

## **Domain Knowledge in Knowledge Discovery and Privacy-Aware Intelligent Systems**

### **Preface**

---

The Knowledge Discovery in Databases (KDD) comprises a number of stages, such as understanding its goals, acquiring, cleaning and integrating relevant data sources, modeling the obtained data sets by means of feature extraction and selection in order to prepare them as inputs to intelligent data analysis algorithms, and finally interpreting the results of those algorithms in practice. The success of KDD depends, among many other factors, on the ability to interact with the domain experts and the availability of effective data and information sharing. The domain knowledge is crucial for the design and tuning of practically all KDD stages. Privacy-preserving is crucial to run intelligent systems without violation of individual privacy by exposing potentially sensitive person-specific information.

This special issue aims at elaborating on the state of the art in the areas of privacy-aware intelligent systems and the utilization of the domain knowledge in KDD. It is inspired by two workshops held at the 20th International Symposium on Methodologies for Intelligent Systems in Macau, China, 4-7 December 2012. The issue contains six carefully extended and thoroughly peer-reviewed articles, which were originally presented at the workshops. Their topics refer to data clustering and similarity, association rule induction and multi-label classification, as well as database engines, search systems and web services. Their results can contribute to a number of areas in our everyday lives, such as medical and healthcare systems, social and scientific networks, administrative services and so on.

The first paper, by Jan Rauch, is titled “Formal Framework for Data Mining with Association Rules and Domain Knowledge – Overview of an Approach”. The author outlines a methodology for data mining with association rules. The framework is based on a logical, appropriately enhanced calculus of rules. It allows to describe the whole knowledge discovery process, including formulation of analytical questions, application of the proposed analytical procedure and interpretation of its results. The domain knowledge is expressed using a formalized language aimed at deriving truly relevant rules.

The second paper, by Nicolas Anciaux, Danae Boutara, Benjamin Nguyen and Michalis Vazirgiannis, is titled “Limiting Data Exposure in Multi-Label Classification Processes”. The authors consider decision making processes based on multi-label classification models in the area of administrative services. They continue their previous research on minimizing data required from applicants as an input to online forms. They develop an experimental platform able to transform arbitrary multi-label data sets into collection rules and to measure the gain obtained in terms of minimizing data exposure.

The third paper, by Haruko Iwata, Shoji Hirano and Shusaku Tsumoto, is titled “Maintenance and Discovery of Domain Knowledge for Nursing Care using Data in Hospital Information System”. The

authors discuss how to support the customization of clinical pathways by using data about nursing actions stored in a hospital information system. They cluster data sequences in order to visualize the similarities between temporal patterns of clinical actions. The method is evaluated on real data sets of ten frequent diseases extracted from the information system in Shimane University Hospital.

The fourth paper, by Wuhui Chen, Incheon Paik and Patrick C.K. Hung, is titled “Privacy Issues in SOAP Message Exchange Pattern for Social Services”. The authors introduce a mathematical model for construction of the privacy policies in SOAP message exchange patterns and security tokens with SOAP messages. The proposed framework includes a formalization of interaction patterns, which describe how social services interact with each other according to data correlations. The interaction patterns can be illustrated by using the rules including sequential, parallel and conditional routing.

The fifth paper, by Michał Meina and Hung Son Nguyen, is titled “Search Result Clustering Based on Query Context”. The authors introduce an interactive clustering-based approach to information retrieval. It allows users to change the clustering structure by applying a free-text clustering context query that is treated as a criterion for document-to-cluster allocation. Exploration mechanisms are delivered by allowing users to interact with data at the level of topic discovery or cluster labeling. A similarity measure used while clustering is based on a special query-summarize graph structure.

The sixth paper, by Junpei Kawamoto, is titled “A Locality Sensitive Hashing Filter for Encrypted Vector Databases”. The author presents a new approach to reduce tuples that need to be compared by an encrypted vector database during similarity-based query processing. Vector databases can be especially useful, e.g., in text or image information retrieval. Encryption provides a security level which is important for real-life applications. The proposed method utilizes locality-sensitive hashing and whitening transformation. Query processing improvements are documented experimentally.

Let us take this opportunity to acknowledge all authors for important contributions and reviewers for insightful comments. We would also like to express our appreciation to Zbigniew W. Raś, the General Chair of ISMIS 2012, for encouraging us to work on this special issue, as well as Damian Niwiński, the Editor-in-Chief of *Fundamenta Informaticae*, for making this special issue possible to get published.

Guest Editors:

*Dominik Ślęzak*

Institute of Mathematics, University of Warsaw

Banacha 2, 02-097 Warsaw, Poland

and

Infobright Inc.

Krzywickiego 34 lok. 219, 02-078 Warsaw, Poland

slezak@mimuw.edu.pl; slezak@infobright.com

*Benjamin C.M. Fung*

School of Information Studies, McGill University

Montreal, QC, Canada, H3A 1X1

ben.fung@mcgill.ca

*William K. Cheung*

Department of Computer Sci., Hong Kong Baptist University

Kowloon Tong, Hong Kong

william@comp.hkbu.edu.hk