

Editorial

Dear Colleague:

Welcome to volume 19(2) of Intelligent Data Analysis Journal.

This issue of the IDA journal, is the second issue of 2015 and consists of twelve articles, all covering a range of topics related to the theoretical and applied research in the field of Intelligent Data Analysis.

The first three articles of this issue are about learning and classification. Abdoos *et al.* discuss holonic multi-agent systems and what role these systems can play in modeling complex structures. They present a reinforcement learning system that is suitable for holonic systems. Their proposed method includes a self-similar structure in which the learning processes of the holons are independent of their actual positions in the holoarchy. Their experimental results show the performance of the proposed holonic reinforcement learning method. Salama and Freitas in the second article of this issue discuss the application of ant colony algorithms for constructing Bayesian multi-net classifiers. Ant colony optimization is a meta-heuristic global search method for solving combinatorial optimization problems inspired by the behavior of real ant colonies. The authors propose two algorithms suitable for building Bayesian multi-net classifiers. They present the experimental evaluation of their algorithms on 33 benchmark classification data sets where the performance of their algorithms is presented. In the next article, Dudas *et al.* use decision trees for post-analysis of multi-objective optimization problems. The authors focus on two multi-objective problems where their results show that useful relationships may be identified by employing decision tree analysis of the solution set from multi-objective optimization.

Tan *et al.* in the fourth article of this issue propose a new co-training algorithm that is based on modified Fisher's Linear Discriminant Analysis. The author's target area is semi-supervised learning where one has a small set of labeled samples to train classifiers. They apply two classifiers, one aiming to maximize the normalized between-class diversity and the other to minimize within-class diversity. They also proposed a method with a confidence criterion for selecting unlabeled data to expand the training data set. Their experimental results show that the new co-training algorithm outperforms some other methods and its average classification accuracy is improved through each iteration. In the fifth article of this issue, Brito *et al.* address the problem of clustering interval data through adapting a model-based approach. They parametric models for interval valued variables are based on variance-covariance matrix and directly takes the nature of the data into account. Their results that are based on synthetic and empirical data show the well-founding of their proposed approach. Their results also emphasize that there is a clear need to explicitly considering the intrinsic variability in interval data.

The next two articles are on the subject of clustering. Wang proposes a matrix factorization technique in order to compute a solution which can maximize between-base disjointness and volatility difference. The idea is based on a series of quantitative measures that are for evaluation of bases and their volatility. The studies use several data sets where clustering consistencies are compared as performance measure. In the study on some financial data, the clustering results of bases and persistent pairs has helped in understanding the contents of the data. Fu and Yang in the next article of this group propose a batch-mode active learning technique which is based on adaptation of SVM where unlabeled data are used for clustering. The approach then uses from each cluster points that are most similar to the other clusters

for labeling. Points near the boundary of clusters are considered as support vectors. The experiments presented in this paper show that the proposed method is efficient and robust.

Bouasker *et al.* in the eighth article of this issue discuss discovering rare patterns and explain the two main problems of rare pattern discovery which are the number of patterns and the quality of mined patterns. The authors introduce a new concise representation of rare correlated patterns as well as new derivation process for the generic bases of rare correlation association rules. Their experiments show the compactness rates offered by the concise representations of discovered patterns. Flores *et al.* in the next article of this issue focus on anomaly based techniques that are used in monitoring bandwidth consumption of a sub-network of computers. Detected anomalies could indicate what is missing in the network. Their proposed system is tested with a number of univariate and bivariate observation sequences to analyze and detect anomaly behaviours. Their proposed approach outperformed comparable methods. The next article by Lan *et al.* is also on mining pattern discovery for frequent item sets. The authors present a novel approach for weighted itemset mining where the idea is based on enhancing the effectiveness of reducing uncompromising candidates. The idea also includes projection-based pruning in order to tighten the upper bounds of weighted supports for item sets in the mining process. The performance of the proposed approach in pruning effectiveness and execution efficiency under various parameter settings is presented in the article.

The last two articles of this issue are on analyzing log files data. Saneifar *et al.* discuss the importance of discovering useful patterns from log file data, in particular in industrial applications where log files are not structured. They argue that passage retrieval methods would be useful and propose a new approach for automatically splitting log files into relevant segments based on their logical units. They then characterize the complex logical units according to their syntactic characteristics. Experiments based on real data sets presented in this article demonstrate the efficiency of their approach. The final paper of this issue by Ghanbari and Beigy propose an incremental learning algorithm that is based on ensemble learning and use this algorithm for spam filtering. The main assumption of this algorithm is that it assumes that the environment is stationary and it trains new weak classifiers for newly arriving data. The results presented in the paper show that the proposed algorithm can learn incrementally and can efficiently learn new cases.

In conclusion, with this issue of the IDA journal, which is Volume 19(2), we are glad to report an excellent submission rate for the manuscripts submitted to our journal. In 2014, in addition to our six regular issues, we published one special issue that was related to the CIARP-2013 (Ibero American Congress on Pattern Recognition) that our colleagues produced. In 2015, we expect to publish other special issues, one on BAFI-2014 (Business Analytics in Finance and Industry) and another one related to CIARP-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