Editorial

Dear Colleague:

Welcome to volume 19(3) of Intelligent Data Analysis Journal.

This issue of the IDA journal is the third 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 two articles of this issue are about clustering and unsupervised learning. Zhu *et al.* in the first article of this issue discuss the problem of optimal number of clusters and propose a method based on consistency criterion to address this issue. Their proposed approach requires little prior information about the data and it can automatically find the optimal number of clusters. Their experiments reported in this article consists of analyzing 22 real-world data sets form different domains in which four well-known clustering algorithms in combination with six clustering indices are used as benchmark methods. Their results show the superiority of their approach in proper determination of the number of clusters. The second article by Alizadeh *et al.* is a cluster ensemble approach in which the authors propose a feed-back framework which they call wisdom of crowd cluster ensemble. Their proposed approach is based on the analysis of conditions necessary for a crowd to exhibit the collective wisdom. The three main conditions in this approach are decentralization criteria, independence criteria and diversity criteria. The authors also suggest appropriate procedures for the evaluation process, which includes a new measure to assess the diversity.

The next four articles are on various forms of classification, a supervised form of learning. Zhao *et al.* in the first article of this group, inspired by multiple criteria linear and quadratic programs propose a novel optimization approach suitable for binary classification and regression problem. Two main advantages of this approach are (i) discovery of a feasible solution and (ii) handling linear inseparable case and nonlinear relationship through introduction of kernel functions in the entire framework. Using various classical approaches and multiple data sets, their proposed approach is evaluated where the results on both synthetic and real data sets are presented along with the effectiveness of their approach. Luor in the second article of this group report on a comparative study of data standardization and its influences in the performance of support vector machines in classification problems. The topic is particularly discussed in the domain of kernel based methods and SVM learning algorithms. The proposed approach is evaluated using a number of synthetic and real-world data sets. The experiments reported in the paper show that a suitable standardization processing has significant improvement on the performance of classification techniques. Lee in the next article of this group presents a novel methodology for sequential classification, a two stage process, which involves sequential pattern generation and classification. The author also proposes a number of theorems to reduce the computational complexity that is common in generating sequential patterns. The article reports on some experiments performed with synthesized sequence databases. In the last article of this group Kamaei and Altancay discuss the issue of extrapolation and interpolation inaccuracies that are common in the performance of the nearest feature line classifiers and propose a novel scheme that is based on editing the feature line segments. Their proposed approach involves two major strategies which involve generalization-based and representation-based elimination. The proposed approach is evaluated through its applications in different orders. Experimental results

reported in the paper demonstrate the better performance of the approach in terms of imorived accuracy when compared with other nearest feature line methods.

Castro and Azevedo in the seventh article of this issue discuss the deficiencies of symbolic aggregate approximation, a common technique in analyzing time-series data. This approach requires specifying two key parameters which are symbolic length and the alphabet size. However proper selection of these parameters is a key issue in the performance of the approach. The authors propose an approach to automatically estimate the best values for these parameters and suggest that their proposed technique can be embedded in existing data mining tasks. The concept is evaluated from several points of view, among which is its impact on visualization interpretability. Fanaee and Gama in the next article of this issue discuss the topic of complex data streams in syndromic surveillance applications and propose a new algorithm to efficiently analyze data in this domain. Their proposed algorithm exhaustively searches for contrast sets in the multivariate data and generates an alert when it finds statistically significant rules. To solve the issue of false alerts, the authors also propose a new approach called EigenEvent that enables detection of both overall and dimension-level changes. Their experimental results on some benchmark data sets demonstrate the better performance of this approach.

The last four articles of this issue are mostly on data preprocessing and results post-processing. Valizadeh and Brazdil discuss the topic of density-based graph model summarization and propose a new approach for multi-document summarization, which is basically through adding the concept of density to existing methods. They evaluate their approach on a number of text summarization applications where they demonstrate that by adding the density concept one can achieve better results and faster than existing methods. Ros *et al.* in the next article propose a new instance selection algorithm that is primarily suitable for classification in non-trivial data sizes. The proposed algorithm is hybrid and runs with only a few parameters that directly control a typical classification application that has the three objectives of low errors, minimum storage requirements and the best runtime. The whole idea behind this algorithm is that instead of using instant selection from the whole database, it is applied to strata that is derived from the regions that represent a set of patterns selected from original training set. Experiments performed with various synthetic and real data sets demonstrate the advantages of the proposed approach. Peng *et al* in the eleventh article of this issue, which is also on data preprocessing, argue that feature selection has a strong influence on discovery of accurate prediction models. The authors propose a unified feature selection method that is primarily to improve robustness and accuracy of prognostics under both supervised and unsupervised applications. They report the performance of this algorithm on two engineering applications. And finally in the last article of this issue, Byun *et al.* propose a multi-currencies trading algorithm that applies a stock-trading method to foreign exchange market. Their proposed algorithm is a combination of principal component analysis and neural nets and its objective is to build a classifier from the foreign exchange market data. The authors demonstrate the profits that could be gained from this approach and discuss a basic procedure that the currency-trading algorithm must follow.

In conclusion, with this issue of the IDA journal, which is Volume 19(3), we are glad to report an excellent rate of manuscript submission for our journal. In addition to our six regular issues, we expect to publish one special issue each year which is normally related to a scientific conference similar to CIARP-2013 (Ibero American Congress on Pattern Recognition) for which our special issue was published at the end of 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