SEE Forum
On Data Science

2016 South East European Forum on Data Science

June 20 – June 21, 2016 Belgrade, Serbia

Organized by:

University of Belgrade, Faculty of Organizational Sciences,

Jove Ilica 154, Belgrade, Serbia

www.datascienceforum.fon.bg.ac.rs

Book of Abstracts

Editors:

Boris Delibašić

Ljupčo Kocarev

UNIVERSITY OF BELGRADE
FACULTY OF ORGANISATIONAL SCIENCES

ISBN: 978-86-7680-327-9

2016 South East European Forum on Data Science (2016 SEE FDS), June 2016, Belgrade, Serbia

Publisher

University of Belgrade, Faculty of Organizational Sciences, Jove Ilića 154, 11000 Belgrade, Serbia

Editors

Boris Delibašić

Ljupčo Kocarev

Technical editor

Milan Vukićević

Cover and design

Milan Vukićević University of Belgrade, Faculty of Organizational Sciences, Jove Ilića 154, 11000 Belgrade, Serbia

CONTENTS

PREFACE

Data science is an interdisciplinary research area that includes methodologies of machine learning, data mining, pattern recognition, visualization, data engineering and other disciplines. In recent years, data science methods were successfully applied to a large range of areas including health, biology, climate, marketing, etc. The number of data science-related activities and publications in the US and in the North-West Europe has grown dramatically, but South-East Europe had fewer data science related activities. In Spring 2015 a successful one-day workshop on Mathematics of Data Science was organized at the Mathematical Institute of the Serbian Academy of Sciences and Arts.

This Book of Abstracts includes the seventeen invited lecturer talks presented at the South-East European Forum on Data Science.

The goal of 2016 South-East European Forum on Data Science is to bring data science researchers and experts from South-East Europe region together to discuss the current state-of-the-art, collaboration opportunities and challenges of data science research, as well as explore funding opportunities. The Forum will be an opportunity to business people to meeting highly-ranked data scientists from South-East Europe. The industry is invited to join the Forum's workshop, panel discussion, and to visit invited lectures of their interest.

You can find more information related to the Forum at the Forum webpage:
www.datascienceforum.fon.bg.ac.rs

Forum co-chairs

Prof. Dr. Boris Delibašić

University of Belgrade – Faculty of Organizational Sciences

Prof. Dr. Ljupčo Kocarev

Macedonian Academy of Sciences and Arts

FROM DATA MINING TO DECISION SUPPORT USING QUALITATIVE MULTI-CRITERIA MODELS

Abstract: *Data is an excellent source of information for decision support. Viewed as a historical record of past decisions, data can be analyzed in order to understand those decisions, to find out the main principles that governed them, and to develop models that can guide or predict future decisions. Unfortunately, models developed by machine learning algorithms from data often turn out not to be fully suitable for decision support, because they are, for instance, incomplete or inconsistent. To make them suitable for decision support, such models have to be verified, revised and/or extended by human experts, with the aim to improve their completeness, consistency and comprehensibility.*

In this talk, I will present an approach based on a qualitative, rule-based, hierarchical multi-criteria modelling method DEX. In the context of using data for decision support, this approach facilitates combining knowledge from different sources (such as data mining and expert modeling) and supports fulfilling the desired quality criteria for decision models. I will use real cases from the areas of food production, risk assessment and health-care management, to show: (1) the inadequacy of models developed only from data, (2) the need for their improvement through expert modelling, and (3) possible ways of adapting them for decision support through hierarchical multi-criteria models.

About the lecturer: Marko Bohanec is a leading Slovenian researcher in the field of decision support models and systems. He works at the Department of Knowledge Technologies at the Jožef Stefan Institute. Also, he is a professor of computer science at the University of Nova Gorica and a visiting professor at the University of Belgrade. He is one of the pioneers of qualitative multi-attribute modeling, which combines the fields of decision analysis and artificial intelligence to support people in making complex decisions. He is a co-author of the method DEX and computer program DEXi, which are used for decision support nationally and internationally. In the last ten years, he has been involved as a decision support expert in European projects on the production, distribution and quality control of food and feed (projects ECOGEN, SIGMEA, Co-Extra and DECATHLON), and disease management in health-care (PD_manager and HeartMan). He has co-authored more than 60 journal papers. According to Google Scholar his work has been cited about 2500 times.

Homepage

http://kt.ijs.si/MarkoBohanec/mare.html

DATA SCIENCE FOR CLIMATE AND ENVIRONMENTAL SCIENCES

Abstract: *Climate and Environmental Sciences are presenting exciting challenges for data sciences community to not only innovate on algorithms and computational frameworks, but also make a significant impact for the common good. In this presentation, I will outline some of the challenges in mobilizing data, information, networks, and knowledge to drive climate change adaptation and present our recent work in this space. I will also introduce the problem of invasive species management, and present our work on integrating diverse sources of data and innovations in network and data sciences to help take the important and critical steps in this direction.*

About the lecturer: Nitesh Chawla is the Frank M. Freimann Professor of Computer Science and Engineering, and director of the research center on network and data sciences (iCeNSA) at the University of Notre Dame. He started his tenure-track career at Notre Dame in 2007, and quickly advanced from assistant professor to chaired full professor position in 2016. He has brought in over $18.5M dollars in research funding to the university. He has received numerous awards for research, scholarship and teaching. He is the recipient of the 2015 IEEE CISOutstanding Early Career Award; the IBM Watson Faculty Award, the IBM Big Data and Analytics Faculty Award, National Academy of Engineering New Faculty Fellowship, and his PhD dissertation also received the Outstanding Dissertation Award. In recognition of the societal and community driven impact of his research, he was recognized with the Rodney Ganey Award and Michiana 40 Under 40. He is a two-time recipient of Outstanding Teaching Award at Notre Dame. He is a Fellow of the Reilly Center for Science, Technology, and Values;, Fellow of the Institute of Asia and Asian Studies; and Fellow of the Kroc Institute for International Peace Studies at Notre Dame. He is the founder of Aunalytics, a data science company.

Homepage

http://www3.nd.edu/~nchawla/

PROCESS ANALYTICS – MAKE IT WORK!

Abstract: *Business Process Analytics is a set of techniques that can be applied to event datasets created by logging the execution of business processes, and emerges as a promising decision support field. In this lecture, we will walk through a roadmap about delivering a process analytics project. After putting Process Analytics in context, we will discuss the main phases and the most frequent actions of a project, along with practical examples. Rather than presenting a generic methodology, the lecture aims at bringing forward common obstacles that researchers/practitioners meet, and discuss the relevant solutions. Finally, some considerations about the future trends conclude the talk.*

About the lecturer: Pavlos Delias is an Associate Professor of Management Information Systems at the Eastern Macedonia and Thrace Institute of Technology (formerly TEI of Kavala), Greece, and a visiting Professor at University Paris Dauphine ((PARIS IX). He holds a jointly supervised PhD (September 2009) in Informatics from both Technical Univesrity of Crete and University Paris Dauphine. His research interests are in the areas of process analytics, decision support, and multiple criteria analysis. He has contributed to numerous projects as a research associate, published several works, and he is member of the scientific organizations EURO and HAICTA.

Email
pdelias@teiemt.gr
Web-page
https://pavlosdeliassite.wordpress.com

MINING BIG AND COMPLEX DATA

Abstract: *Increasingly often, data mining has to learn predictive models from big data, which may have many examples or many input/output dimensions and may be streaming at very high rates. Contemporary predictive modeling problems may also be complex in a number of other ways: they may involve (a) structured data, both as input and output of the prediction process, (b) incompletely/partially labelled data, and (c) data placed in a spatio-temporal or network context. Data mining methods that can handle such problems are being developed within the EU funded MAESTRA project (http://maestra-project.eu/). The talk will first give an introduction to the different tasks encountered when learning from big and complex data. It will then present some methods for solving such tasks. It will focus on structured-output prediction, semi-supervised learning (from incompletely annotated data), and learning from data streams. Some illustrative applications of these methods will also be described.*

About the lecturer: Sašo Džeroski is a scientific councillor at the Jozef Stefan Institute and the Centre of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, both in Ljubljana, Slovenia. He is also a full professor at the Jozef Stefan International Postgraduate School and the Faculty of Computer and Information Science, University of Ljubljana. His research group investigates machine learning and data mining (including structured output prediction and automated modeling of dynamic systems) and their applications (in environmental sciences, incl. ecology, and life sciences, incl. systems biology). He has co-authored/co-edited more than ten books, including "Inductive Logic Programming", "Relational Data Mining", "Learning Language in Logic", "Computational Discovery of Scientific Knowledge" and "Inductive Databases and Constraint-Based Data Mining". He has participated in many international research projects and coordinated two of them in the past. He currently leads the FET XTrack project MAESTRA (Learning from Massive, Incompletely annotated, and Structured Data) and is one of the principal investigators in the FET Flagship Human Brain Project. Saso Džeroski received his Ph.D. degree in computer science from the University of Ljubljana in 1995. He was awarded a Jožef Stefan Golden Emblem Prize for his outstanding doctoral dissertation. Immediately thereafter, he received a fellowship from ERCIM, The European Research Consortium for Informatics and Mathematics, awarded to 5% of applicants. In 2008, he was awarded the title ECCAI fellow by the European Association for Artificial Intelligence (at that time called European Coordinating Committee on Artificial Intelligence) for "Pioneering Work in the field of AI and Outstanding Service for the European AI community". In 2015, he became a foreign member of the Macedonian Academy of Sciences and Arts.

Homepage

http://www-ai.ijs.si/SasoDzeroski/

MASSIVE GRAPHS: ALGORITHMS, TECHNIQUES AND CHALLENGES

Abstract: Recent advances in information and communication technologies have shown dramatic improvement and unparalleled development of our networked world. Almost all humans are interconnected and beyond that, all objects of our environment are becoming networked, forming huge global system. Massive graphs have been changing the traditional notion of efficient algorithms (in terms of processing time) from polynomial to sub-linear solutions by considering only a portion of the network. In this talk, I will overview algorithms, techniques and challenges for analyses of massive graphs by considering those methods and tools that allow massive graphs problems to be tackled not only on expensive supercomputing infrastructure, but also on more available, off-the shelf equipment.

About the lecturer: Ljupco Kocarev is a full professor at the Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University in Skopje, Macedonia, and research professor at University of California San Diego (UCSD), USA. His scientific interests include network science, machine learning, nonlinear systems and circuits; dynamical systems, mathematical modeling, and computational biology. He has co-authored more than 150 journal papers in 30 different international journals, ranging from mathematics to physics and from electrical engineering to computer sciences. According to Science Citation Index his work has been cited more than 6000 times, while according to Google scholar his work has been cited almost 12000 times. He is a fellow of IEEE, member of the Macedonian Academy of Sciences and Arts, and member of several other academies worldwide.

Homepage

http://www.cs.manu.edu.mk/people/faculty/ljupco-kocarev

A FRAMEWORK FOR DATA ENCRYPTION BASED ON JOINT EMPLOYMENT OF CRYPTOGRAPHY AND CODING

Abstract: *An important topic of Data Science is Data Security where data confidentiality appears as a very important issue. When a heavy employment of encryption is necessary, minimization of the overheads and fit into the implementation constraints are required which preserve cryptographic security as well. Accordingly, this talk addresses an approach for design of compact encryption which supports minimization of the overheads, fits into asymmetric implementation constraints and provides certain level of the provable security. The addressed approach is based on a combination of traditional encryption and coding in order to provide security enhancement of lightweight encryption algorithms which fits into the implementation constraints*

About the lecturer: Miodrag J. Mihaljević has received his B.Sc. and M.S. degrees in electrical engineering from University of Belgrade, Serbia (Yugoslavia), and received his Ph.D. degree in 1990. He is a Research Professor and the Projects Leader at the Mathematical Institute, Serbian Academy of Sciences and Arts, Belgrade, and serves as Deputy Director of the Institute. His main research areas are cryptology and information security. He has published more than 100 research papers in the leading international journals, books and conference proceedings (including over 50 papers in IEEE journals, Journal of Cryptology, Phys. Rev. A, Computing, IET Information Security, Inform. Process. Lett., LNCS, IEICE Transactions, and as certain book chapters), and over 200 publications in total. He is co-inventor of 6 granted patents in U.S, Japan and China. His research results have been cited more than 2000 times in the leading international publications. He has participated in over 10 international research projects and has served over 150 times as the reviewer for the leading international journals and conferences. He has held long-term visiting positions at the University of Tokyo, IMAI Lab (1997-2001 and 2004-2005), Sony Computer Science Labs (2001-2002), Sony Corporation Labs (2002-2003), Tokyo, the Research Centre for Information Security (RCIS), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan (2006-2012), Invited Senior Researcher at the Research Institute for Secure Systems (RISEC), National Institute AIST, Tsukuba, Japan (2012-2013), Invited Researcher and Professor at the Chuo University, IMAI Lab., Tokyo, Japan (2013-2014) and Project Professor at IIS, The University of Tokyo (2014-2016). Dr. Mihaljević is a recipient of the 2013 Award of Serbian Academy of Sciences and Arts for ten years achievements, and is an elected member of the Academia Europaea from 2014.

Homepage

http://www.mi.sanu.ac.rs/cv/cvmihaljevic.htm

DATAFLOW SUPERCOMPUTING FOR BIGDATA

Abstract: *This presentation analyses the essence of DataFlow SuperComputing, defines its advantages and sheds light on the related programming model. DataFlow computers, compared to ControlFlow computers, offer speedups of 20 to 200 (even 2000 for some applications), power reductions of about 20, and size reductions of also about 20. However, the programming paradigm is different and has to be mastered. The talk explains the paradigm, using Maxeler as an example, and sheds light on the ongoing research in the field. Examples include SignalProcessing, GeoPhysics, WeatherForecast, OilGas, DataEngineering, DataMining, etc. A recent study from Tsinghua University in China reveals that, for Shallow Water Weather Forecast, which is a BigData problem, on the 1U level, the Maxeler DataFlow machine is 14 times faster than the Tianhe machine, which is rated #1 on the Top 500 list (based on Linpack, which is a small data benchmark). Given enough time, the talk also gives a tutorial about the programming in space, which is the programming paradigm used for the Maxeler dataflow machines (established in 2014 by Stanford, Imperial, Tsinghua, and the University of Tokyo). The talk concludes with selected examples and a tool overview (appgallery.maxeler.com and webIDE.maxeler.com).*

About the lecturer: Prof. Veljko Milutinovic received his PhD from the University of Belgrade, spent about a decade on various faculty positions in the USA (mostly at Purdue University), and was a co-designer of the DARPAs first GaAs RISC microprocessor. Now he teaches and conducts research at the University of Belgrade, in EE, MATH, and PHY/CHEM. His research is mostly in data mining algorithms and data flow computing, with the stress on mapping of data analytics algorithms onto fast energy efficient architectures. For 7 of his books, forewords were written by 7 different Nobel Laureates with whom he cooperated on his past industry sponsored projects. He has over 40 IEEE or ACM journal papers, and about 4000 Google Scholar citations. Prof. Veljko Milutinovic is Life Member of the ACM, Fellow Member of the IEEE, Member of Academia Europaea, Member of the Serbian Academy of Engineering, Member of the Scientific Advisory Board of MindGenomics, Member of the Scientific Advisory Board of MaxelerTechnologies.

Homepage
http://home.etf.rs/~vm/

MATHEMATICAL DATA SCIENCE – OPPORTUNITIES AND CHALLENGES

Abstract: *In this talk, I will give an overview of the Mathematical Data Science program and describe the ONR funding model. I will discuss some of the scientific challenges that the program is addressing including: new tools for the analysis of large and complex datasets, importance of discovering causal dependences, and the reproducibility crisis in science. I will also discuss some of the new opportunities including: advances in human computing and collaboration, personalized learning, and collective decision making.*

About the lecturer: Dr. Pedja Neskovic is a program officer at ONR where he oversees the mathematical data science, and computational methods for decision making programs. He is also an adjunct associate professor of brain science at Brown University and a visiting professor at Johns Hopkins University. Within the scope of the mathematical data science program, he is addressing various basic research problems. These include methods for the analysis of big and small data, analysis of complex networks such as social and brain networks, reproducibility in science, and multi-modal and multi-scale information integration. He is also interested in developing methods for computational decision making that are utilizing novel crowdsourcing and collaborative techniques.

Homepage

http://www.onr.navy.mil/en/Science-Technology/Departments/Code-31/All-Programs/311-Mathematics-Computers-Research/Mathematics-Data-Science.aspx

PROPOSITIONAL FEATURE EXTRACTION USING RANDOM FORESTS AND DEEP NEURAL NETWORKS

Abstract: *Binary attributes are an important special case of attributes used in machine learning. In this talk we will present our ongoing research in feature extraction for such data. One proposed approach relies on transformation of tree models to propositional formulae and another on imposing propositional structure on deep neural networks by problem specific regularization. By imposing propositional structure on extracted features, we hope to improve interpretability of those features, and yet keep prediction quality. Our methods are primarily intended for analysis of medical healthcare record data in which binary attributes indicate presence or absence of the disease or indicate if a treatment has been performed on the patient. However, they are not limited to that specific domain.*

About the lecturer: Mladen Nikolić is an assistant professor of computer science at the Faculty of Mathematics of the University of Belgrade. He got his PhD in computer science from the University of Belgrade in 2013. His interests include machine learning, data mining, and automated reasoning. He worked as a visiting researcher at the Data Analysis and Biomedical Analytics (DABI) Center at Temple University for six months. He participated in several national and international research projects including two projects funded by Swiss National Science Foundation and one funded by DARPA.

Homepage
http://poincare.matf.bg.ac.rs/~nikolic/

PREDICTIVE MODELING OF INFORMATION NETWORKS BY EXPLOITING HETEROGENEOUS KNOWLEDGE WHILE LEARNING FROM DATA

Abstract: *In structured regression on networks, the response is predicted at multiple nodes given explanatory variables while taking into account the network's structure. This problem is challenging when there are multiple kinds of dependencies among nodes, some influences are negative and others are positive, uncertainties propagate for longer term predictions and representation of attributes and dependencies among nodes are learned jointly. Our solutions for these problems proposed at 4 papers published at AAAI-16 and SDM-16 will be discussed in this presentation.*

About the lecturer: Zoran Obradovic an Academician at the Academia Europaea (the Academy of Europe) and a Foreign Academician at the Serbian Academy of Sciences and Arts. He is a L.H. Carnell Professor of Data Analytics at Temple University, Professor in the Department of Computer and Information Sciences with a secondary appointment in Department of Statistical Science, and is the Director of the Center for Data Analytics and Biomedical Informatics. His research interests include data science and complex networks applications in health management and other complex decision support systems. Zoran is the executive editor at the journal on Statistical Analysis and Data Mining, which is the official publication of the American Statistical Association and is an editorial board member at eleven journals. He was general co-chair for 2013 and 2014 **SIAM International Conference on Data Mining and was** the program or track chair at many data mining and biomedical informatics conferences. In 2014-2015 he chaired the SIAM Activity Group on Data Mining and Analytics. His work is published in more than 320 articles and is cited about 16,500 times (H-index 48).

Homepage
http://www.dabi.temple.edu/~zoran/

SERVICE QUALITY ASSURANCE FOR CLOUD-BASED DATA-INTENSIVE APPLICATIONS

Abstract: *Since the software development market expects to be dominated by data-intensive cloud applications in the next years, there is now an urgent need for novel, highly productive, software engineering methodologies. Methods and tools that help satisfying quality requirements in data-intensive applications by iterative enhancement of their architecture design are under development in research projects like DICE (www.dice-h2020.eu). The latest proposes to monitored data acquired during testing and operation is deeply analysed to find quality pitfalls and outliers, which will lead to identify quality anti-patterns in the architecture and downstream design. The talk will be focused on the state-of-the-art and the DICE solution to the service quality assurance problem.*

About the lecturer: Dana Petcu (Mrs., PhD) is Professor at Computer Science Department and vice-rector responsible with international relationships at West University of Timisoara, scientific manager of its supercomputing center, and CEO of the research spin-off Institute e-Austria Timisoara. Her interest in distributed and parallel computing is reflected in more than two hundred papers about Cloud, Grid, Cluster or HPC computing. She is and was involved in several projects funded by European Commission and other research funding agencies, as coordinator, scientific coordinator, or local team leader. She is chief editor of the open-access journal Scalable Computing: Practice and Experience

Homepage
http://web.info.uvt.ro/~petcu

ACADEMIC VERSUS INDUSTRIAL RESEARCH IN MACHINE LEARNING

Abstract: *We face an increased trend of interest and competitiveness in machine learning, for both academic and industrial research. However, the aims are completely different: in industry, there is a constant pressure to augment the existing functionalities through smart services, while the academic research mainly targets development and study of specific building blocks. Practical examples will emphasize the gap between the two cultures. The talk touches the challenges of delivering the results in both academic and industrial environments. The issues exposed might provide potential hints for the academic curricula, aiming to speed up graduates' involvement in academic and/or industrial R&D, with a special emphasis on machine learning.*

About the lecturer: Lucian M. Sasu is associate professor at the Department of Mathematics and Informatics, Faculty of Mathematics and Informatics, Transilvania University of Brasov, Romania; he is teaching and doing academic research in this university since 2000, and he holds a Ph.D. degree in Computer Science since 2006. He also joined Siemens Corporate Technology Romania in 2012 in the R&D area, being involved in both FP7 and in internal industrial projects. His research focuses on machine learning – with an emphasis on incremental learning, pattern recognition and bio-informatics, and was published in peer-reviewed journals and conferences. He is a member of IEEE and Romanian Mathematical Society.

Homepage: https://www.researchgate.net/profile/Lucian_Sasu

TOWARDS DATA-DRIVEN PREDICTIVE MODELS IN HOSPITAL CARE

Abstract: *Electronic medical records and lately electronic health records serve as a rich source of data for analytical and research purposes. This talk will focus on two key ideas that allow the integration of data-driven predictive models in everyday hospital care. We will address the data privacy issues that can limit the performance of the developed predictive models due to limited sample sizes. On the other hand, once the large repositories of data are available we can target patient conditions more individually by searching for similar cases, retrospectively or in real time. Some characteristics of data and types of problems that can be addressed by personalized predictive models will be presented in the second part of the talk.*

About the lecturer: Gregor Štiglic is a Vice-Dean for Research, an Associate Professor and Head of Research Institute at the University of Maribor, Faculty of Health Sciences. He worked as a Visiting Researcher at the Data Analysis and Biomedical Analytics (DABI) Center at Temple University (2012) and as a Visiting Assistant Professor at Shah Lab, Stanford School of Medicine (2013). His research interests encompass application of knowledge discovery and predictive modelling techniques to large healthcare datasets. Specific areas of his technical interest include comprehensibility of classifiers, stability of feature selection algorithms, meta-learning and longitudinal rule discovery. His work was published in multiple conferences, journals and book chapters. Gregor gave talks on different topics from his knowledge discovery in healthcare and bioinformatics research at renowned research institutions such as IBM Watson Research Center, Stanford University and University of Tokyo. He currently serves as the Program Chair of the IEEE International Conference on Healthcare Informatics (ICHI) 2016 conference and as chair of the Data Mining for Medicine and Healthcare (DMMH) 2016 workshop.

Homepage
http://ri.fzv.uni-mb.si/gstiglic/

LINKED DATA BASED ANALYTICS FOR DRUG DATA ON A GLOBAL SCALE

Abstract: *The Web is currently comprised of vast amounts of heterogeneous data, most of which are unstructured or semi-structured. Health-care data are available as open data on the Web, and the integration and consolidation of such data can provide unprecedented results regarding price comparison of medicine products between countries, finding similar drugs and other relevant statistics, but the different structure of the data makes this a difficult task. The issue that impedes this task is that health-care data for most countries are merely published as 2-star open data, thus forming silos that don't intercommunicate. Additionally, there is not a central directory where this data can be found, so finding the right drug dataset for each world country by itself can be a challenging job.*

Our aim is to create a single dataset incorporating medicine products data from each country, using the Semantic Web technologies and Linked Data principles. These principles enable us to create a dataset of structured data, interlink them with other public health-care datasets and do this in an open way so that efforts were not duplicated. The dataset would allow users to access world drug data from a single endpoint and provide them with an opportunity to examine unseen statistics in the medical field previously. For this purpose, we have created an automated system for collecting the drug data for each country. So far we have gathered data from 25 countries, but our system is highly scalable, and we plan to incorporate data from all country where it is publicly available on the web. The drug data we use from each country is data extracted from medical databases from the official drug registry in the associated country. This system extracts, cleans and transforms the data from 2-star Open Data to 5-Star Linked Data. Also, we present a web application which manipulates the data from our dataset in order to provide end users with extensive information about the world drugs and useful statistics.

About the lecturer: Dimitar Trajanov Ph.D. is a professor at Faculty of Computer Science and Engineering – Cyril and Methodius University –Skopje. From March 2011 until September 2015, he was the Dean of the Faculty of Computer Science and Engineering, and in his tenure, the Faculty has become the largest technical Faculty in Macedonia. Dimitar Trajanov obtained his master and Ph.D. in Computer Science and Engineering from the "Ss Cyril and Methodius University" Skopje in 1998 and 2006 respectively. From 1999 to 2010 he was employed at the Faculty of Electrical Engineering and Information Technologies, first as a research and teaching assistant, and the as an Assistant Professor and form 2010 as an Associate Professor.

He is an author of more than 120 journal and conference papers and seven books. He has been involved in more than 40 international and national scientific and applicative projects as a project lead or a participant. Dimitar Trajanov is the leader of Regional Social Innovation Hub established as a cooperation between UNDP and the Faculty of Computer Science and Engineering. He is also, a member of the National Committee for Innovation and Entrepreneurship where he is working on creating strategies and legislation to encourage innovation and entrepreneurship in the Republic of Macedonia through inspiring the use of technology for economic development. His research interests include Semantic Web, Data Science, Big Data, Open Data, Social Innovation, Mobile Development, Ad hoc Networks, Parallel Processing, Reliability, System on Chip Design.

Homepage

http://www.finki.ukim.mk/en/staff/dimitar-trajanov

MULTI-TARGET PREDICTION AND APPLICATIONS IN THE PUBLISHING, ENERGY AND RETAIL INDUSTRIES

Abstract: *This talk is structured in two parts. In the first one, I will very briefly introduce the area of multi-target prediction, highlight its challenges and summarize some of our recent work on exploiting dependencies among multiple targets. In particular, I will briefly touch upon our work on: (i) discovering deterministic positive entailment and mutual exclusion relationships among multiple labels, representing these relationships as a Bayesian network and exploiting them for post-processing the output of multi-label models through probabilistic inference [1], and (ii) transferring successful multi-label classification approaches that use target variables as inputs (stacking, classifier chains) to the multi-target regression setting and extending them in order to deal with the discrepancy of the values of the target variables between training and prediction [2]. In the second part of the talk, I will discuss three ongoing data science collaborations with the private sector. In particular, I will talk about our work on: (i) semantic indexing of scientific publications [3], (ii) natural gas consumption forecasting, and (iii) business intelligence for the retail industry [4].*

References

1. Papagiannopoulou, G. Tsoumakas, I. Tsamardinos (2015) Discovering and Exploiting Deterministic Label Relationships in Multi-Label Learning. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15). ACM, New York, NY, USA, 915-924.
2. Spyromitros-Xioufis, G. Tsoumakas, W. Groves, I. Vlahavas (2016) Multi-Target Regression via Input Space Expansion: Treating Targets as Inputs. Machine Learning Journal.
3. Papanikolaou, D. Dimitriadis, G. Tsoumakas, M. Laliotis, N. Markantonatos, I. Vlahavas (2014) Ensemble Approaches for Large-Scale Multi-Label Classification and Question Answering in Biomedicine, Proceedings BioASQ 2014 Workshop, Sheffield, UK, 2014.
4. Fachantidis, A. Tsiaras, G. Tsoumakas, I. Vlahavas, (2016) Segmento: An R-based Visualization-rich System for Customer Segmentation and Targeting. In Proceedings of the 12th Hellenic Conference on Artificial Intelligence.

About the lecturer: Grigorios Tsoumakas is an Assistant Professor of Machine Learning and Knowledge Discovery at the Department of Informatics of the Aristotle University of Thessaloniki (AUTH) in Greece. He received a degree in computer science from AUTH in 1999, an MSc in artificial intelligence from the University of Edinburgh, United Kingdom, in 2000 and

a PhD in computer science from AUTH in 2005. His research expertise focuses on supervised learning techniques (ensemble methods, multi-target prediction) and text mining (semantic indexing, sentiment analysis, topic modelling). He has published more than 80 research papers and according to Google Scholar he has more than five thousand citations and an H-index of 30. Dr. Tsoumakas is a member of the Data Mining and Big Data Analytics Technical Committee of the Computational Intelligence Society of the IEEE and a member of the editorial board of the Data Mining and Knowledge Discovery journal. He is an advocate of applied research that matters and has worked as a machine learning and data mining developer, researcher and consultant in several national and private sector funded R&D projects

Homepage

http://intelligence.csd.auth.gr/people/tsoumakas

CHALLENGES OF BIG DATA ANALYTICS IN HEALTHCARE

Abstract: *Constantly increasing number of chronic diseases (such as diabetes, cardiovascular diseases, chronic renal failure, cancer etc.) and lack of doctors worldwide, led to the need for paradigm change from delayed interventional to Predictive, Preventive and Personalized Medicine (PPPM). Success stories of the Big Data paradigm and Predictive Analytics in many application areas led to the wide recognition of their high potential impact and benefits (both human and economic ones) in healthcare. However, there is still a large gap between actual and potential data usage in healthcare, because of numerous challenges: demand for highly accurate and interpretable models, high dimensionality, privacy concerns, the need for collaboration between domain experts and data scientists etc.*

In the first part of the talk, I will discuss the challenges and potential benefits of Big Data exploitation in healthcare. In the second part, I will present recent methodological approaches that were applied on real data and published in last several years. These approaches will include sparse predictive modeling, data and knowledge driven feature selection, "white-box" (or "glass box") algorithm design and meta-learning.

About the lecturer: Milan Vukicevic is an Associate Professor at the University of Belgrade, Faculty of Organizational Sciences. He worked as a Visiting Researcher at the Data Analysis and Biomedical Analytics (DABI) Center at Temple University (2014-2015). His research interests encompass design and development of predictive algorithms and their application in health care predictive modeling. Specific areas of his technical interest include sparse predictive models, fusion of domain knowledge and data-driven methods , parameter and feature selection, predictive analytics on heterogeneous data sources and meta-learning. His work was published in multiple conferences, journals and book chapters. Milan also had several invited talks on bioinformatics and healthcare predictive analytics topics.

Homepage

https://www.researchgate.net/profile/Milan_Vukicevic

DATA FUSION [OF EVERYTHING [FOR EVERYONE]]

Abstract: *In everyday life, people prefer to make decisions by considering all the available information, and often find that inclusion of even seemingly circumstantial evidence provides an advantage. We thought it would be great if a computational method could in a similar manner infer knowledge from large sets of coarsely related data. I will overview a recently proposed approach for large-scale data fusion that leaves all the data in its original domain space and requires little or no data engineering of the input. It assembles the data in a mosaic called data fusion graph and uses compression and chaining through the compressed system to yield a model wherein every piece of evidence counts, even if it is only distantly related to the prediction task. In the second part of the talk, I will introduce a visual programming tool that could be used to support data fusion and other data science tasks, and allow lay users to intuitively use otherwise computationally complex data science approaches.*

About the lecturer: Blaž Zupan is a professor at the Faculty of Computer and Information Science, University of Ljubljana, where he heads the Laboratory for Bioinformatics. He is also a visiting professor at Baylor College of Medicine in Houston, USA. His main research interests are data mining, machine learning, data fusion, and interactive data visualization. His lab develops Orange (http://orange.biolab.si), a popular open source data mining suite. Using visual programming, users of Orange can combine basic and advanced data science components and build powerful workflows for data analytics.

Homepage

http://www.biolab.si/en/blaz