General Sponsors

Research Council of Lithuania

algoritmų sistemos

Main Sponsor

Critical IT infrastructure

balp

Sponsors

NRD Cyber Security

NetCode

BALTIC AMADEUS

LIKS

LITHUANIAN COMPUTER SOCIETY

VILNIUS UNIVERSITY

INSTITUTE OF DATA SCIENCE AND DIGITAL TECHNOLOGIES

LITHUANIAN ACADEMY OF SCIENCES

DAMSS

DATA ANALYSIS METHODS FOR SOFTWARE SYSTEMS

9th International Workshop on

DATA ANALYSIS METHODS FOR SOFTWARE SYSTEMS

Druskininkai, Lithuania, Hotel "Europa Royale"

http://www.mii.lt/DAMSS

November 30 –December 2, 2017

Contacts:

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

DOI: http://dx.doi.org/10.15388/DAMSS.2017

ISBN 978-9986-680-64-2

Preface

DAMSS-2017 is the 9th international workshop on data analysis methods for software systems, organized in Druskininkai, Lithuania, at the end of the year.

History of the workshop starts from 2009 with 16 presentations. The number of this year presentations is 68. The idea of such workshop came up at the Institute of Mathematics and Informatics. Lithuanian Academy of Sciences and the Lithuanian Computer Society supported this idea. This idea got approval both in the Lithuanian research community and abroad.

This year is special for the initiator of DAMSS conferences. On 2010, the Institute of Mathematics and Informatics became a member of Vilnius University, the largest university of Lithuania. Now, the institute changes its name into the Institute of Data Science and Digital Technologies. This name reflects activities of the institute much better. The renewed institute has seven research groups: Cognitive Computing, Image and Signal Analysis, Cyber-Social Systems Engineering, Statistics and Probability, Global Optimization, Operations Research, Education Systems.

The workshop has several main goals: to facilitate scientific networking that will lead to joint research projects and to initiate connections with foreign research institutions and scientists. Seeking to facilitate relations between science and business, computer science and IT business community will be introduced with research undertaken at Lithuanian and foreign universities in the fields of data science and software engineering.

Topics of the conference cover big data, bioinformatics, data science, deep learning, digital technologies, high-performance computing, visualization methods for multidimensional data, machine learning, medical informatics, ontological engineering, optimization in data science, business rules, and software engineering.

This book gives an overview of all presentations of DAMSS-2017.

Heuristic Extension for Multi-Criteria Decision Making in Autonomous Ground Robot Navigation

R. Baušys, R. Semėnas

Vilnius Gediminas Technical University
rokas.semenas@vgtu.lt

Autonomous ground robot navigation is an important research field that has many real-life applications. Over the past decade human-uncontrolled search and rescue, exploration and area mapping problems grew in numbers, but there is still a lack of robust and optimal algorithms to tackle them.

Many different approaches were suggested, and among them Multi-criteria Decision Making methods show feasible results and huge potential, considering the amount of different parameters that can be presented to the robot. In this paper, the main algorithms and problems of so-called, state of the art methods are presented and discussed, several possible heuristics to extend them are proposed.

Combined Permutation Tests and Finite Sample Consistency

Stefano Bonnini

Department of Economics and Management
University of Ferrara
stefano.bonnini@unife.it

In several applications, the dataset presents a large number of variables and small sample sizes. A typical solution for a location problem where two multivariate distributions are compared, is the Hotelling's T-square test. It requires, among the other assumptions independent samples and multivariate normality. In the presence of small sample sizes, Hotelling's T-square test is not very powerful, and it cannot provide a suitable solution for one-sided (directional) alternative hypotheses.

Combined Permutation Tests represent a nonparametric proposal in order to deal with a large number of variables, especially in the presence of small sample sizes. This testing method mainly requires exchangeability under H0, and it is distribution free. Thus it is much more flexible and robust than other testing methods, especially parametric solutions (e.g., Hotelling's T-square test). This method is suitable for both two-sided and one-sided alternatives and with numeric, categorical or ixed variables.

Finite sample consistency is one important property of this methodology. According to this property, under mild conditions, the power of the test increases with the number of variables under study. The result holds even when the number of variables is much larger than the sample sizes.

It has been proved that the, for the multivariate two-sided problem, the combined permutation test is more powerful than the Hotelling T2 test. Some Monte Carlo simulation studies prove that similar results hold under Normal, Cauchy, Student's t and Pareto distributions. Thus it is robust with respect to the underlying multivariate distribution. The test is well approximated under H0, unbiased and consistent (in the classic sense) when sample sizes diverge.

We present some advances in the study of the mentioned methodology and property, through the results of a simulation study and an application example.

An Investigation of Early Cyber Threat Detection Using Ensembles of Machine Learning Methods

V. Bulavas, G.Dzemyda, V. Marcinkevičius

Institute of Data Science and Digital Technologies
Vilnius University
viktoras.bulavas@ittc.vu.lt

According to PwC's Global economic crime survey, cybercrime has overall evolved into the second place after an asset misappropriation. According to Lithuanian National Cyber Security Centre Annual report for 2016, scanning of surveilled network devices since 2015 has increased fivefold. Lithuanian academic network LITNET is no different, observing persistent multiple step intrusion activities.

As nowadays it is impossible to detect and mitigate all threats manually, automatic tools are used on a 24/7 basis. The techniques utilized by current network intrusion detection appliances in use fall into three main categories: anomaly detection, misuse detection, and hybrid. Misuse detection systems use signatures that describe already known attacks and require regular ruleset update.

Machine learning based anomaly detection requires supervision and specialist review due to currently still high false positive rate of detecting previously unseen system behaviors. With a n increasing frequency of cyber-attack, reviews take more and more time of cyber security specialists, which is a challenge. This indicates the highly demanded area for research aiming to increase threat detection accuracy and training