EDITORIAL

Special issue on "context-aware data mining (CADM)"

Chiara Renso · Vania Bogorny · Hui Xiong

Knowledge discovery is generally described as the process of automatic extraction of interesting and previously unknown patterns from large amounts of data. However, intelligent methods that explicitly incorporate semantic information into the discovery process are not yet well established in the literature. The exploitation of context during the mining process is one of the most common techniques to take advantage of the semantics of different application domains.

Contextual aspects may intervene in several steps of the discovery process, but the question is how these aspects interact within the process, and the forms that this can take. For instance, in data pre-processing, enriching the data with additional context information can lead to the generation of more meaningful and human understandable patterns, once the generated patterns are directly related to the input data. Also, during the mining task, context may be used as constraints, thus allowing search space reduction, pattern pruning, and the development of more efficient algorithms. Context information and domain knowledge may intervene in the post-processing steps, helping in the explanation of the results.

Recent years have witnessed increasing interests in exploiting context information in data mining. Also, data mining has been successfully applied in a number of different fields, such as bioinformatics, business intelligence, finance data analysis, location based services, network data analysis, and social link analysis. This special issue is an example of this, since the selected papers are related to different data mining areas such as privacy, sensor data, text mining, graph mining, and ontologies. As a consequence, context-aware data mining is a never-ending resilience field and attracts growing research efforts from different fields.

C. Renso
KDD Laboratory, ISTI-CNR, Pisa, Italy

V. Bogorny
Universidade Federal de Santa Catarina, Florianopolis, Santa Catarina, Brazil

H. Xiong (✉)
Rutgers, The State University of New Jersey, Piscataway, NJ, USA
e-mail: xionghui@gmail.com

The objective of this special issue on context-aware data mining is to bring together high quality papers facing different aspects of context-aware data mining. We received a total of 27 submissions from which the 7 papers have been selected for publication after a rigorous peer review process.

The first paper, entitled "COAT: COnstraint-based anonymization of transactions," faces the problem of producing anonymizations of transactional data proposing COAT (COnstraint-based Anonymization of Transactions) to transform data using specific context-aware constraints.

A framework for efficiently discover relevant association rules between user context data and services is introduced in "CAS-Mine: providing personalized services in context-aware applications by means of generalized rules".

Taxonomies (e.g., a geographic hierarchy on spatial coordinates, a classification of provided services) drive the rule generalization process, and after these rules are classified into groups according to their semantic meaning.

Time is an essential element of the context. User behaviors may vary from one period to another. The Paper "Discovering frequent behaviors: time is an essential element of the context" aims on finding coherent and compact behaviors frequent over a specific period.

The paper "Sensor data analysis for equipment monitoring" focused on a specific application for equipment monitoring. It proposes a model called (SDAEM) to mine the sensor time series data to understand oil plant operation status and predict failures in devices.

The paper "Statistical semantics for enhancing document clustering" is an extension of the paper that won the "Best Paper award" at the Second IEEE International Workshop on Semantic Aspects in Data Mining in 2009. This paper proposes new models for document representation that capture semantic similarity between documents based on measures of correlations between their terms. The paper uses the proposed models to enhance the effectiveness of different algorithms for document clustering.

Ontologies representing contextual background knowledge are exploited in "On ontology-driven document clustering using core semantic features" to improve clustering algorithms that make use of ontologies as background knowledge.

The method proposed in paper "An efficient graph-mining method for complicated and noisy data with real-world applications" introduces a graph-mining algorithm with approximate matching policy that takes advantage of previously known domain knowledge, thus providing a general framework to evaluate qualified frequent-induced graph patterns.

Finally, we would like to acknowledge the authors of the submitted papers and the reviewers who actively collaborated in reaching the high quality selection of papers in this issue. In addition, we would like to thank the Editor-in-Chief, Dr. Xindong Wu, for his help and strong support on this special issue!