might be relevant. European Commission within the Education and Training Monitor initiative calculates indices towards Europe 2020 strategy. Each country indices are calculated normalizing data with respect to EU28 average. Though calculated indices are not aggregated to a single index for a country. The objective of the research is to calculate Education monitoring index, as integral indicator, for Lithuania, Latvia, Estonia, United Kingdom, Finland and Germany available over time. We construct Education Monitoring Index for a country following structural CIPO framework, which describes relationships between Input, Process and Output in education within a certain Context. Data reduction methods are applied for an index construction. For methodologically consistent data for a country, we use EUROSTAT as data source.

Acknowledgment

This work has been partially funded by Research Council of Lithuania the Researchers team's project Nr. MIP-024/2015.

On Economic - Technological Optimization of High-Power Electric Cables

G. Jankevičiūtė, R. Čiegis, M. Meilūnas, O. Suboč, A. Ilgevičius

Department of Mathematical Modelling
Faculty of Fundamental Sciences
Vilnius Gediminas Technical University
Gerda@vgtu.lt

However, cross-section (metal) area, obtained by taking into account these restrictions, usually is appropriate to guarantee a safe and good quality of energy supply, but it is too small in order to minimize the total cost of power cable usage during the long time period.

In the past years remarkable progresses have been made in understanding of this problem and in working out mathematical models, which include the economic dimension as well (see (Wachter et al., 2011), (Tanides et al., 2003)).

The most important task in modelling the total costs, is to evaluate energy losses of power cable.

Here we use our original software for calculation of heat transfer in power cables under different loads (Ciegis et al., 2014).

References

Bruno De Wachter, Walter Hulshorst, Rodolfo di Stefano (2011). Whitepaper Cable Conductor Sizing for Minimum Life Cycle Cost. European Copper Institute, Cu0105.

Carlos G. Tanides, Maria J. Canepa, Osvaldo D. Petroni, Juliana Rizzone, Hector L. Soibelzon (2003). Economix Optimisation of Electric Conductors in Commercial and Public Buildings Using Methodology Proposed by IEC 287-3-2.ECEEE 2003, Summer Study.

Raimondas Ciegis, Olga Suboc, Andrej Bugajev (2014). Parallel algorithms for three-dimensional parabolic and pseu- doparabolic problems with different boundary conditions. Nonlinear analysis: modelling and control, 19(3).

Experimental Investigation of Chromatic Aberration Elimination in Digital Images of Eye Fundus Camera

V. Jakštys, V. Marcinkevičius, P. Treigys

Institute of Mathematics and Informatics

Vilnius University

Vitaakas@gmail.com

Chromatic aberration effect is a result of lens or objective focus dependency upon the light wavelength. The effect occurs when a lens or an objective brake different length waves in different angles, resulting in several image focal planes; finally, this results in an image with poor contrast. Mechanically this effect can be eliminated by adding extra lenses with a negative distance to the focus. According to the Abbe number, lenses can be situated in a manner that red and blue colour focal planes would match each other while wavelength of other colours refraction would be as same as possible. In this way only the axial chromatic aberration problem can be solved while contrasting the objective. In order to minimise the effect of photo camera systems designed without achromatic lenses, it is necessary to apply the image processing algorithms for chromatic aberration effect elimination. This study analyses images obtained with Optomed SmartScope M5 portable eye orbital camera which does not have achromatic lenses. Authors of the study present an investigation of algorithms published in academic literature which eliminates the effect of chromatic aberration and compare these algorithms.

Computer-Aided Localization of Prostate in mpMRI

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

¹ Insitute of Mathematics and Informatics

Vilnius University

² National Cancer Institute

Vilnius University

Justinas.Jucevicius@mii.vu.lt

Lithuanian cancer registry data from 2012 shows that prostate cancer is the most frequent not only when analysing cancer cases amongst men, but

it is also the most frequent tumour amongst both men and women combined. World Cancer Research Fund International states that prostate cancer is the second most frequent tumour amongst men and fourth most common among both sexes in the same year worldwide. Lithuanian cancer registry data also shows that prostate cancer is the second most common cause of cancer death and comes after lung cancer. Biopsy is the main way that can unambiguously detect prostate cancer if performed on the right location. However, only 70-80% of clinical cancer cases are detected by biopsy.

Nowadays multi parametric magnetic resonance imaging (mpMRI) is used to determine the location to perform biopsy on. Usually localization is done by hand, but it takes a lot of time and is inaccurate. This causes the need of software to aid locating abnormal prostate areas.

This work is dedicated to analysing currently used algorithms for locating prostate in mpMRI views, their effectiveness and arising problems. Further study will be conducted to determine possible ways of improving prostate localization algorithms.

Computerized Similarity Detection in Ceramic Assemblages Based on Morphological Properties

A. Kaliszewska, M. Syga

Institute of Archaeology and Ethnology
Polish Academy of Sciences, Poland
Agkaliszewska@gmail.com

This paper concerns the computerized classification methods of ceramic assemblages with exclusive focus on morphological properties of the vessels. We propose a method of mathematical representation of shapes of vessels, using 2D images as the input, and creating typologies of vessels. We compare our method to the already existing methods, and their usefulness in archaeological practice.

The choice of a method to representing curve impacts on establishing the similarity between the vessels. This leads to the generation of a typology by applying cluster analysis and discriminant analysis. In archaeology, ceramic assemblages are essential in studying past societies and setting chronologies. The archaeological practice is to study the shapes of the vessels and fragments excavated and create a typology of the vessels. Such a typology, when inserted in a wider regional framework, allows to

establish relative chronologies as well as to study relations between regions and sites and infer about the organization of past societies. However, the crucial step, which is creating the typological groups, is based on the experience and subjective judgment of the pottery analyst, therefore making it biased. This is where mathematical tools and computer science come with assistance.

For the numerical experiment we chose the pottery from Crete, of the Early to Middle Minoan period. The experiment is based on the already published archaeological material, which allows to compare the typologies obtained using our computer method to the typologies created by archaeologists using a traditional method.

Acknowledgment

The result present on the paper were obtained during the internship at the Systems Research Institute of the Polish Academy of Science, financially supported by the project "Information technologies: research and their interdisciplinary applications"(PO KL 04.01.01-00-051/10-00).

Multiple Criteria Decision Making decision making support - in a quest for a universal toolbox

I. Kaliszewski

Systems Research Institute

Polish Academy of Sciences, Poland

`Ignacy.Kaliszewski@ibspan.waw.pl`

In the presentation we shall be concerned with the following issue: what decision making support tools are offered by Multiple Criteria Decision Analysis and Multiple Criteria Decision Making for the unassisted use by laymen? It is nothing unusual to expect such tools from a domain researched vastly and deeply for more than forty years, the domain which is claiming itself to be a matured field.

We shall focus on a trickle of researches intentionally devoted to produce decision making support tools of low cognitive complexity. On the base of efforts of that sort, in this presentation we are able to propose a universal decision making support *toolbox*, ready for operation with minimal prerequisites.

We shall share our experience with the toolbox, gathered from applications.

Modelling and Control of Human Emotion Signals as Reactions to a Dynamic Virtual 3D Face

V. Kaminskas, E. Ščiġlinskas

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

This paper introduces how predictor-based control principles are applied to the control of human emotion – excitement and frustration – signals. 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 signals that characterize excitement and frustration. An input - output type model building method is proposed that allow building stable models with the smallest output prediction error. A predictor-based control law is synthesized by minimizing minimum variance control criterion in an admissible domain. Admissible domain is composed of control signal boundaries. Relatively high prediction and control quality of excitement and frustration signals is demonstrated by modelling results. Control signal boundaries allow decreasing variation of changes in a virtual 3D face.

Relation between the Number of Latent Variables and Degrees of Freedom of the Object Motion in the Image

R. Karbauskaitė, G. Dzemyda

Institute of Mathematics and Informatics
Vilnius University
Gintautas.Dzemyda@mii.vu.lt

One of the problems in the analysis of the set of images of a moving object is to evaluate the degree of freedom of motion of the object and the angle of its rotation in a separately taken image. Here the intrinsic dimensionality of multidimensional data, characterizing the set of images, can be used. Usually, the image may be represented by a high-dimensional point, the dimensionality of which depends on the number of pixels in the image. The results deal with the papers (Karbauskaitė et al., 2011), (Karbauskaitė

et al., 2015). The maximum likelihood estimator of the intrinsic dimensionality is explored experimentally.

The investigation is performed with the high-dimensional data points, corresponding to photos of the same person's face observed in different poses with different lighting direction. In a result, we have discovered that the number of latent variables is highly related to the number of degrees of freedom of the object motion. Therefore, the minimal possible intrinsic dimensionality of a data set of images is equal to the number of degrees of freedom of a possible motion of the object. However, the true intrinsic dimensionality may be larger than the number of degrees of freedom of a possible motion of the object. Illumination (lighting direction) has been discovered as the third latent variable, however its significance is much less than that of two directions of motion (two poses): left-and-right pose and up-and-down pose. We have also explored that the maximum likelihood estimator provides the right estimate of the intrinsic dimensionality, if the geodesic distances between the data points are calculated.

References

Karbauskaitė, R., Dzemyda, G. and Mazėtis, E. (2011). Geodesic distances in the maximum likelihood estimator of intrinsic dimensionality, *Nonlinear Analysis: Modelling and Control* 16(4): 387–402.

Karbauskaite, R., Dzemyda, G. (2015). Optimization of the maximum likelihood estimator for determining the intrinsic dimensionality of high-dimensional data. *International Journal of Applied Mathematics and Computer Science (AMCS)*, 25(4).

Statistical Analysis of Time Series Based on Incomplete Discrete Data

Yu. Kharin, H. Ageeva, H. Rudakouskaya

Research Institute for Applied Problems of Mathematics and Informatics
Belarusian State University

`kharin@bsu.by`

This paper is devoted to a new type of distortion in observation of time series $y_t \in \mathbf{R}$ that takes place in data mining: instead of true values of time series only incomplete discrete data $z_t = f(y_t) \in K$ is observed, where $f(\cdot): \mathbf{R} \rightarrow K$ is some known distortion function, K is a finite set. Examples of this distortion type: rounding errors, grouping, classification, interval censoring. We consider two typical models for hidden process y_t : multiple linear regression time series; $AR(p)$ time series. Statistical estimators for regression and autoregression coefficients based on incomplete discrete

data $\{z_i\}$ are constructed and analyzed. Computer algorithms are developed. Results of computer experiments on simulated and real data are presented.

Methods and Means for Generalization of Information in Specialized Information Systems

S. Konovalenko

Department information systems and technologies
University of Customs and Finance
Ukraine, Dnipropetrovsk
Customslab@ieee.org

The work is dedicated to research and development methods and means of the processing information in automated information system of the Customs Service of Ukraine. The main part of the customs information system is a subsystem of risk management. The main task of which is to collect, analyze, compile information on violations of the Ukrainian customs legislation. In the development of information technology to support decision-making there is very important problem of constructing models generalize the information on the basis of fairly large array (a heterogeneous input vector). These data contain hidden knowledge, which can be useful for making decisions for inspector of the Customs Service of Ukraine. In developing the algorithm of information processing must perform several steps: to make preprocessing of training input vector, choose the model of processing information, carry out interpretation of results. As a rule, large arrays of data may contain incomplete, inconsistent or noisy data, which strongly influence on the quality of recognition and could even lead to the unsuitability of certain algorithms for use. To overcome this kind of problem may the algorithms based on the rough sets theory. The paper discusses theoretical aspects applying methods of generalization based on the rough sets theory, whose task to receive optimal classification rules allow us accurately identify classes risks of violation of the Ukrainian customs legislation.

Quantitative Evaluation of Morphologic Changes in Histologic Samples by Means of Mathematical Morphology

A. Kriščiukaitis¹, R. Petrolis², R. Ramonaitė³, R. Pečiulis⁴,
D. Jančiauskas⁵, A. Kiškis⁴, L. Kupčinskas²

¹ Department of Physics, Mathematics and Biophysics
Lithuanian University of Health Sciences

² Neuroscience Institute
Lithuanian University of Health Sciences

³ Institute for Digestive Research
Lithuanian University of Health Sciences

⁴ Kaunas University of Technology

⁵ Clinic of Pathology
Lithuanian University of Health Sciences
Algimantas.Krisciukaitis@lsmuni.lt

The efficacy of histological analysis used for evaluation pathologies can be improved by means of digital imaging, giving quantitative estimates of known diagnostic features used by the experts, enriching them by mathematical morphology based estimates. This study illustrates proposed methodology by formation of diagnostic features set reflecting inflammation severity in colon tissue and elaboration of automatic evaluation method for computer-aided diagnostics.

Acute and chronic colon inflammation in the experimental animals was induced by oral administration of 3.5 % dextran sulphate sodium (DSS, TdB Consultancy, Uppsala, Sweden). Histological images were taken by means of OLYMPUS IX71 light microscope ($\times 20$ magnification) equipped with Q IMAGING EXI aqua camera at 1392×1040 pixels resolution ($0.6 \mu\text{m}/\text{pixel}$). Image processing algorithms were realized as programs in MATLAB computation environment.

Image features formation was performed on 512×512 pixel colon image cutouts selected by the experts, representing as much as possible homogeneous and typical tissue pattern without any gaps. Main diagnostic features in histologic images characterizing ulcerative colitis include crypt distortion, branching, and appearance of lymphoplasmacytic infiltrate deep in the crypts. Every feature has its representation in digital image: for example, crypts appear as elliptic white spots varying about 180–350 pixels long and 50–130 pixels wide, eosinophils, are rounded spots of 7–25 pixels etc. We used Gabor filters for detection and evaluation of such

morphological changes. Principal component analysis was used to construct optimal representation of such morphological changes. First principal component (PC1) was representing the major part of features variation (97 %) and was selected as a measure of inflammation severity. Validation of new measure was performed by means of custom-made software realizing double blind comparison of differences in PC1 with expert's opinion about inflammation severity presented in two compared pictures. Overall congruence of estimations was 80 %.

Conclusion: Quantitative estimates of known histological image features, enriched by mathematical morphology based estimates can form informative diagnostic features set. Principal component analysis can produce one combined quantitative estimate for evaluation of certain pathology.

Energy Efficient Computing in PC Clusters on Demand

O. Kurasova¹, T. Petkus¹, E. Filatovas², A. Igumenov²,
S. Ivanikovas²

¹ Faculty of Science and Technology
Lithuanian University of Educational Sciences

² Faculty of Fundamental Sciences
Vilnius Gediminas Technical University
Ernestas.Filatovas@vgtu.lt

Recently, high performance computing (HPC) centres with huge resources are constantly established and their computational power continually increases. According to the concept of such centres, they are designed for solving large scale problems, where extremely high performance computational resources are required. However, there are many types of problems that could be solved by using clusters consisted of personal computers (PCs) available in business and public offices as well as in universities, that could be used for computations on demand. Moreover, computing performance of the current PC processors enables to solve complicated problems with high enough computational burden. Energy consumption and savings are ones of the main factors when drawing up cluster usage strategies and developing energy efficient algorithms. Our research is devoted to investigation of energy consumption in PC clusters when solving time-consuming problems in order to identify the most energy efficient strategies when distributing tasks between processors. Moreover, such PC clusters are usually used on demand, therefore the

problems could be solved while the computers are used by office workers/students for everyday works. Thus, it is also important to consider user's satisfaction with PC response to his/her request when the computer is highly loaded.

Multistage Speech Emotion Recognition for Lithuanian: Experimental Study

T. Liogienė, G. Tamulevičius

Institute of Mathematics and Informatics

Vilnius University

Tatjana.Liogiene@mii.vu.lt

Intensive search for reliable and robust speech emotion features introduced huge sets of acoustical features. Employment of these features sets for straightforward classification of emotions complicates the task of speech emotion identification. Therefore, various feature transformation and selection techniques, alternative classification schemes were proposed for recognition task.

The multistage classification scheme for emotion recognition is based on idea that all emotions can be arranged into predefined emotional groups (low-pitch and high-pitch emotions for example). These groups enable us to simplify emotion classification task by organizing the process in stages. During the first stage all unknown emotional speech patterns are classified into foregoing emotion classes. Afterwards, all these groups are classified into lower level classes or particular emotions during the second stage. The process is terminated in the last stage where the entire set of analyzed emotions is obtained. Each classification stage is characterized by employed feature subsets and analyzed emotions groups.

In this study we applied the multistage classification scheme for Lithuanian speech emotion identification task. Three different feature selection techniques were employed for this scheme to compare: maximal efficiency criterion, minimal cross-correlation criterion and sequential forward selection (SFS) technique. SFS technique shows clear superiority in emotion recognition rate over two remaining selection criterions. Nevertheless, SFS produced feature sets of higher order (in case of larger emotion sets) and caused much longer training process.

Multicriteria Modelling of Financial Markets Efficiency

M. Liutvinavičius, V. Sakalauskas, D. Krikščiūniene

Kaunas Faculty of Humanities

Vilnius University

Marius.Liutvinavicius@khf.vu.lt

General instability of financial markets and recent financial crises led to intensive search for new ideas and methods in asset pricing and risk management. Recent research works confirm the idea of irrationality of financial markets versus their efficiency. Financial markets are influenced not only by economic, social or geopolitical factors. Expectations and fears of investors, manipulation of information in media streams and social networks play very important role as well. Huge amount of information emanating from sources of various origin create factors which should be incorporated in financial crisis prediction. This is growing the demand for application of analytical methods capable to process information streams characterized as Big Data. Our research object is financial crises and factors determining them. Financial markets can be perceived as dynamic systems of interacting agents which modeling and prediction are based on computer technology. The problem of increasing the effectiveness of identification of anomalous situations in financial markets is solved in this article. The research approach deals with system dynamics and behavioural finance areas. The main objective is modelling complex economic processes, taking into account the influence of different factors and their interdependencies. The models of real life processes are designed and simulated in order to experimentally analyze various possible scenarios. The research goal is to offer new methods for identification of anomalous situations and factors causing them. We investigate the specific behaviour patterns influenced by rational and irrational factors to the professional and non-professional investment cases. There is also a need to analyze the investment process in accordance to such aspects as market risk evaluation, investors' expectations and risk tolerance. The research results reveal that emotions-based investment decisions lead to losses. They also confirm that dynamic simulation of financial systems can be applied to discover hidden patterns of investors' behaviour. Application of Big Data analytics allows to extend the scope of analyzed factors and increase the efficiency of prediction.

Effects of the Parameters on the Air Flow in a Four-Channel Cyclone

T. Leonavičienė, P. Baltrėnas

Department of Mathematical Modelling
Faculty of Fundamental Sciences
Vilnius Gediminas Technical University
Terese.Leonaviciene@vgtu.lt

The problems of air pollution in today's world are very importance. Cyclones are the most popular equipment for removing solid particles from air. To increase the efficiency of removing solid particles multi-channel cyclones are used. In this research a three-dimensional numerical model of air flow in a new generation four-channel cyclone was analyzed. The cylindrical part of this cyclone contains 3 cylindrical half-rings that form channels with adjusted spacing. Due to the specific structure of the cyclone particulate matter is removed in cyclone channels 1 and 2. In order to investigate air flow in such a cyclone, a detailed numerical analysis was performed. The model of a virtual four-channel cyclone was created according to the data provided by the VGTU Research Institute of Environmental Protection. Air flow movement inside a cyclone is described by a system of differential equations which is formed according to the laws of continuity (material mass) and conservation of momentum. We analyzed the influence of the cyclone parameters on the results. The cyclone's cylindrical half-rings can be regulated by selecting different spaces for the peripheral (reversible) and transit (entering another channel) flows. Due to such structure of the cyclone and the set positions of half-rings results in an increase in air flow velocity in the channel 1 and channel 2. This increase in velocity is important for the removal of particulate matter from the air as efficiently as possible. Air flow in the channels of a four-channel cyclone was modelled. A numerical analysis was carried out using the finite volume method. The numerical results for air flow velocity were fixed at the characteristic points of each channel. The numerical results are presented and analyzed. The obtained results can be used for the analysis of air cleaning in a multi-channel cyclone.

Visualization of Differences in Language Usage of Parliamentary Position and Opposition in the Lithuanian Parliament 2008–2012

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

¹ Vilnius University

² Baltic Institute of Advanced Technology

³ Vytautas Magnus University

Justina@bpti.lt

This study explores and visualizes language variation in the political debates of the Lithuanian Parliament (2008–2012 term) based on the parliamentary position and opposition speeches transcripts. The language is studied and visualized with respect to style, i.e. "style of the parliamentary position" and "style of the parliamentary opposition" of the language usage by applying computational stylistics or stylometry. Stylometry is based on two hypotheses. One of them states that each individual has a unique style. The second hypothesis proposes that this unique style can be measured. If you look from the Information Retrieval perspective, stylometry allows gaining meta-knowledge, i.e., what can be learned from the text about the author, eg, gender, age, psychological characteristics, political affiliation, etc. Part of the corpus of parliamentary speeches from the Lithuanian Parliament ("Automatic Authorship Attribution and Author Profiling for the Lithuanian Language" (ASTRA) (No. LIT-8-69), 2014–2015) was used for this study. It contains of parliamentary speeches from March 1990 till December 2013 that have at least 100 words. Experimental setup for this study consists of stylistic features that do not require external linguistic tools (in this case most frequent words) and they are combined with unsupervised machine learning algorithms, namely (hierarchical) clustering. The findings - language variation in terms of word usage - are visualized using dendrograms, thus mapping positions of the text samples in relation to each other according to the parliamentary position and opposition. This experimental setup leads to a successful capture, exploration as well as visualization of language variation in terms of word usage in the political debates of the Lithuanian Parliament (2008-2012 term) based on parliamentary position and opposition.

Automatic Thematic Classification of the Titles of the Seimas Votes

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

¹ Vytautas Magnus University

² Baltic Institute of Advanced Technology

³ Kaunas University of Technology, Institute of Public Policy and Administration

Vytautas.Mickevicius@bpti.lt

Statistical analysis of parliamentary roll call votes is an important topic in political science as it reveals ideological positions of members of parliament and factions. However, these positions depend on the issues debated and voted upon as well as on attitude towards the governing coalition. Therefore, analysis of carefully selected sets of roll call votes provides deeper knowledge about members of parliament behaviour. However, in order to classify roll call votes according to their topic automatic text classifiers have to be employed, as these votes are counted in thousands.

Automatic classification of voting titles is a rather known subject in the world, however it is quite new issue in Lithuania.

This is a work in progress with the aim to determine the most suitable methods for automatic classification of short political Lithuanian texts – topics of votings in the Seimas. Different text classifiers (such as *Support Vector Machines*, *Multinomial Naive Bayes*, *k Nearest Neighbors*) and natural language processing techniques (such as *bag-of-words*, *n-gram*, *tf-idf*) are being tested.

This work is a part of a larger project – an infrastructure in the form of website, dedicated for monitoring and analysis of Seimas' voting results. A wider variety of statistical and machine learning methods are used in the aforementioned infrastructure, such as multidimensional scaling, cluster analysis and text classifiers.

Hybrid Approach for Model Renovation: New Life of Classical Chanter-Thornley Model of Mushroom Cultivation

O. Myronycheva ¹, V. Vigont ², A. Topaj ², V. Denisov ³

¹ Tavria State Agrotechnological University
Melitopol, Ukraine

² Agrophysical Research Institute
Saint-Petersburg, Russia

³ Department of Informatics and Statistics
Klaipėda University, Lithuania
Vitalij.Denisov@ku.lt

Hybrid (multi-paradigm) modelling approach allows reviving of old classical models, making them more consistent with reality. In this work well-known Chanter-Thornley mathematical model of mushroom cultivation was modified by combining elements of system dynamics, discrete event and agent-based modelling methods. The model was implemented using a multi-paradigm modelling approach and then studied numerically using the AnyLogic simulation environment. The main change is multi-agent description of sporophores component instead of single compartment presentation in original model. The built-in procedures of automatized computer simulation (calibration, sensitivity analysis, and optimization) have been applied for improvement and investigation of developed model. In particular, it allows solving practically applicable task - optimization of the harvesting time of fruit bodies in order to provide the maximum total yield during mushrooms fruiting.

On the Inequality in Open Queueing Networks

S. Minkevičius^{1,2}, E. Greičius¹

¹ Faculty of Mathematics and Informatics
Vilnius University

² Institute of Mathematics and Informatics
Vilnius University
Minkevicius.Saulius@gmail.com

The paper is devoted to the analysis of queueing systems in the context of the network and communications theory. In this paper, we investigate an open queueing network model in heavy traffic. Also we investigate the

inequality in an open queueing network and its applications to the theorems in heavy traffic conditions. Note that inequality is the key inequality to prove several laws (fluid approximation, functional limit theorem, and law of the iterated logarithm) for a queue of customers in open queueing networks in heavy traffic conditions.

A Method of Comparing Business Rule Sets

L. Paliulionienė

Institute of Mathematics and Informatics

Vilnius University

Laima.Paliulioniene@mii.vu.lt

Structural and operative business rules can be expressed in different forms, using modalities (e.g. obligation and necessity claims) and involving negation. In order to increase the efficiency of processing business rules it is important to detect redundant rules in the set, for example, semantically equivalent or overlapping rules. To this aim, we need to have a possibility to compare business rules and find differences between them or prove their equivalence.

On the other hand, it would be valuable to have a possibility to compare not only separate business rules but also the sets of rules. These can be alternative business rules sets that have the same purpose but possibly give different results in some situations, or these can be subsets of a set. The detection of differences is also useful in the managing changes in business rules systems.

This paper proposes a method to find semantic differences between business rules or their sets. The method supposes that rule statements are expressed in some declarative way (for example, structured natural language) and can be transformed into first-order predicate logic with some restricted extensions. The method uses the generating of elementary test situations, logic-based derivation and the analysis of results produced by the derivation.

An Evaluation Model for Web-Service-Based Information System Integration

L. Pavlič, M. Vaupotič, M. Heričko

University of Maribor, FERi, Slovenia

Luka.Pavlic@um.si

Modern techniques for integrating components into a single, or even across multiple, information systems are primarily based on HTTP protocol. However, two main HTTP-based competing techniques can be employed at the integration layer. One is based on traditional web services (using WSDL and SOAP technologies), while the other is REST-enabled web services. They are both well established and highly standardized. However, the information system context and priorities strongly determine which technique architects should choose, since different web services offer different architectural properties and quality of service (QoS) parameters. Selecting an inappropriate technology, or using a correct technology with an inappropriate approach, can have a bad influence on a system's overall quality characteristics.

This is why we focused on the challenge of selecting an appropriate architecture for the web services integration layer. Based on a detailed analysis of the web services characteristics and their impact on the quality of information systems, we developed a comprehensive hierarchical multi-attribute decision model which enables us to select the optimal architectural approach for the given design problem. Alternatives were evaluated based on QoS parameters. These parameters are: performance, modifiability, scalability, reliability, integrity and security. The decision model consists of criteria that describe the presented QoS parameters.

We have proposed a generalized decision model by evaluating alternatives for basic decision-making criteria, which were derived from existing surveys and expert knowledge. With the help of the generalized decision model, and the descriptions of the decision-making criteria, a designer can apply a variant decision model that corresponds to the presented design problem. The usefulness of the proposed solution was evaluated based on a case study.

Determining the Number of Topics in a Semi-Supervised Setting

M. Pavlinek, M. Heričko, V. Podgorelec

Faculty of Electrical Engineering and Computer Science
University of Maribor, Slovenia
Miha.Pavlinek@um.si

In the field of text mining, the topic-based representation of documents is often used as an alternative to a unigram model. As semantically related clusters of words, topics can describe an entire text corpus in much smaller feature space while preserving most intrinsic information. The latter is correlated with the quality of the topic models, which depends on several parameters, where the number of topics is the most influential among them. This number must be specified in advance and there is no conventional method to determine it automatically. In general, it is related to the corpus size but also depends on the task in which topics are used. Nevertheless, some proposed methods and metrics already exist but mostly do not yield comparable results. To find an appropriate one, we evaluated some of them on various data sets with Latent Dirichlet Allocation, one of the most successful topic models. In this presentation, we will show which of the proposed methods provided the best number of topics for two different settings. Inside document categorization task topic modeling was used in a supervised manner and in semi-supervised settings with a self-training method.

Analysis and Forecast of Teachers' Demand in Lithuania

T. Petkus, O. Kurasova

Faculty of Science and Technology
Lithuanian University of Educational Sciences
Tomas.Petkus@leu.lt

Declining Lithuanian population, optimizing the network of the general education schools, decisions are made on the assumption that there is a large surplus of teachers, and annually the number of state-funded places for future teachers is reduced. In order to assess the teachers' demand, to draw the proper conclusions, and to make the right decisions, the analysis of the teachers' demand and forecast should be continuously performed. In this research, the teacher's demand has been forecasted up to 2024 year.

The criteria that influence this forecast have been chosen. It has been determined how the number of pupils are forecasted taking into consideration the birth rate and the number of pre-school children, the change of pupils moving from a lower to a higher class. The distribution of teacher age that influence the forecast of the teachers' numbers has been also investigated. The number of new prepared teachers has been evaluated. The teachers' demand (surplus or deficiency) has been calculated taking into account the forecasted number of pupils, the age of teachers, the number of students, graduated teacher training programs. The forecasting shows when the deficiency of teachers of the physics, mathematics, Russian language, chemistry, information technology and other subjects appear. These facts should be taken into account immediately when allocating state-funded places for future teachers.

Optimal Management of Forest-to-Biofuel Supply Chain: a Case Study

D. Podkopaev^{1,2}, A. Marques³, A. Mäkinen¹, B. Oliveira³,
J. Rasinmäki¹

¹Simosol Oy, Finland

²System Research Institute, Polish Academy of Sciences

³Institute for Systems and Computer Engineering, Technology and Science, Portugal

Dmitry.Podkopaev@ibspan.waw.pl

Good management practices in forest-based value chains contribute to economic and environmental sustainability of many regions in Europe as well as across the world. Some optimization problems in this field are very complicated due to complex structure of business processes. One of such problems is optimal planning of company's activities aimed at production and distribution of biofuels. We consider a case study of a Finnish company working in this field.

Given forest residue piles located at roadside and demand for biofuel by power plants, the company uses portable chippers to produce biofuel (wooden chips) at piles, and uses trucks to transport chips to power plants. Company's own terminal(s) can be used as seasonal storage. Trucks' movements have to be synchronized with chippers' works at piles. Power plants pay for energy contained in chips, which increases in time as residue piles dry. These as well as other dynamic aspects make the problem time-dependent and computationally complex.

A consortium of several European institution addresses the case study in a systematic way. We formulate and integrate together different levels of the planning problem aiming at developing a computer system of integrated planning and control. At the tactical level, the problem is modeled as a mixed-integer linear programming problem. We describe this model and discuss computational approach to its solution.

Classification and Visualization Algorithms on Cloud Computing: Issues and Advantages

T. Pranckevičius, V. Marcinkevičius

Institute of Mathematics and Informatics

Vilnius University

`Tomas.Pranckevicius@mii.vu.lt`

The analysis of large data and data mining deals with understanding and using new ways and tools how to process, operate, and reuse large multi-dimensional statistical or technical data sets, generated by many different sources, such as science, sales, insurance, research, transport, finance, social media and anything found on the internet. The purpose of this research is to investigate the issues and advantages of the Cloud Computing technology impact on parallel data processing. So far multidimensional data research methods may be successfully integrated to the Cloud computing technology based large scale data processing platforms and computing capabilities could be provided as a service. Existing parallel computing algorithms are increasingly applied to large scale data processing platforms and more realization solutions could be found, but unfortunately transition of data analysis algorithms is still slow and long process. The analysis systematically overviews and compares different tools and technologies used to deal with big data and helps to understand issues, challenges and advantages.

The Challenges for Recommendation Systems of a New Dreams Meanings Dataset

A. Rapečka, G. Dzemyda

Institute of Mathematics and Informatics
Vilnius University
Aurimas.Rapecka@mii.vu.lt

With a rapid development of technologies and the increasing number of internet users, more and more products and services are transferred into the virtual space. Various offers to buy something by means of internet or to use a certain service without leaving the home should save the client's time. There are a lot of methods to recommend products or services. Each method has its advantages and disadvantages. Also, there are a lot of different datasets and some of them are very specific. The aim of this work is to introduce a new dataset and to present the first investigations with this dataset.

Dreams meanings dataset was collected in 2012–2015 by using most popular dreams meanings website in Lithuania. Dataset consists of 27000000 search records by 1470000 users in 4200 dreams meanings.

Complex challenges for recommendation systems occur in this dataset. First challenge is to determine if it is possible to forecast dreams for future. Second challenge is to determine if there are similar groups of users in dataset. And largest challenge is to determine relations between user's dreams and events of nature, like moon phases, magnetic storms etc. This dataset is available for scientific researches and is published online.

An Application of ARAP Method for Parametrization of Shoetrees

M. Sabaliauskas, V. Marcinkevičius

Institute of Mathematics and Informatics
Vilnius University
Martynas.Sabaliauskas@mii.vu.lt

Currently, custom footwear production flattening of individual lasts is a manual job. Standard manufacturing process consists of three main steps: segmentation of shoetrees, flattening of segmented parts using wax foil, and digitization of 2D results. First two steps in this process can be performed by computer software without using human work force or wax

foil. In this research, we present a new method, with automatically perform segmentation and parametrization (flattening) of shoetrees. At the beginning we scan shoetree with 3D scanner, than perform shoetree surface specific segmentation and parametrization of the segmented surfaces. The segmentation part of the method is based on grouping triangles of the surface according normal directions of it. In the parametrization part of the method, we are using one of authalic mapping methods - ARAP. The method was accelerated by introducing algebraic method for fast computation of large inverse sparse matrices. An appropriateness of this method was conducted by comparing its results to manual standard manufacturing method.

Activity Based Costing by Two Stage Stochastic Programming

L. Sakalauskas, A. Ušpurienė

Institute of Mathematics and Informatics

Vilnius University

Leonidas.Sakalauskas@mii.vu.lt

The activity-based costing in service companies is considered modelling the uncertainty of service demand in the statistical probabilistic way. The two-stage ABC stochastic programming model has been developed. The model involves a stochastic demand and a relative minimum level to be satisfied. To solve that stochastic model, a modified L-shaped algorithm has been implemented.

New Computational Technologies in Numerical Analysis

Y. D. Sergeev

Numerical Calculus Laboratory

University of Calabria, Italy

Yaro@dimes.unical.it

In this talk, a recent computational methodology is described. It has been introduced with the intention to allow one to work with infinities and infinitesimals numerically in a unique computational framework. It is based on the principle ‘The part is less than the whole’ applied to all quantities (finite, infinite, and infinitesimal) and to all sets and processes (finite and

infinite). The methodology uses as a computational device the Infinity Computer (patented in USA and EU) working numerically with infinite and infinitesimal numbers that can be written in a positional system with an infinite radix. A number of numerical examples is presented.

Model-driven Development and Application of Model-to-Model Transformations in a CASE Tool

T. Skersys, P. Danėnas, R. Butleris

Kaunas University of Technology

Rimantas.Butleris@ktu.lt

One of the main features of Model Driven Architecture is model-to-model (M2M) transformations, which improve the overall model-driven systems development process by speeding up the development process itself and also enabling reusability of existing models within a single or even multiple projects. However, CASE tool-supported M2M transformations quite often lack so needed flexibility and customization options – in many cases, this is due to the fact that those transformations are implemented as a hard-coded solutions, rather than a model-based ones; moreover, number of such transformations in a CASE tool might also be quite limited and not always satisfy the needs of a user.

Our current practical experience with various CASE tools, as well as the feedback from our professional partners indicates that a certain subtype, i.e. user-interacted partial M2M transformations are in particular demand by many, especially when it comes to agile, highly iterative model-driven development process. By the term “partial M2M transformation” we assume the case where only a user-defined part (fragment) of the model (i.e. a set of selected source model concepts) is being transformed into a fragment on another model (i.e. a set of target model concepts). A common practice shows that a CASE tool user usually prefers to have a certain degree of “freedom of choice” when it comes to selecting a process of work, ability to create few models at a time while reusing each other’s concepts, modifying these models on demand, etc. This is where the advantage of partial M2M transformations over traditional complete M2M transformations is the most obvious.

The main goal of our research is to develop a practical approach for the development and application of user-interacted partial M2M transformations, which would improve the overall usability of M2M transfor-

mations in a CASE tool environment. One of the key features of this approach is the model-driven development and customization of M2M transformations, which greatly increases the flexibility and usability of this capability when used in a CASE tool environment. Another distinctive feature of the approach is that it can be applied to any graphical modelling language that is based on UML metamodel. The approach is currently being experimented with four modelling languages: UML, BPMN, SoaML and SBVR. The latter three have UML metamodel-based implementations.

Our current implementation prototype of the approach works as an extension to the CASE tool MagicDraw; however, it can be adopted by any other advanced CASE tool that meets certain capability requirements.

Automated Blood Vessel Detection and Characteristic Evaluation in Eye Fundus Images

G. Stabingis¹, J. Bernatavičienė¹, G. Dzemyda¹, D. Imbrasienė^{2,4}, A. Paunksnis², V. Slavinskytė³, P. Treigys¹

¹ Institute of Mathematics and Informatics
Vilnius University

² Hospital of Lithuanian University of Health Sciences Kauno klinikos

³ Republican Hospital of Kaunas

⁴ Lithuanian Sports University

Giedrius.Stabingis@mii.vu.lt

Eye fundus image analysis is important in diagnosis of diseases especially in early diagnosis of diabetic retinopathy, glaucoma, and hypertension and so on. Very often, the image analysis is performed on high quality fundus images, which are gathered by ophthalmologist using high quality fundus cameras. The image quality of mobile fundus cameras is growing, and these cameras are becoming available for family doctors. The adjusted automatic eye fundus image analysis algorithm for images gathered with mobile eye fundus camera is presented here. Images are pre-processed to achieve higher quality. Blood vessel network is detected using common morphological operations. The detected blood vessel network is used for better optic disk localization. The located optic disc is parametrized and is used for the automatic measurements for the width of main blood vessels. These measurements are performed for the images of both eyes and images are taken before and after the sport load. The data

of 11 different people is used in this experiment. The automatic measurements are compared with the measurements performed by the expert.

Image Classification for Quality Prediction after JPEG Compression

J. Tichonov¹, O. Kurasova¹, E. Filatovas²

¹Institute of Mathematics and Informatics
Vilnius University

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

Nowadays digital cameras allow us to get high quality images which take a lot of external storage places. Therefore, images are usually compressed by lossy algorithms, and the JPEG algorithm and format is the most common one for such image storage. Digital image processing tools allow us to vary many parameters of the JPEG algorithm when selecting the most suitable values for a particular image. These parameters control a compression level, thereby a quality of the compressed image. However, even in the case of the same values of the parameters, depending on the image content, the images of different quality and sizes are obtained after compression. It is difficult to determine in advance how the values of the parameters in the JPEG algorithm will influence a certain image. Therefore it is purposeful to predict prospective image quality before usage of the compression algorithm. If an image falls into a class of the high-quality images, it can be stated that the chosen values are suitable the image compression. Otherwise, if the image is assigned to the low quality image class, it is necessary to look for other values of these parameters.

In this research, the image classification has been investigated, where the goal is to classify images in order to predict how the JPEG algorithm will affect the images. We aim to determine which features describing the original images have to be used in the image classification into two classes with high and low quality after compressing them by the JPEG algorithm. Quality of images after compression is estimated by a structural similarity (SSIM) index method. We conclude that when the support vector machine is used for the classification the most accurate classification results have been obtained with images pre-processed by edge detection filters and thresholding.

Presymptomatic Genetic Risk Assessment Methods

V. Tiešis, A. Lančinskas, V. Marcinkevičius

Institute of Mathematics and Informatics

Vilnius University

Vytautas.Tiesis@mii.vu.lt

The purpose of the genetic risk assessment methods is to evaluate the quantitative index which indicates a lifetime risk of the disease determined by an individual's genotype. The presymptomatic genetic risk assessment is analysed in the paper. The chronic non-communicable diseases are caused by a combination of multi-locus genetic risk factors. In this case the main problem is to evaluate mutual impact of numerous disease-associated loci of genes because the number of different genotypes increases exponentially with the number of loci. Therefore, none of such very expensive statistical evaluations have been conducted until this time. So, in research papers the strong assumption is usually made that impacts of different loci are mutual independent. This and other assumptions are used by the genetic risk assessment companies, e.g. Navigenics and 23andMe, to calculate a lifetime risk of a disease. The object of the paper is to compare such risk assessment methods. It is difficult to calculate the error due to assumptions; therefore, the evaluation of risk assessment methods was done experimentally. The theoretical disease model that describes both environmental and genetic factors has been used for evaluation of assessment methods. The system of nonlinear equations for tuning model's parameters to real statistical parameters of the disease has been developed. The Receiver Operating Characteristic curve has been used to evaluate the quality of the methods as predictive tests. The experiments have demonstrated that investigated risk indexes produce very similar ROC curves. However, the quality of the risk indexes as a predictive test is low in the case when indexes are based only on genetic factors.

Respiratory Motion Compensation by Model-based Functional Target Tracking

R. Užupyte^{1,3}, T. Krilavičius^{2,3}, I. Žliobaitė^{3,4},
H. Simonavičius⁵

¹ Vilnius University

² Vytautas Magnus University

³ Baltic Institute of Advanced Technology

⁴ Aalto University, Finland

⁵ UAB Rubedo Systems

Ruta.Uzupyte@bpti.lt

The main goal of radiation therapy is to destroy cancer cells while minimizing exposure to healthy cells. Tumor motion caused by respiration limits the effectiveness of radiation therapy. Various techniques, such as Pearson correlation, linear interpolation, partial-least squares, can be used in order to compensate for tumor location uncertainties. However, performance of these approaches still leaves plenty of space for the improvement. In this research we are trying to predict position of the functional target (tumor) from an external marker by applying different types of regressions models.

Experimental results show that multiple linear regression model is the most suitable method for functional target motion prediction from the analyzed methods (simple linear regression, linear regression models with first-order of autoregressive errors). Furthermore, more accurate predictions are obtained using external markers with a greater range of movement (external markers placed in the abdomen area).

High Frequency Statistical Arbitrage Strategy Engineering and Algorithm for Pairs Trading Selection

M. Vaitonis, S. Masteika

Kaunas Faculty of Humanities

Vilnius University

Mantas.Vaitonis@khf.stud.vu.lt

There are many trading strategies that do incorporate algorithms and computers. Statistical arbitrage is one of them. When using this strategy profit arises from pricing inefficiencies between assets. Statistical arbitrage or

pair trading is strategy where you have to find two assets that move together and take long/short positions when they diverge abnormally, hoping that the prices will converge in the future. The most popular statistical arbitrage strategies were tested using high frequency gas future market data. One strategy which was used in the research first was implemented by Perlini M. in Brazilian market, the other one was based on J.F. Caldeira and G.V. Moura implemented strategy in Sao Paulo stock exchange and the last one used was covered by D. Herlemont. As the result best performance was shown with D. Herlemont pairs trading strategy. The main objective of this research is to check if this strategy does work as good on Baltic region stocks as it did on gas future market. All previously used strategies did work with predefined pairs for trading. During this research there are no predefine pairs, thus it was necessary to incorporate pair selection algorithm in order to find best pairs for each trading period. Least squares method was used for defining pairs. Data base for testing did cover trading signals from 2014-10-01 till 2015-03-31. At the end of research we could see the pairs that the algorithm did find for trading. These pairs were used for trading strategy which was measured by the profit it did generate.

Knowledge Based UML Dynamic Models Generation Method

I. Veitaitė, A. Lopata

Kaunas Faculty of Humanities

Vilnius University

`Ilona.Veitaite@khf.vu.lt`

The main scope is to present UML dynamic diagrams generation from enterprise model method. As organizations grow and information systems get bigger it is needed to create accurate and complicated analyses for developing systems. The quality of these analyses is decisive for the success of information systems development process, because the later a problem is found the more expensive is to solve it. Designers every day face new challenges when they need to encounter and understand other designers' and analytics' made models. Process like that creates additional problems and mistakes. Automatization of information system engineering process lets create better and more qualified models with less mistakes. To achieve this purpose there is analysed ISO standards influence to information systems engineering process, ISO standards and

UML models integration with knowledge-based enterprise model, MOF architecture's role in IS engineering process, possibilities to enrich MOF architecture's composition with new knowledge-based layer and improved MOF architecture's usage in UML dynamic models generation from enterprise model process.

Visual Analytics for Big Data

J. Zubova, O. Kurasova, V. Medvedev

Institute of Mathematics and Informatics
Vilnius University

Jelena.Zubova@mii.vu.lt

Data and information have been collected and analyzed right through history. New technologies and Internet have boosted the ability to store, process and analyze the data. Recently, the amount of data being collected across the world has been increasing at exponential rates. Various data come from devices, sensors, networks, transactional applications, web and social media – much of them are generated in real time and very large scale. The existing technologies and methods that are available to store and analyze data cannot work efficiently with such amount of data. That is why big data should speed-up technological progress. Big data become a major source for innovation and development. However, the huge challenges, related to the storage of big data, their analysis, visualization and interpretation, arise. Size is only one of several features of big data. Other characteristics, such as the frequency, velocity or variety, are equally important in defining big data. A variation of big data definition is wide of range, therefore it is purposeful to overview various definitions used by academia and industry.

Big data analytics become mainstream in the era of new technologies. It is the process of examining big data to uncover hidden and useful information for better decisions. Big data analytics assists data researchers to analyze such data which cannot be processed by conventional analysis tools. Big data analytics involves multiple analysis including visual presentation of data. Thus, visual analytics enables to see hidden relations which cannot be detected using data analysis methods. The usage of visual analytics for solving big data problems brings new challenges as well as new research issues and prospects.

The primary focus of the research has been on visual analytics developed by the world's leading IT companies (Microsoft, SAS, HP, IBM, Amazon, etc.) to gain valid and valuable insights from big data. The visual analytics are overviewed considering to the adaptability for visual representation of big data. It is important to note that the visual analytics are

sensitive to data of the large volume, especially to big data, and they are being constantly improved by considering big data challenges.

On One-Step Worst-Case Optimal Trisection in Univariate Bi-objective Lipschitz Optimization

A. Žilinskas, G. Gimbutienė

Institute of Mathematics and Informatics
Vilnius University
Grazina.Gimbutiene@mii.vu.lt

The bi-objective Lipschitz optimization with univariate objectives is considered. The concept of the tolerance of the lower Lipschitz bound over an interval is generalized to arbitrary subintervals of the search region. The one-step worst-case optimality of trisectioning an interval with respect to the resulting tolerance is established.

The theoretical investigation supports the previous usage of trisection in other algorithms. The trisection-based algorithm is introduced. Some numerical examples illustrating the performance of the algorithm are provided.

On Branch and Bound for Non-convex Multi-objective Optimization

J. Žilinskas

Institute of Mathematics and Informatics
Vilnius University
Julius.Zilinskas@mii.vu.lt

Real-world optimization problems often involve more than one criteria, what leads to solution of multi-objective optimization problems. Multi-objective optimization is a research area rich with various approaches. Many methods convert the multi-objective optimization problem into a set of single-objective problems. Apart from general disadvantages of such approaches, in non-convex multi-objective optimization even the scalarized single-objective optimization problem is not easily solved - global optimization must be used. Although it could be possible to solve non-convex multi-objective optimization problem solving a set of global optimization problems by branch and bound algorithm, a more advantageous approach is to develop multi-objective branch and bound algorithms which solve multi-objective optimization problem in one run.

7th 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



Research
Council of
Lithuania



Sponsors

