Intelligent Data Analysis 21 (2017) S1–S3
DOI 10.3233/IDA-170868
IOS Press

Editorial

Dear Colleague:
Welcome to the 2017 special issue of Intelligent Data Analysis (IDA) Journal.

The contents of this issue present intelligent analysis techniques for big data and social informatics in various domains. The phrase "big data" indicates large, diverse, dynamic, and complex data sets. They may be generated from social media, internet of things, multi-media files, server logs, and other digital sources. Especially, social networks have become essential in daily life due to the popularity of mobile devices in recent years. Both the big data and social informatics need efficient and effective data analysis techniques to find relevant but not inundating knowledge. There are totally eleven articles included, mainly selected from The Fifth ASE International Conference on Big Data & The Fourth ASE International Conference on Social Informatics, which were jointly held by National University of Kaohsiung, Taiwan on October 7th to 9th, 2015. A few of them were selected from the 2015 Conference on Technologies and Applications of Artificial Intelligence, which was held in Tainan, Taiwan. The contents of the articles are briefly introduced below.

The first four papers are on various forms of analysis for different big data types. The first work authored by Pham *et al.* deals with single channel source separation, which is to recover speech signals from mixtures as accurate as possible. The authors propose an improved approach based on the non-negative matrix factorization (NMF) to separate mixed signals. They incorporate the group sparsity NMF with β-divergence, graph regularization, and an adaptive dictionary to achieve the purpose. Their experimental results show that the proposed method is outstandingly effective in speech separation in various scenarios.

Wang *et al.* in the second paper discuss the trade-off between privacy and utility for anonymized mobile context streams. Because only a few privacy-preserving techniques were proposed to anonymize sensitive mobile contexts before they are released to service providers, the authors propose a unified strategy to define both privacy gain and utility loss when anonymizing sensitive contexts on mobile user data. They provide an analysis framework, perform an extensive numerical evaluation on well-known anonymization techniques, and compare their trade-off between privacy and utility.

Jian and Chen next present a game-based learning platform on a cloud environment for improving the learning performance of students. The learning states of individual learners (students) can be analyzed from their game-playing data, and appropriate suggestion can then be fed back to learners. The dynamic assigning algorithm is adopted to classify individual learners into different levels according to their learning ability, with suitable teaching scenario provided to the learners based on the levels. Besides, learners can connect to the system through different remote devices.

In recent years, the short text media such as Twitter is very popular. The fourth work from Chen and Kao thus provide a solution to avoid some disadvantages of the LDA topic model on analyzing topics of short texts. The approach generates new virtual documents by re-organizing the words in documents and uses them to increase the amount of word co-occurrence relation. No additional external data is needed and the proposed method is easy to extend to some other existing LDA models.

The next five papers are about various kinds of pattern mining from big data. In the fifth work, Chen *et al.* explore the mining of spatio-temporal chaining patterns from non-identity event databases. Different

from the previous work relying on the existence of identity information for the accumulation of pattern appearances, the authors devise two novel algorithms to effectively process spatio-temporal data mining without identity information due to privacy concern. Two real criminal datasets without the identity information are also tested for the performance, with some interesting hidden phenomenon revealed.

Zihayat *et al.* then consider mining high utility sequential patterns in static and streaming data. They propose two efficient data structures, item utility lists and high utility sequential pattern tree, to maintain essential information in both offline and online fashions. Two mining algorithms based on the two data structures respectively are then presented. In addition, a novel utility model called sequence-suffix utility is designed for effectively pruning unpromising search space in the mining process. They also show their algorithms perform well in terms of execution time, memory usage and number of generated candidates.

Elkhani and Muniyandi in the seventh paper design a feature-selection model for microarray cancer data based on membrane computing. They use the kernel P system, which is a variant of membrane computing, combined with the multi-objective binary particle swarm optimization to get to the target. The proposed model is evaluated on the cell line data of colorectal cancer through Weka. The results indicate that the selected marker genes by the proposed model have high classification accuracy, precision and recall metrics in comparison with pure multi-objective binary particle swarm optimization feature selection.

In the era of big data, data cannot be totally fed into main memory for mining. Lin *et al.* thus propose a novel algorithm for fast mining association rules with secondary memory. It saves in-process mining information either in a hard disk or in main memory depending on the execution condition. Through the use of mixed hard disk and memory, data mining with limited memory can be handled smoothly. The empirical evaluation under various simulation conditions shows the proposed approach delivers good performance in terms of execution efficiency and scalability.

In the ninth paper of this issue, Lin *et al.* present a generic approach for mining indirect association rules in data streams. Indirect association rules are an interesting type of knowledge hidden in many applications. Instead of handling stream data mining by investigating different types of streaming models, the authors treat the problem from a generic viewpoint. They propose a generic window model that can represent all classical streaming models and retain user flexibility in defining new models. They also develop a corresponding algorithm, which guarantees no false positive rules and bounded support errors as long as the window model is specifiable by the proposed model.

The last two papers in this special issue are about recommender systems. Su *et al.* designed an effective social content-based collaborative filtering mechanism for music recommendation. Traditional recommender systems usually discover users' musical preferences by rating, which causes some disadvantages. The authors thus attempt to integrate social and collaborative information to predict users' preferences for avoiding the drawbacks. A multi-modal music recommender system is also presented. Through optimizing the integrated social-and-collaborative information, the users' preferences can be inferred more accurately and efficiently.

In the last paper of this special issue, Chen *et al.* proposed an approach to find group stock portfolios using the MapReduce architecture. The purpose of group stock portfolios is to partition stocks with similar properties in the same groups. The authors adopt the group genetic algorithm to find solutions, but the evaluation process is time-consuming. They thus further propose a map-reduce-based approach to speed up the evolution process while keeping good results. The chromosomes in a population are divided into subsets and sent to respective mappers, while the reducers execute fitness evaluation and genetic operations.

We hope the special issue can bring some interesting ideas and recent advances in the intelligent analysis of big data and social informatics. At last, we are grateful to all the authors for their contributions

and the referees for their vision and efforts. We would also like to express our thanks to Dr. A. Famili, the editor-in-chief of the journal, for his great support to realize the special issue.

With best regards
Dr. Tzung-Pei Hong, National University of Kaohsiung, Taiwan
Dr. Been-Chian Chien, National University of Tainan, Taiwan
Dr. Wen-Yang Lin, National University of Kaohsiung, Taiwans
Dr. Shyue-Liang Wang, National University of Kaohsiung, Taiwan
Guest Editors