

Editorial

Dear Colleague

Welcome to volume 18(6) of Intelligent Data Analysis Journal.

This issue of IDA journal, the last of 2014, consists of eleven articles, all covering various topics related to the theoretical and applied research in the field of Intelligent Data Analysis.

The first four articles are on frequent pattern discovery. Muino and Borgelt, in the first article of this issue, discuss synchrony in neuronal spike trains and present a framework for characterizing spike synchrony. In addition, they introduce a methodology and algorithms for identification of frequent parallel episodes in sequences of events. Results of their evaluation are presented in the paper. Line *et al.*, in the next article of this issue, discuss methods for efficient updating of sequential patterns and present a study that extends a sequential-pattern-tree algorithm that is targeted for efficient handling of sequential patterns. Their experimental results show that the proposed algorithm has good performance for incrementally handling newly inserted transactions in large databases. Ryang *et al.* in the third article of this issue discuss the complexity of rule mining in items database and argue that in real world, items in the database have different nature. This point emphasizes that association rule mining needs to consider multiple minimum supports as a result of which one can expect to discover all item rules by reflecting their characteristics. The authors propose an algorithm for mining high utility itemsets with multiple minimum supports and their experimental results show that their algorithm outperforms comparing algorithms on both real and synthetic datasets. Abdul-Kadir *et al.* look at the frequent item sets from the point of view of outlier detection in which they propose an approach based on frequent negative itemsets. The idea is based on generating knowledge and discovering interesting patterns from frequent positive and negative itemsets to identify the outliers, which are essentially transactions that are common or rare. Their proposed idea is evaluated using a set of UCI data sets with interesting results for outlier detection.

The next three articles are on the topic of data pre-processing and handling imbalanced data. Than and Ho discuss diversity and log-normality of data. Diversity is representative of variations in the attributes while log-normality is the situation where data follows a log-normal distribution. Methods such as Latent Dirichlet Allocation are investigated for this problem which results in their limitations for modeling these two issues. A vibrational method, called Dirichlet lognormal topic model, is proposed in which model learning and inference are reduced to solving complex optimizational problems. The experiments reported in this article suggest that the predictive power of this proposed approach is consistent with diversity and log-normality. Cao *et al* in the next article discuss the issue of class imbalance in various forms and propose a method that is a combination of hybrid probabilistic sampling and diverse random subspace. The idea behind the combination of these two is apparently to solve both class imbalance and within class imbalance problems in certain classes of data. Their experimental results demonstrate that their proposed approach performs very well in imbalanced data sets. The last article of this group by Prachuabsupakij and Soonthornphisaj is also about imbalanced data. The authors introduce a new resampling technique that is based on clustering with sampling and it is primarily for multiclass imbalanced classification using ensemble techniques. Essentially their proposed approach is based on three steps which consist of: (i) a clustering step to create training sets, (ii) applying two resampling techniques to

rebalance the class distribution and (iii) use of ensemble approaches to combine the models. The authors report on a carefully designed experiment to evaluate their proposed method.

The next two articles are on classification and learning. Adeli-Mosabab in the eighth article discuss the issue of multi-label classification in large scale data and propose a simple distributed algorithm for minimizing the nuclear norm of any data matrix in order to recover its low rank representation. Several synthetic and real data sets are used by the authors to evaluate the two elements of their approach which consists of nuclear norm minimization and distributed matrix completion. Taherian and Shiri in the next article of this issue discuss state abstraction and knowledge discovery in reinforcement learning and propose a framework to aggregate the results of state abstraction in multiple tasks of selected domains to be reused in future tasks of the same domain. In their approach, the authors show theoretically how abstraction speeds up learning and then they examine theoretically and algorithmically how to use the knowledge extracted by fuzzy value approximation of a single task in the future. The authors also show empirically that batch learning based on similarity measures can speed up learning in the future tasks of a given setting.

The last two articles of this issue are mostly on applied perspectives. Silva and Zarate in the tenth article of this issue provide a review of several approaches to deal with missing data. The main emphasis of their review is that one has to pay attention to the identification of the absence mechanism in order to choose the most appropriate method for missing values. And finally Thomaidou *et al.* in the last article of this issue, discuss ideas for an automated online advertising. They propose a methodology, an architecture and a framework for semi- and fully-automated creation, monitoring and optimization of pay-per-click advertising. Their prototype is experimentally evaluated on real world campaigns where the results are presented in the paper.

In conclusion, with this issue of the IDA journal, which is Volume 18(6), we are glad to report an excellent submission rate for the manuscripts submitted to our journal. Among the special issues that our colleagues have been working on and will be published soon, there is one that is related to the CIARP-2013 (Ibero American Congress on Pattern Recognition) and another one on BAFI-2014 (Business Analytics in Finance and Industry). We look forward to receiving your feedback along with more and more quality articles in both applied and theoretical research related to the field of IDA.

With our best wishes,
Dr. A. Famili
Editor-in-Chief