

Traceable Use of Emerging Technologies in Smart Systems

J.UCS Special Issue

Wolfram Luther

(University of Duisburg-Essen, Duisburg, Germany,

 <https://orcid.org/0000-0002-1245-7628>, wolfram.luther@uni-due.de)

Gregor Schiele

(University of Duisburg-Essen, Duisburg, Germany,

 <https://orcid.org/0000-0003-4266-4828>, gregor.schiele@uni-due.de)

Keywords: AI-based data mining, application troubleshooting, authentication, automated root cause analysis (RCA), BRCA1/2 mutation probability, cloud infrastructures, cloud-native applications, cyclical patterns, data fusion & integration, Dempster-Shafer evidence theory (DST), distance learning, distributed tracing, explainable AI (XAI), hypothesis testing, interval analysis, key performance indicators (KPI), machine learning (ML), project-based learning, RED metrics, rule-induction systems, sales goals and operation planning, security, social networks, spatio-temporal data, video activity report, work engagement

Categories: B.8, E.2, E.4, G.1, G.3, H.3.1, H.3.2, H.3.3, H.3.7, H.4.2, H.5.1, H.m, I.1.2, I.2, I.2.6, I.5.3 I.5.4, I.6.4, J.3, K.6, L.1.2. L.3.6

DOI: 10.3897/jucs.112577

This volume presents a selection of invited papers from the 3rd Workshop on Collaborative Technologies and Data Science in Smart City Applications (CODASSCA 2022): *From Data to Information and Knowledge*, held in Yerevan, Armenia, August, 23-25, and further articles from a free call for papers JUCS-CODASSCA-2023 published by EasyChair. The workshop continues the cooperation between the University of Duisburg-Essen (UDE) and the American University of Armenia (AUA) funded by the German Academic Exchange Service (DAAD) and the German Research Foundation (DFG). The workshop took place together with a one-week summer school on the topic *Enhancements of Deep Learning for Intelligent Applications and the Connected Society*.

In two rounds of review, 15 of the papers submitted in the three formats of extended abstract, short paper, and full paper were selected by the program committee and published in conference proceedings *Data Science, Human-Centered Computing, and Intelligent Technologies* released with a CC BY license by Logos, Berlin 2022, ISBN 978-3-8325-5520-7, <https://doi.org/10.30819/5520>.

Authors submitted significantly extended and improved versions of their contributions to be considered for a J.UCS special issue *Traceable Use of Emerging Technologies in Smart Systems*. These articles are grouped in two thematic areas:

- *Data science with intelligent technologies;*
- *Human-centered computing with intelligent technologies.*

There was also a J.UCS open call so that any author could submit papers on the highlighted subjects. The invitation to review the 20 contributions received was ac-

cepted by 22 experts, and, after three rounds, eight articles were finally accepted for publication in the special issue.

This special volume aims to show ways in which scientists and users can jointly move beyond the current practice of publishing scientific results to more traceable methods and results that facilitate understanding of the goals defined and technologies used to achieve the results and their explainability and validation, that is, to move from "results" to "Xresults" (i.e. the traceable results and explainable technologies used).

The paper written by members of the Armenian National Academy of Sciences is about the two stages of tracing. First, Bob decides that the data is a stegotext and uses the extraction algorithm to identify the hidden message. Then, he uses the key to decide whether it was Eve or Alice who was active.

A group of Armenian and American scientists and developers discuss the application of rule-learning approaches to automated, real-time, and intelligent root cause analysis (RCA) in cloud applications. In the investigation, they use the terms tracing and traces in the sense in which they are listed in, for example, the Oxford Dictionary. A single trace shows an individual request passage through the microservices. Systems trace traffic passing through a malfunctioning microservice to identify and remediate performance degradation issues, finding or describing their origin, development and duration in terms of trace types, spans, and tags.

A group of Armenian, German, and American researchers and R&D specialists deal with ML approaches that help to automate the management of complex systems such as virtual machines, hosts, and datastores by monitoring millions of time series metrics, terabytes of logs, and application traces to capture a high-resolution "image" of the entire stack. Recently, self-diagnostics for issues with such intelligent monitoring and analytics solutions has become another fundamental problem in customer environments. They require time-intensive expert analysis of these traces.

Researchers from Algerian universities use word embedding vectors and multi-grained scanning—a sophisticated extension of text tracing and tracking with sliding windows, Deep Random Forest, and contextual embedding model AraBert—to detect and remove hate speech in Arabic tweets.

Researchers from the Japanese university of Tsukuba consider a prototypical web-based communication system to enable teachers to track student engagement in remote environments. Students upload video activity reports via the system. By automatically fetching a video activity report and analyzing the content of the video, the system can automatically identify students whose engagement is declining.

German researchers from Wismar Applied University describe the requirements for traceable open-source data retrieval to bound mutation probabilities in two tumor-suppressor genes responsible for hereditary breast and/or ovarian cancer. They use a Dempster-Shafer model to consider the family history and first age of onset, as well as epistemic uncertainty.

Chilean scientists and Japanese R&D experts introduce retail sales performance indicators that should be regularly traced and analyzed to identify improvement opportunities and make informed decisions to optimize sales performance. This approach allows users to choose an optimized indicator prediction model for each retail store. Users can use a web tool to estimate improvements in the indicators based on desired sales goals.

Researchers from Chile and the UK deal with periodic pattern mining and look for efficient algorithms to find cyclical patterns in spatio-temporal databases.

Here is a more detailed overview of the contributions:

Data science with intelligent technologies

Two-Stage Optimal Hypotheses Testing for a Model of a Stegosystem with an Active Adversary by Mariam Haroutunian, Parandzem Hakobyan, and Arman Avetisyan.

This paper considers an information-theoretic model of a stegosystem with an active adversary. Unlike a passive adversary, an active adversary could modify the data, whether a covertext or a stegotext, sent by the legitimate transmitter over a public channel. The receiver's task is to decide whether the communication is a covertext or a stegotext and, in case of a stegotext, to further decide whether the message was sent by a legitimate sender or by an adversary. From the receiver's point of view, the authors suggest a two-stage statistical hypothesis testing approach that is logarithmically asymptotically optimal. This approach addresses the functional dependence of errors in both stages.

Distributed Tracing for the Troubleshooting of Native Cloud Applications via Rule-Induction Systems by Arnak Poghosyan, Ashot Harutyunyan, Naira Grigoryan, and Clement Pang.

The authors of this paper consider rule-induction classification methods for the root cause analysis of performance degradations of native cloud application problems. In their experiments, the researchers considered various approaches to determine their suitability for explaining performance degradation and unveiled the main benefits of each method.

Challenges and Experiences in Designing Interpretable KPI-diagnostics for Cloud Applications by Ashot N. Harutyunyan, Arnak V. Poghosyan, Lilit Harutyunyan, Nelli Aghajanyan, Tigran Bunarjyan, and A.J. Han Vinck.

The authors describe the existing technology challenges and their experiences while designing problem root cause analysis mechanisms that are automatic, application agnostic, and interpretable by human operators. They focus on diagnosis of cloud ecosystems through the Key Performance Indicators (KPI). Those indicators are utilized to build automatically labeled data sets and train explainable AI models for identifying conditions and processes "responsible" for misbehaviors. Experiments on a large time series data set from a cloud demonstrate that those approaches are effective in obtaining models that explain unacceptable KPI behaviors and localize sources of issues.

Deep Random Forest and AraBert for Hate Speech Detection from Arabic Tweets by Kheir Eddine Daouadi, Yaakoub Boualleg, and Oussama Guehairia.

This paper addresses the problem of hate speech detection in Arabic tweets. It proposes a method known as Contextual Deep Random Forest (CDRF) by combining Arabic contextual embeddings with Deep Random Forests. The proposed CDRF improves the accuracy rate of hate speech detection and outperforms existing classification approaches.

Human-centered computing with intelligent technologies

Using Video Activity Reports to Support Remote Project-based Learning by Kosuke Sasaki, Zhen He, and Tomoo Inoue.

This paper reports on the authors' experience with video activity reports in a project-based remote-learning activity in higher education. Characteristics of the video reports (length, pauses, negative statements) are contrasted with self-reported engagement scores using a minimalist instrument (UWES-3 questionnaire) developed in a

work-related context. The findings are conceived as steps towards designing an AI-enabled support system.

Towards a Traceable Data Model Accommodating Bounded Uncertainty for DST Based Computation of BRCA1/2 Mutation Probability with Age by Lorenz Gillner and Ekaterina Auer.

The authors present a novel uncertainty-aware model for determining the risk of cancer diseases. Their model is based on interval extensions of Dempster-Shafer evidence theory and applied successfully to data about breast cancer available in open-access publications and databases.

Retail Indicators Forecasting and Planning by Nelson Baloian, Jonathan Frez, José A. Pino, Cristóbal Fuenzalida, Sergio Peñafiel, Belisario Panay, Gustavo Zurita, and Horacio Sanson.

The article presents a methodology for forecasting plausible business goals for a particular retail store for the next two months. This methodology supports managers in planning sales goals and operation.

Efficiently Finding Cyclical Patterns on Twitter Considering the Inherent Spatio-Temporal Attributes of Data by Claudio Gutiérrez-Soto, Patricio Galdames, and Daniel Navea

This paper introduces HashCycle, a new and efficient algorithm for finding cyclical patterns over spatio-temporal databases.

The editors would like to thank the authors and express their gratitude to the AUA and Dr. Aram Hajian, dean of the College of Engineering, for hosting the CODASSCA workshop series, to the J.UCS consortium for accepting the special issues in the Journal of Universal Computer. Science, and to Johanna Zeisberg and the publishing team for their kind assistance throughout the process of writing and compiling the issue.

Thanks also to the German Research Foundation (DFG), the German Academic Exchange Service (DAAD), the UDE publication fund, and the sponsors involved for funding our common activities since 2015. Without the help of the CODASSCA 2022 editors and their ongoing encouragement, the present volume would not have been possible.

The guest editors are especially grateful to the anonymous reviewers who spent a significant amount of time reading the articles and making substantive suggestions for improvement., We are particularly indebted to Dr. Yanling Chen, Prof. Nelson Baloian, and Dr. Ashot Harutyunyan for their invaluable help during the preparation of this special issue.

Wolfram Luther
Gregor Schiele
Duisburg
September 2023