SBORNÍK

prací účastníků vědeckého semináře
doktorského studia
Fakulta informatiky a statistiky
Vysoké školy ekonomické

Abstrakty

Vědecký seminář se uskutečnil dne 11. února 2021
pod záštitou děkana FIS
prof. Ing. Jakuba Fischera, Ph.D.

Sestavení sborníku
prof. Ing. Petr Doucek, CSc.
proděkan pro tvůrčí činnost a zahraniční vztahy

Nakladatelství Oeconomica – Praha 2021

ISBN 978-80-245-2410-8

OBSAH

Předmluva

„Den doktorandů“ patří mezi tradiční akce, které Fakulta informatiky a statistiky pořádá pro studenty doktorského studia. V tomto roce se jednalo již o dvacátý šestý ročník. Seminář se konal 11. února 2021 pod gescí děkana Fakulty informatiky a statistiky prof. Ing. Jakuba Fischera, Ph.D. Vzhledem k obtížným podmínkám, které s sebou přinesla epidemická situace proběhla některá vystoupení on-line. I hodnoticí komise „zasedaly“ za podpory informačních technologií. Tento fakt ovšem nenarušil kvalitu vystoupení doktorandů, byť pro mnohé z nich to bylo první vystoupení před odbornou veřejností, na němž získávají cenné zkušenosti a tříbí si tak i formulace názorů a hypotéz. Kromě toho si doktorandi vyzkoušeli presentaci závěrů výzkumné práce a argumentaci na jejich podporu.

V letošním roce byly příspěvky, vzhledem k celkovému počtu třinácti přihlášených účastníků ze všech studijních programů doktorského studia, rozděleny do dvou sekcí – Aplikovaná informatika a Kvantitativní metody – studijní programy Ekonometrie a operační výzkum a Statistika.

Nedílnou součástí „Dne doktorandů“ je i práce hodnoticích komisí, jejichž členové pečlivě sledují jednotlivá vystoupení a potom vybírají nejlepší práce k ocenění. Hlavními kritérii pro jejich rozhodování byly zejména kvalita a aktuálnost zpracovaného tématu, přístup k řešení vybraného problému, způsob použití metodiky, úroveň práce s reálnými daty a v neposlední řadě i schopnost prezentovat a argumentačně své výsledky obhájit v diskusi. Ti nejlepší z účastníků získávají prestižní „Cenu děkana FIS“, s níž je spojena i symbolická finanční odměna.

Za práci v hodnoticí komisi studijního programu Aplikovaná informatika chci poděkovat všem jejím členům – prof. Ing. Vojtěch Svátek, Dr. (KIZI), Mgr. Ing. Zdeňku Smutnému, Ph.D. a Ing. et Ing. Michalu Doleželovi, Ph.D., za práci v hodnoticí komisi pro Kvantitativní metody – studijní programy Ekonometrie a operační

výzkum a Statistika pak prof. Ing. Josefovi Jablonskému, CSc. (KEKO), prof. Ing. Haně Řezankové, CSc. (KSTP) a doc. RNDr. Ivaně Malé, CSc. Obě komise se zhostily své práce na výbornou.

V letošním roce získali ceny za nejlepší příspěvky následující studentky a studenti:

Studijní program – Aplikovaná informatika
1. místo: Ing. Lukáš Švarc: Anomaly Detection Algorithms in University Environment
2. místo: Ing. Pavel Strnad: Towards complex anomaly detection algorithms
3. místo: Ing. Tereza Zichová: COVID-19 Czech Financial e-Administration Innovation Challenge

Studijní programy – Ekonometrie a operační výzkum a Statistika
1. místo: Ing. Jiří Koudelka: Extreme value theory for capital requirement calculation in Insurance
2. místo: Ing. Bc. Petra Zýková: Dynamic DEA models
3. místo: Ing. Filip Habarta: Simulation of impact of changes in debtors' behavior within modeling the time to loan repayment

Oceněným studentům doktorského studia upřímně blahopřeji a pevně věřím, že získané zkušenosti uplatní při své další práci, ať už vědecké nebo v praxi. Uznání také patří všem vědeckým a pedagogickým pracovníkům FIS – školitelům doktorandů, kteří se „Dne doktorandů“ zúčastnili a svým vedením a radami byli nápomocni při zpracování příspěvků.

Zvláštní poděkování pak patří studijní referentce doktorského studia paní Jitce Krajíčkové, díky níž byl seminář skvěle organizačně zajištěn, dále paní Petře Šarochové za administrativní podporu akce a Mgr. Lee Nedomové za práci při editaci a sestavení tohoto sborníku abstraktů.

prof. Ing. Petr Doucek, CSc.

proděkan pro tvůrčí činnost a zahraniční vztahy

STUDIJNÍ PROGRAM
APLIKOVANÁ INFORMATIKA

Comparison of selected enterprise architecture modeling techniques from the perspective of IT services

Jan Buriánek

jan.burianek@vse.cz

Ph.D. student of Applied Informatics

Supervisor: prof. Ing. Václav Řepa, CSc. (repa@vse.cz)

Focus of this paper is set on a mutual comparison of selected popular enterprise architecture modeling techniques from the perspective of IT services. Particular frameworks and notations that has been researched are ArchiMate, Unified Architecture Framework, SoaML, NATO Architecture Framework and Unified Modeling Language. In order to compare and evaluate these techniques, a method presented by Framework for Evaluating BPM/ISM Techniques has been utilized. This method suggests evaluating modeling techniques by their breadth (typical modeling goals) and depth (modeling perspectives). Furthermore, notations used by techniques in focus have been analyzed and compared with standard ISO 20000 used as a reference point. According to the methods used, Unified Architecture Framework was classified as the most versatile and comprehensive enterprise architecture modeling technique among researched frameworks and notations.

Keywords: conceptual modeling, enterprise architecture, IT service, modeling notation, UML, SoaML, UAF, ArchiMate, NAF, Framework for Evaluating BPM/ISM Techniques

JEL Classification: L86, M15

MIRAQL – a collaborative approach for the agile development of OWL ontologies

Viet Bach Nguyen
nguv03@vse.cz
Ph.D. student of Applied Informatics
Supervisor: prof. Ing. Vojtěch Svátek, Dr., (svatek@vse.cz)

This paper deals with the review of ontology editors and the problems with large ontology development projects where a collaborative approach is required. Based on the reviewed features of ontology editors and the new requirements coming from the collaborative environment of ontology development, a new systematic approach for Git-based management of OWL ontologies is proposed to streamline the development and negotiation process between ontology engineers and domain experts. The system makes use of several Semantic Web technologies such as RDF triplestores, OWL API, graph model API, SPARQL Update query language in combination with several DevOps techniques to create a comprehensive workflow that enables code-review-based evolution for large, modularized, and multilingual ontology projects with high update frequency that require the collaboration between domain experts and ontology engineers.

Keywords: continuous deployment, continuous integration, ontology development, ontology editor, OWL, RDF, SPARQL, version control

JEL Classification: L86

Towards complex anomaly detection algorithms

Pavel Strnad

pavel.strnad@vse.cz

Ph.D. student of Applied Informatics

Supervisor: prof. Ing. Petr Berka, CSc. (berka@vse.cz)

This paper introduces recently developed synthetic dataset generator and its use case as an anomaly detection algorithm benchmarking tool. Generator was developed in cooperation with other colleagues in project IGA 12/2019. Generator contains anonymized data from Prague University of Economics and Business information system logs and is able to scale these data in terms of time and also perform injection of data with attackers behaviour patterns. Anonymized data still contain user behaviour patterns therefore individual anomalous behaviour could be detected. Different types of real attack behaviour patterns from university environment were selected and these are used to demonstrate attackers behaviour in synthetically created system logs. Mentioned features allow other researchers to benchmark their anomaly detection algorithms with complex data. Because the more complex a model is the more data is needed.

Keywords: anomaly detection, synthetic dataset generator, machine learning

JEL Classification: C61

Web security perception of educational institutions in V4 countries

Jaroslav Svoboda

jaroslav.svoboda@vse.cz

Ph.D. student of Applied Informatics

Supervisor: doc. Ing. Stanislav Horný, CSc.
(stanislav.horny@vse.cz)

ICT security is a highly complex topic which for non-ICT specialist is quite challenging to grasp. As this contribution shows, on the example of school headmasters and their deputies, the opinion of a manager can be quite far from the truth. Based on website scan and subsequent survey addressed to the management of schools providing ISCED level 0-3 education, this article presents the data describing people managing schools and their opinions about their ICT security skills and the security of the school website. This is put in contrast with the scan results.

The personality of the school manager was created from a survey of 1630 respondents. The average school manager in V4 countries is approximately 51 years old, with a background in pedagogy, with a positive attitude towards ICT and considers himself/herself quite skilled in the ICT security area. The manager thinks the school website is relatively well managed, maintained and secured. This is in direct contrast with the security scan which found at least one significant security issue on more than 80 % of websites. The presence of on-campus ICT employee or employees is relatively insignificant. Since all the tracked security issues are not a matter of software or hardware cost, this leads to the belief that the problem is the matter of awareness.

Keywords: security, web, education, V4, school, HTTPS, CMS

JEL Classification: L86

Anomaly Detection Algorithms in University Environment

Lukáš Švarc

lukas.svarc@vse.cz

Ph.D. student of Applied Informatics

Supervisor: prof. RNDr. Jiří Ivánek, CSc., (jiri.ivanek@vse.cz)

This paper focuses on anomaly detection approaches based on machine learning techniques. It covers a description of anomaly detection algorithms used in this research. These approaches can be divided into two groups - supervised and unsupervised. However, some methods combine both approaches. Several different supervised and unsupervised approaches were benchmarked. To validate usefulness of specific methods in university environment two different datasets were used and results from them were compared. First experiment was performed on publicly available KDDCUP'99 dataset, specifically on password guessing on terminal connection. This dataset was prepared by capturing nine weeks of LAN traffic in U.S. Air Force network and although it contains about five million connection records it has been striped down to telnet traffic only for purposes of this research. In order to be able to compare its results to university environment, a dataset from Prague University of Economics and Business was prepared with other colleagues from IGA project. This dataset contains traffic from information system InSIS and includes several different types of attacks. Several supervised and unsupervised algorithms were benchmarked using this dataset and results were compared to KDDCUP'99 dataset results. All experiments were run on BigML platform, which is robust cloud-based platform that offers Machine Learning as a Service (MLaaS). In university environment there was a novel method proposed for automated attacks detection. Using BigML framework in combination with real university dataset, K-means,

Isolation Forest and Logistic Regression algorithms were used in order to validate proposed novel method. Results verified that approaches that work well in general are useable in university environment as well if parameters are set precisely.

Keywords: Anomaly Detection, Machine Learning, Network Attacks, University Environment, BigML

JEL Classification: Y40, C60, C61

Selection of IT Processes Within COBIT 2019 for Process Mining Utilization

Ivan Vassilyev

vasi00@vse.cz

Ph.D. student of Applied Informatics

Supervisor: doc. Ing. Vlasta Svata, CSc. (svata@vse.cz)

The article discusses the problems of modern IT audit, in particular, the limited amount of information for making auditor´s decision or expressing judgment in an extremely short time. To solve the problem, the use of process mining methods is proposed. Based on system event logs from Process-Aware Information Systems, process mining techniques allow IT auditors and process owners to automatically discover and simulate real IT processes in an enterprise and to obtain consolidated information on processes, such as process performance indicators.

The COBIT 2019 framework was chosen as a canonical model of IT processes. The purpose of the article is to find out exactly which IT processes described and standardized in the COBIT 2019 framework are suitable for the implementation of process mining methods.

To achieve this goal, an algorithm was developed for selecting IT processes and one-to-one corresponding governance and management objectives suitable for further implementation of process mining techniques. The developed algorithm is based on the collection, analysis and subsequent systematization of disparate data on standardized IT processes and instrumentaria used to perform these processes according to COBIT 2019.

The article presents an empirical approach to determining criteria on the basis of which it is proposed to decide whether a

governance component of the "Services, infrastructure and applications" type meets the necessary and sufficient conditions for process mining methods utilization. Gradual elimination method was applied.

Keywords: Audit Software, IT Audit, COBIT 2019, Enterprise Governance of IT, Process Mining, Process-Aware Information Systems

JEL Classification: M15, M42

Intelligent Data Structuring for Facility Management as Basis for Smart Applications

Adrian August Wildenauer

wila03@vse.cz

Ph.D. student of Applied Informatics

Supervisor: prof. Ing. Josef Basl, CSc. (josef.basl@vse.cz)

The thesis combines the structured ordering and practical use of data in the construction industry on a long-term perspective. It transfers the principles of Industry 4.0 and the cyber-physical foundations inherent in it to the not digitally mature construction industry. It deals with the management and consistent further use of information in digital building models. These shall be used in the early planning and realisation phase of construction projects, but also in later use in the operational phase which represents approximately 80% of the life cycle. This is particularly important from the point of view of an owner or client (hereafter owner) of a facility to be built who has to maintain buildings usually for decades. The purpose of this thesis is to identify the necessary structuring framework for the use of practical applications in the operational phase without having to subsequently record this data after completion of an asset. This can be, for example, the provision of "smart applications" in which the user of an office building is shown the next available meeting room, the way there and all necessary information such as the last disinfection, cleaning and equipment in real time. The data ordered so far in a project, that has not been consistently implemented digitally, is not sufficient for this, as it is not standardised and mostly done on a project-specific base.

For the purpose of this thesis, the approaches of information management by using digital building modelling (Building

Information Modelling BIM) are compared to previous implementations of other industries. BIM has been widely used in the European construction industry since about 2015. An end-to-end solution for information management is therefore only in development in the industry. However, BIM has very significant parallels to the digital twin ecosystems of the manufacturing industry (Industry 4.0) and the consequent use of cyber-physical systems. Industry 4.0 requires a continuous flow of data from the project idea to the demolition, re-use or recycling – and so does the owner of a built asset. However, due to the phase- and person-driven construction industry, a holistic life cycle based data management is mostly not existing and therefore no starting point available. Owners are faced with the dilemma of missing data, although they have ordered an error-free, complete work including a digital representation of a project.

The approach of this work is structured into four interdependent parts. As a first step, existing literature and research work were reviewed. The topic has not yet become the focus of scientific work. Only a few works have so far addressed the specifics of the construction industry with regard to information management using BIM. As a second step, existing three-dimensional building models of projects in realisation from the Czech Republic, Germany and Switzerland were examined with regard to existing and necessary data fields. The received 15 projects are counting to approximately 4% of the current projects dealing with information management based on Building Information Modelling in Switzerland and Czechia. It turned out that these digital models were primarily created for evaluating options, alternative and variants for the planning and realisation phase. Data fields or special attributes for the significantly longer operational phase of a building were usually neither specified in project documents nor taken into account in the creation of digital models nor mentioned in handover records. Due to the prevailing sequential processing of tasks in the construction industry, among other things, the needs of facility management are usually only retrieved very belatedly or not at all. To identify these necessary data fields, 15 interviews

have been conducted so far with proven experts in the digital transformation of the construction industry. The aim was to identify and verify necessary data fields and to elicit necessary data governance topics. Their input built the base for the third step, the creation of an online questionnaire with which the statements made there will be verified. So far, 132 persons have participated in this questionnaire, which was spread via professional associations and institutions, direct contacts and social networks. As a fourth point, the data mentioned will be examined in a real construction project (demo project) and tested on a use case, the automatic availability of meeting rooms. A necessary framework will be created with the aim is to test the framework on a real project and to make suggestions for further use.

Keywords: Building Information Modelling, BIM, data structuring, facility management, construction

JEL Classification: L74, N64, R33

COVID-19 Czech Financial e-Administration Innovation Challenge

Tereza Zichová

tereza.zichova@vse.cz

Ph.D. student of Applied Informatics

Supervisor: doc. Ing. Stanislav Horný, CSc.
(stanislav.horny@vse.cz)

The paper deals with the analysis and evaluation of an innovative approach in the electronic communication of the Czech Financial Administration related to the needs that suddenly arose from the COVID-19 pandemic. There has been a significant change in the use of information technologies, the potential of which has not yet been realized, which has prevented the streamlining of services offered to taxpayers. An important part of the study is the comparison and evaluation of digital interaction by the electronic society before and after the outbreak of the coronavirus crisis. The following electronic Tax Administration form filling interfaces are chosen for detailed comparison: a) before the pandemic – Tax Portal, b) during the pandemic – application for the Provision of a compensation bonus for self-employed persons (so-called Twenty-five for self-employed persons). The comparison is based on a quantitative questionnaire survey composed of questions, the evaluation factors of which are based on previous interviews within the qualitative approach to the study of phenomena in society. The purpose of the article is to verify by means of a paired t-test the hypothesis that there is a different perception of the user environment of the Financial Administration applications offered before the pandemic and developed during the pandemic. The article analyzes these changes in the discussion, puts them into context and deduces from the results of the study the possible

contribution to the future direction of online communication of the Financial Administration.

Keywords: e-administration, COVID-19 innovation, financial administration, e-governance, e-government innovation

JEL Classification: H11, O38

STUDIJNÍ PROGRAM EKONOMETRIE A OPERAČNÍ VÝZKUM

Comparison of Optimal Combination Approaches for Hierarchical Time Series Forecasting

Jiří Georgiev

jiri.georgiev@vse.cz

Ph.D. student of econometrics

Supervisor: prof. RNDr. Václava Pánková, CSc.
(vaclava.pankova@vse.cz)

Hierarchical time series, represents aggregation structures between multiple time series and are commonly present among macroeconomic time series. Traditional forecasting approaches for hierarchical time series provide coherent forecasts. These approaches are based on forecasts only for one level of aggregation structure and therefore are omitting relationships between series and other valuable information. A new approach called MinT aims to rectify this issue as it requires all series from the hierarchy to be forecasted. This contribution evaluated different MinT approaches on data for GDP of four European countries: Germany, Italy, Spain and Poland. The MinT (Shrink) variant seems to provide the best results across evaluated countries and aggregation levels.

Keywords: hierarchical time series, forecasting, reconciliation

JEL Classification: C53

Dynamic DEA models

Petra Zýková

petra.zykova@vse.cz

Ph.D. student of econometrics

Supervisor: prof. Ing. Josef Jablonský, CSc., (jablon@vse.cz)

This paper deals with dynamic efficiency analyses based on Data Envelopment Analysis (DEA) models. The aim is to formulate new dynamic DEA models with time series, which compute the unit's overall efficiency with the impact of all units' inputs and outputs in all periods. Many researchers have studied DEA models. This paper contributes to the DEA broad theory. The proposed models are compared with already known models invented by Park and Park (2009). We introduce four new dynamic DEA models with quadratic objective function and nonlinear constraints that differ in every unit's time weights every year. The first proposed model has a decreasing vector of the weights; the second model uses the weights' ratio scale. These models divide the set of units into two subsets: efficient and inefficient units. The inefficient units can be easily rank according to the efficiency scores. The efficient units have identical efficiency score equal to one and cannot be raked. Therefor two super-efficiency models are proposed. The super-efficiency models compute the super-efficiency scores greater than one for the efficient units, and the efficient unit can be ranked according to this super-efficiency score.

All models are illustrated and compared on a dataset, and then their results are discussed. The dataset contains 38 German NUTS 2 (Nomenclature of Territorial Units for Statistics) regions. The aim is to find the most efficient regions and their ranking between the years 2008 – 2016. The dataset contains information about the NUTS 2 regions for nine following years starting by 2008. There are used two inputs – employment (thousand hours worked) and

gross fixed capital formation (million euros) and one output – gross domestic product (million euros). All calculations are realised using original procedures written in the LINGO modelling language.

Keywords: DEA models, dynamic models, efficiency analysis, time series

JEL Classification: C44, C61

STUDIJNÍ PROGRAM
STATISTIKA

The possibility of using minimum spanning tree in cluster analysis

Jakub Danko

jakub.danko@vse.cz

Ph.D. student of Statistics

Supervisor: doc. Ing. Diana Bílková, Dr. (diana.bilkova@vse.cz)

The paper deals with the design of a method for classifying objects into a predetermined number of groups using the minimum spanning tree of a graph. Using this tool of discrete mathematics, we propose a clustering algorithm that tries to minimize the total distance in this spanning tree under predetermined conditions. The obtained results are compared with other known hierarchical clustering methods, while for correct comparison the same distance metrics are considered when defining the distance between objects. Specifically, we considered the often-used Euclidean distance, from which we applied a complete graph by applying graph theory, whose edge evaluations represent the distances between individual objects. From this graph, we then calculated the minimum spanning tree of the graph, which was the basis for the application of our proposed algorithm. In this paper, we work with the well-known dataset Iris suitable for evaluating the success of classification methods.

Keywords: minimum spanning tree, clustering, Iris, Euclidean distance

JEL Code: C38

Simulation of impact of changes in debtors behaviour within modelling the time to loan repayment

Filip Habarta

filip.habarta@vse.cz

Ph.D. student of Statistics

Supervisor: doc. RNDr. Ivana Malá, CSc., (malai@vse.cz)

Developed financial markets with credit loans are extremely competitive, take for example current Czech mortgage market, such competition puts even more pressure on financial institutions if they want to be successful in the retail business with credit loans. One of the keys of success is to be able to properly estimate risk arising from each provided loan which is covered mainly by the default of issued loans. In this paper, we present construction of a "clock-forward" markovian multi-state model which captures client movements between states of regular loan repayment, payment delays and default. The focus is on the clients' unable to meet regular payments whose loans end in default. Loan default necessarily does not need to lead to the total loss of provided credit for the lender but brings at least delay before obtaining provided funds back from the client, thus the time to default plays a crucial role in the whole time to loan repayment. To obtain the estimate of the model open data from peer to peer lending platform Bondora were used. Leveraging the estimated model, we further show the expected development of client-specific default in time for a simulated portfolio of clients. Such portfolio simulations can be further used for better portfolio management in financial institutions.

Keywords: Multi-state model, Time to loan repayment, Default modelling, Cox proportional hazard model, Time dependent covariates

JEL Classification: C14, C15

Extreme value theory for capital requirement calculation in Insurance

Jiří Koudelka

kouj04@vse.cz

Ph.D. student of Statistics

Supervisor: doc. RNDr. Luboš Marek, CSc., (marek@vse.cz)

The aim of this case study, based on the real unique dataset of operational risk claims of an anonymous Central and Eastern European insurance company, is to examine the sufficiency of the Standard Formula. We applied Extreme Value Theory to build up a more complex model for estimation of operational risk events, using the available historical claims portfolio. The dataset was split into separate parts using the selected threshold and for each group i.e., for the body and the tail the best frequency and severity distribution was chosen. The frequency distribution was found at first. The number of claims follows the Poisson distribution. As a result of previous analysis, the outliers over the threshold follow Frechet distribution and GPD. Claims below the threshold are suited to commonly used distributions and the Lognormal distribution came out as the best. Our findings show applicability of Extreme Value Theory to real claim portfolio. The solvency capital requirement for operational risk events based on the Standard Formula may not be sufficient in comparison to the Extreme Value Theory approach. Thus, the insurance company with a specific risk profile using Standard Formula may be exposed to higher potential losses.

Keywords: Operational risk, Extreme Value Theory, Solvency Capital Requirement, Value at Risk, Generalized extreme value distribution, Bootstrapping

JEL Classification: C46, G22