

Editorial

Dear Colleague:

Welcome to volume 21(6) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal contains 11 articles representing 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 aspects of unsupervised learning methods in the IDA. Wang *et al.* in the first article argue that existing privacy preserving mechanisms for trajectory clustering contend with the problems of narrow applicability, low-level utility, and difficulty in being applied to real scenarios. To deal with this problem, the authors propose a differential privacy preserving mechanism, called Cluster-Indistinguishability, to support trajectory clustering. Their experimental results show that their proposed approach has general applicability and better performance compared to existing methods. Xiao *et al.* in the next article of this issue discuss the two main problems with the K-means algorithm: how to provide appropriate number of clusters and how to determine initial cluster centers automatically. To solve the problem of appropriate number of clusters automatically, the authors propose the Davies Bouldin Index (DBI) based hierarchical K-means (DHMK) algorithm on the basis of their previous work. The proposed algorithm can integrate DBI metric into their hierarchical K-means algorithm and can determine the number of clusters with low time cost. Their experiments results on a number of UCI datasets and synthetic data demonstrate the effectiveness and feasibility of their proposed algorithm. Li and Nagan in the next article of this issue propose a weight-adjusted-voting framework that combines an ensemble of classifiers for improving sensitivity of prediction. In this framework, the authors first adjust each individual classifier's weight in the ensemble based on their ability of making correct predictions, and then use the weight of classifiers and a voting strategy to make final predictions. To compare the sensitivity of the proposed weight-adjusted-voting, and two other approaches of combining classifiers – voting, and stacking, the authors use two different datasets in the UCI machine learning repository for evaluation. Their results have demonstrated their proposed approach performs better in sensitivity than other approaches compared in the experiment. Roostaiyan *et al.*, in the fourth article of this issue, discuss that in many real-world applications, data contain heterogeneous input modalities and learning a distance metric or similarity measure that originates from all input modalities or views is essential for many tasks such as content-based retrieval ones. Using properties of multi-modal data, the authors design multi-modal deep networks and propose a pre-training algorithm for these networks. Their proposed network has the ability of learning intra- and inter-modal high-order statistics from raw features. Their experimental results show that the proposed method outperforms recent methods on image retrieval tasks. He *et al.* in the last article of this group propose a probabilistic generative model, called label correlation mixture model, intended to depict multi-labeled document data, which can be utilized for multi-label text classification. The approach is based on two stochastic generative processes, which correspond to two submodels: 1) a label correlation model; and 2) a label mixture model. The authors also introduce an inference algorithm for calculating the generative probability of a multi-label class and present the results of extensive multi-label classification experiments on three standard text

data sets. The experimental results show significant performance improvements comparing to existing approaches.

The next five articles are about novel applications of IDA methods. Zhang *et al.* in the first article of this group discuss the problem of emotion classification of Weibo's posts in a hierarchical way using a constrained topic model and Support Vector Regression (SVR). The proposed approach is based on employing a hierarchical emotion structure to classify emotions more precisely into multiple classes. This hierarchy can meet different research granularities. The whole architecture is proposed aimed at alleviating the pain of misclassification caused by feature imbalance and decreasing the labor cost. The experimental results validate that their proposed model outperforms traditional methods with precision, recall and F-scores. Pan *et al.* in the next article of this group discuss the topic of frequent itemset mining as one of the most common data mining techniques, which is based on the analysis of the occurrence frequencies of items in transactions. They argue that the approach is inapplicable in real-life situations since customers may purchase several units of the same item and all items may not have the same unit profits. The authors propose to mine the skyline frequent-utility patterns (SFUPs), by considering both the utility and the occurrence frequencies of items. A SFUP is introduced as a non-dominated itemset, where the dominance relationship between itemsets is based on the utility and frequency measures. Substantial experiments presented in this article show that SFU-Miner outperforms the state-of-the-art SKYMINE algorithm for SFUP mining in terms of runtime, memory consumption, number of candidates, and scalability. Sendi *et al.* in the eighth's article of this issue argue that generated content on the microblogging social network (e.g. Twitter) continues to grow with significant amount of information and the semantic analysis offers the opportunity to discover and model latent interests' in the users' publications. The authors focus on the problem of uncertainty in the users' publications and propose a new approach for users' interest discovery from uncertain information that augments traditional methods using possibilistic logic. Their empirical analysis and the comparison with the most known methods proves the significance of this approach. Feizi and Nazemi in the next article of this issue propose a recurrent neural network to support vector machine (SVM) learning in regression with widespread applications in a variety of settings. The SVM learning problem in regression is first converted into an equivalent quadratic programming (QP) formulation. An artificial neural network for SVM learning is then proposed. The presented neural network framework guarantees to obtain the optimal solution of the support vector regression (SVR). The authors provide two illustrative examples to demonstrate of the effectiveness of their proposed method. Peng *et al.* in the last article of this group propose a multi-stage maximum likelihood approach to recover the latent parameters of the stochastic block model, in time linear with respect to the number of edges. The authors also propose a parallel algorithm based on message passing. Their experiments demonstrate that the algorithm can produce high quality results on both benchmark and real-world graphs. An example of finding more meaningful communities is illustrated consequently in comparison with a popular modularity maximization algorithm.

And finally the last article of this issue is a survey article in which Cano and Morisio provide a systematic evaluation of the state of the art hybrid recommender systems introduced over the last decade. They address the most relevant problems considered and present the associated data mining and recommendation techniques used to overcome them. The authors also explore the hybridization classes each hybrid recommender belongs to, the application domains, the evaluation process and propose future research directions. Based on their findings, most of the studies combine collaborative filtering with another technique often in a weighted way.

In conclusion, in the last issue of our volume 21, we would like to thank all the authors who have submitted the results of their excellent research to be published in the IDA journal. 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