

Editorial

Dear Colleague

Welcome to volume 18(5) of Intelligent Data Analysis Journal.

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

The first five articles of this issue are about data understanding and search for patterns. Zimmerman in the first article of this issue discusses episode mining techniques and propose a process to compare these techniques using data simulation that mimics real world applications. This approach is based on recovering patterns that are embedded in noise. The experiments included in this article indicate that temporal constraints are more important in affecting the effectiveness of episode mining than occurrence semantics. Aghabozorgi and Wah in the second article of this issue discuss clustering of time series data and propose a new multi-step approach to accurately cluster time series data. From their experimental results that include data from different domains, the authors claim that their multi-step approach generates accurate clusters that are based on similarity in shape in very large time series data sets. Balasch-Masoliver *et al.* in the next article of this issue discuss the topic of multivariate aggregation in high dimensional data and argue that depending on the attributes chosen and their correlation, the amount of information loss and disclosure risk vary. The authors present a genetic algorithm based approach for multivariate microaggregation with its goal being finding optimal or near optimal attribute groupings taking into account both information loss and disclosure risk. The approach is extensively experimented with results included in this article. Visconti *et al.* in the fourth article of this issue discuss the topic of efficient gene expression profiling through integration of additional information such as functional annotations. The authors propose a new bi-clustering approach that includes gene associations. Their approach supports the extraction of biologically relevant bi-clusters through leveraging additional knowledge. This is demonstrated through experiments performed on several gene expression data reported in this article. The fifth article of this group, by Akdemir and Jannink is about the last step in search for patterns in data analysis. The authors propose some new methods for post-processing a large ensemble of conjunctive rules discovered through supervised, semi-supervised and unsupervised methods. They show through various examples that for high dimensional regression problems the models constructed through post-processing the rules with partial least square regression have significantly better prediction performance than the ones produced by the random forest.

The next three articles focus on a different aspect of learning and classification. Pevec and Kononenko compare and examine two families of non-parametric approaches that are common in construction of prediction intervals for arbitrary regression models in the supervised learning framework. The first of the families introduced in this article is based on the idea of explaining the total prediction error as a sum of model's error and the second is based on the assumption of similarity of data. Their results on a large set of real and artificial data demonstrate one of these proposed methods always generates valid and superior results comparing to others. Similarly, the seventh article of this issue by Kutsuna and Yamamoto is about a parameter-free approach to one-class classification where they argue that parameter tuning is a difficult process in all classes of learning, especially when one class is the subject of study. The authors propose

a one class classifier whose parameters can be tuned automatically. In this classifier, the region of a training data is expressed as a Boolean formula that is constructed by using a binary decision diagram. Their experimental results show that the proposed method works well with both synthetic and real data sets. Romero and de Campos in the last article of this group discuss multilabel classification and argue that due to its complexity, the most common approach is building binary classifiers and combining them afterwards. However, the basic assumption used in this approach which is class independence may not be valid for all problems. Therefore, the authors propose a methodology that is based on a combination procedure with a classifier trained on the co-occurrences of the labels. The results of experimenting the approach is shown in the article using three probabilistic base classifiers.

The last three articles of this issue are mostly on applied research and novel applications. Allahyar and Sadoghi-Yazdi propose a new linear discriminant analysis (LDA) approach. This approach incrementally computes the solution of LDA with the time complexity that is lower than most incremental algorithms. In their experiments included in the article, the efficiency of the proposed algorithm along with its speed and accuracy are compared to several existing incremental algorithms. In the next application article, Movahedian and Khayyambashi discuss the importance of modeling in recommender systems and propose a new system that is based on generating similarities between user and item profiles. These profiles are then extended by association rules discovered through an association rule mining process. Their proposed approach is evaluated using a real data set where their results show better representation of user interests. And finally, the last article of issue by Abreu *et al.* is about model based collaborative techniques. The authors develop a framework capable of automatically calculating the final game statistics through log-files where a feature selection algorithm is used to select the variables that most influence the final game results. The proposed approach that also includes steps such as ranking, clustering and selection of best strategies, is extensively experimented and the results are shown in the paper.

In conclusion, with this issue of the IDA journal, which is Volume 18(5), we are glad to report an excellent submission rate for the manuscripts submitted to our journal. Among the special issues that our colleagues are currently preparing, there is one that is related to the CIARP-2013 (Iberoamerican Congress on Pattern Recognition). We should have more news about the status of this special issue by Sept 2014. 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