Editorial
Selected Research in Data Analysis and Theoretical Computer Science

Jurij Mihelič and Uroš Čibej, Guest Editors

Faculty of Computer and Information Science, University of Ljubljana, Slovenia

With modern computing technology, it is easy to collect a vast majority of data. However, a much bigger challenge is to analyze this data and extract as much knowledge out of it as possible. Therefore, much of the current research attention is focused on the so-called data science, a multi-disciplinary field consisting of areas such as statistics, data analysis, data mining, big data, machine learning, and others. Accordingly, this special issue includes four scientific papers broadly in the scope of data science. The articles focus on analysis techniques that give new insight into the data and give experts in various fields new valuable tools for their decision making.

Underneath every successful scientific and engineering field, there is a strong theoretical foundation. We present six papers that advance the theoretical aspects of computer science. In particular, the papers focus on areas such as graph theory, category theory, programming languages, and operational semantics.

We begin with the four papers focusing on the data algorithms. The first paper **A Case of Churn Prediction in Telecommunications Industry** by the authors Simon Brmež and Martin Žindaršič gathers the information from a telecommunication company and extracts the knowledge using techniques from machine learning and uses it to predict the churn, i.e., the leaving of customers. This is very important information for companies since they can detect unsatisfied customers quickly and by acting on it make the user experience much better.

The authors Kralj Novak, Senja Pollak, Aljoša Valentinčič, Igor Lončarski, Marko Pahor, and Martin Žnidaršič, in their paper **A Study of Importance of Textual Features for Predictive Models of Financial Indicators**, study various features of unstructured financial data, e.g., texts from financial reports, twitter messages, etc. and form a comprehensive study of the correlations and the predictive power of various features extracted from these texts. They demonstrate the relevance of these features and show a large potential for the future.

A surprising application comes from medical imaging, where the author Luka Šajn presents **Automatic Localization of White Spot Lesions Caused by Teeth Braces.** The article presents an improved method for locating a phenomenon arising from orthodontics. The author demonstrates that previously used methods can be significantly improved by using new approaches.

The last paper in the data science group is by Janja Klemenčič and Jurij Mihelič titled **Creation of Attribute Vectors from Spectra and Time-series Data for Prediction Model Development**. The paper explores the area of new attributes creation from the existing ones. In particular, it deals with extracting and summarizing the information from the data available in times series data such as spectral series and sensory process data used in many manufacturing industries.

The second group of papers focusing on theoretical aspects begins with a paper by Valerie Novitzká, William Steingartner and Ján Perháč titled **From Coalgebraic Logic to Modal Logic: An Introduction**. The paper introduces modalities of coalgebraic logic in order to describe the behavior of systems. The paper also presents many examples and uses them to compare coalgebraic and modal logic.

The sixth paper by Valerie Novitzká, Ján Perháč, and William Steingartner focuses on **Categorical Approach to Denotational and Operational Semantics**. Here, the authors apply category theory to constructs functors between operational and denotational semantics. The concepts are tested with a simple programming language.

Another paper from the field of programming languages by the authors William Steingartner, Ján Perháč, and Alexander Biliňski is **A Visualizing Tool for Graduate Course: Semantics of Programming Languages**. Its focus is on an educational software tool used for translation of a code written in a model language into inner code in JSON format.

One of the hottest problems in graph theory in recent years is graph isomorphism. In theory, this problem remains hard, but various practical approaches can make it quite tame. The paper by Luka Fürst titled **Neighborhood-based Pseudo-canonical Representation of Graphs** represents one such approach, where graphs are transformed into a pseudo-canonical representation and isomorphic graphs have the same representation. Since the method is not exact, there are certain cases where this approach does not work, but the author demonstrates that they are rare.

The ninth paper is a survey by Peter Gabrovšek and Jurij Mihelič titled **Graph Covering and Subgraph Problems**. The authors focus on graph problems; in particular, they present four covering problems such as vertex cover and dominating set as well as subgraph problems such as matching and independent set. They also collect relations between different kinds of problems.

The tenth and final paper by Andrej Brodnik and Sandi Režonja model and analyze an interesting problem which originates from chemistry and biology. In the article entitled **Construction of Stable Mesh Using Self-Assembly Chains,** they formalize the problem as the problem of embedding chains onto a mesh structure. They present an algorithm for the problem and perform an empirical evaluation.

All of the presented works give a careful study of the problems as well as focus on the algorithmic and computational approach to solving the problems. We believe that the whole issue is a very interesting read.