Editorial

Dear Colleague:

Welcome to volume 20(4) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal, the fourth issue of 2016, consists of twelve articles, all covering a wide range of topics related to the theoretical and applied research in the field of Intelligent Data Analysis.

The first five articles of this issue are about various IDA methods. Feng and Jing in the first article discuss uncertainty measurement in rough sets and argue that although some uncertainty measures have been investigated, the definitions of existing uncertainty measures in ordered information systems are mainly based on information theory approach. The authors investigate some properties and relationships among six different uncertainty measures. Their theoretical analysis indicates that these measures can be used to quantify the uncertainty in ordered information systems. This is demonstrated through some illustrative examples in the article. Drole and Kononenko in the second article explore the idea of closed world specialisation (CWS) and propose two different approaches to incorporating it into the induction process of a bottom-up inductive logic programming system. Their motivation comes from the fact that using CWS as a post-processing step would be incapable of solving problems in which the negated part of the hypothesis is crucial. The authors apply the proposed approaches in an Inductive Logic Programming system and give examples of problems where classical CWS fails to find a complete and consistent solution, whereas their proposed approaches succeed. Mirzamomen and Kangavari in the next article of this group present a new decision tree learning algorithm, called fuzzy min-max decision tree which is based on fuzzy min-max neural networks. In contrast with traditional decision trees in which a single attribute is selected as the splitting test, the internal nodes of their proposed algorithm contain a fuzzy min-max neural network. In this article it is shown experimentally that the decision trees resulted from the proposed algorithm achieve the highest accuracy and the lowest size and depth in comparison with C4.5, BFTree, SimpleCart and NBTree on the most commonly used UCI data sets. In the next article of this group, Mirisaee et al discuss Binary Matrix Factorization (BMF) and show that an alternating strategy with the L2-norm, used at the core of several methods to solve BMF, can be reformulated as an Unconstrained Binary Quadratic Programming (UBQP) problem. This reformulation allows one to use local search procedures designed for UBQP in order to improve the solutions of BMF. The authors introduce a new local search dedicated to the BMF problem and show that this solution is in average faster than the previously proposed ones. Yu *et al.*, in the last article of this group, discuss that link prediction is a fundamental problem in network data analysis, and explain that network topology information have been applied in this class of work, because of its simple framework. They argue that prediction accuracy of these methods depends on the compatibility between the algorithm definition and the structural characteristics of the target network, so the stability of structure similarity-based algorithms would be low. The authors propose a Choquet fuzzy integral-based link prediction method, evaluate it in a number of empirical experiments using six real networks and demonstrate that the proposed method outperformed the mainstream link prediction baseline methods.

The next group of articles are about various aspects of dynamic learning and item set mining. Yo and Holder explain that graph classification has received considerable interest during the past years where

most approaches focus on designing effective kernels to compute similarities for static graphs. However, these kernels become computationally intractable in terms of time and space when a graph is presented in an incremental fashion with continuous updates, i.e., insertions of nodes and edges. In this article, the authors examine the problem of classification in large-scale and incrementally changing graphs and propose a framework combining an incremental support vector machine (SVM) with the Weisfeiler-Lehman (W-L) graph kernel. They validate the advantages of their learning techniques by conducting an empirical evaluation on several large-scale real-world graph datasets where their experimental results show the benefits of their subgraph extraction method when combined with the incremental learning techniques. Lee in the seventh article of this issue discusses and compares grid proximity measures that use representative data points in grid cells and the average distance between the data points in grid cells. The author applies grid distance measure to the grid-based clustering problem where the number of clusters is dynamically determined by using a threshold value and by maximizing intra-similarity in a cluster and inter-dissimilarity between the clusters. The article reports numerical experiments on randomly generated problems and also for a clustering problem concerning microarray data of human fibroblasts in varying serum concentrations, with the latter data having been taken from a prior study. Zamani and Beigy in the eighth article of this issue argue that the best approach for learning from online data is to exploit an online learning algorithm and apply online ensemble methods. They explain that prediction with expert advice is a well-studied problem in the online ensemble learning literature, where the weighted majority and the randomized weighted majority (RWM) algorithms are two well-known solutions to this problem. The authors aim to resolve this problem by proposing a novel online ensemble algorithm to the problem of prediction with expert advice and propose a cascading version of RWM to achieve not only better experimental results but also a better error bound for sufficiently large datasets. Lin *et al.* in the next article of this issue explain that many approaches have been proposed for High Utility Itemset (HUI) mining from a static database but fewer studies have been developed for dynamic environment. The authors introduce an efficient algorithm to discover HUIs and through experiments show that the proposed maintenance approach for transaction deletion significantly outperforms the previous approaches in terms of execution time, memory consumption and scalability. Song *et al.* in the last article of this group also discuss high utility itemset mining and introduce an expansion method that is based on binary partition. The authors define the transaction utility list and key support count along with a new pruning strategy. Their tests on publicly available datasets show that the proposed algorithm outperforms other state-of-the-art algorithms.

The third group of articles in this issue are on applied IDA research. Oviedo *et al.* in the eleventh article consider the use of graphical probabilistic models in the field of education. They argue that general learning procedures involve a large number of variables that measure different aspects of the same concept, as it can be the case of socio-economic indicators in a population. They also argue that all the variables have some degree of dependence among them, without a true causal structure. They present a new procedure which makes a hierarchical clustering of the data while learning a joint probability distribution. The method is applied to the analysis of two datasets of the educational data where the results are quite interesting. Montazeri *et al.* in the last article of this issue discuss the importance of feature selection in classifying high-dimensional data and propose a novel method based on hyper-heuristic approach to find an efficient proper feature subset. In their proposed method, low level heuristics are categorized into two groups: the first group contains exploiters which cause to exploit the search space efficiently, and the second includes explorer heuristics which explore the solution space by dwelling on random perturbations. Their empirical study of the proposed method on several commonly used data sets from UCI repository indicates that the proposed approach outperforms recent methods in the literature for feature selection.

In conclusion, with this issue of the IDA journal, we would like to let you know that the IOS Press, the publisher of the IDA journal, has several 20th anniversary events planned for 2016. These plans would be held during two related conferences: AI-Stat 2016 in Cadiz-Spain and ECML/PKDD 2016 in Verona-Italy.

During 2016, in addition to our six regular issues, we will also publish a special issue related to CIARP series of conferences. 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