Editorial

Dear Colleague:
Welcome to volume 21(2) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal, the second issue of 2017, represents our twenty first year of publication with the objective of serving the IDA community. This issue contains 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 two articles of this issue are about various aspects of unsupervised learning. In the first article of this issue, Huang argue that mining richly structured heterogeneous datasets are a major challenge for data mining and further discuss that when traditional clustering methods are applied, heterogeneous networks consisting of multiple entities must first be converted to homogeneous networks. The author proposes a three-phase general framework to directly handle the information contained in extended star-structured heterogeneous data. A numerical example is provided in the article to illustrate the modelling idea and working principle of the proposed method, and experiments on a real-world dataset show the effectiveness of the three proposed algorithms. Wan *et al.*, in the next article of this issue, discuss that in the big data era, the efficient indexing of gradually increasing databases has become vitally important for information retrieval and to incrementally adapt to changes of databases, the authors propose a novel clustering based dynamic indexing and retrieval approach. This similarity retrieval method which utilizes a dynamic clustering tree combines one-way search and backtracking strategy to gain good retrieval accuracy and efficiency. The authors applied their approach to example-based image retrieval where the experimental results confirm that their proposed method is effective and promising.

The next three articles are on recommender systems. Mei *et al.*, in the first article of this group, explain that trustworthy computing is important in multi-agent systems, social network analysis, and recommender systems. As an additional dimension of information to past rating history, the authors emphasize that trust would be helpful for improving the accuracy of recommendations. The authors study people's trust and rating behavior with the Epinions dataset, which is a popular product review website allowing users to rate various categories of products. The authors propose a trustee-influence based trust model where a trustee's activeness or trustworthiness is used to determine trust relationships. Their experimental results demonstrate the effectiveness of the proposed trust model for recommendation. In the next article of this group Khalid *et al.* argue that clustering algorithms that are commonly used to increase the scalability of recommender systems, cannot accommodate the incremental updates with the arrival of new data. The authors propose a new clustering algorithm called One-Pass which maintains a good level of accuracy, scale well with data, and build the training model incrementally with the arrival of new data. The algorithm is experimented with four data sets where the results are comparable with K-Means in terms of accuracy and cluster quality. Wu *et al.*, in the fifth article of this issue, propose a trapezoid fuzzy rating model in order to reflect the actual case of users' decisions and rating patterns, and solve the sparsity problem of traditional collaborative filtering recommendation algorithms. Based on this model, the user fuzzy similarity-based collaborative filtering recommendation algorithm is designed. Their experimental results show that the proposed method can obtain better performance than other traditional methods, especially when implemented in a sparse dataset with high ratio between users and items.

The third group of articles in this issue are on data streams and learning. Rodriguez *et al.* in the first article of this group argue that large scale multi-label learning, i.e. the problem of determining the associated set of labels for an instance, is gaining relevance in recent years due to the emergence of several real-world applications. They also argue that a well-known approach for multi-label classification is the Binary Relevance (BR) algorithm for which the serial implementation is not suitable for medium or large datasets due to the time and computational resources required. The authors discuss a parallel implementation of the MULAN BR technique that harnesses the computational power of multi-core processors. Their experiments show significant improvements in speed-up of the classification tasks. Pisani *et al.* in the seventh article of this issue argue that authentication mechanisms already present in devices, such as smart phones, may not provide enough security. Therefore, new authentication methods, such as accelerometer biometrics, have been proposed. The authors investigate the use of adaptation mechanisms to update user models in accelerometer biometrics in a data stream context. Practical issues regarding the usage of accelerometer data are also discussed in this article. Zhuang *et al.* in the last article of this group consider feature transformation as a key step in data mining and propose a new feature transformation method named dynamical feature bundling for decision tree algorithms. The proposed approach that consists of three steps has been evaluated on several data sets where improvements of 2–9% have been obtained on very imbalanced data sets.

And finally, the last group of articles in this issue are about novel application of IDA methods. Pourkazemi and Keyvanpour in the first article of this group argue that community detection is a NP-hard problem and evolutionary algorithms have been considered as one of the most effective approaches. The authors propose a multi-objective particle swarm optimization algorithm that is based on an opposition-based method for producing an initial swarm. The performances of the proposed algorithm and its extended versions on real networks are presented in this article. Their results show that the performance of this algorithm is not effected as the size of social network grows. Weng *et al.* in the next article of this issue discuss that identifying itemsets with higher values (utility and cost-benefit) from frequent itemsets with low utility is appealing. However, past studies on utility itemset mining have focused on high-utility itemsets rather than low-utility itemsets. The authors propose a new algorithm to discover frequent itemsets with low utility. Their experimental results obtained using two real-life datasets show that the top K itemsets determined by the proposed algorithm have higher prediction performance in the measures of utility and cost-benefit compared with other frequent itemsets determined by a traditional algorithm. Nemati *et al.* in the eleventh article of this issue argue that affective video retrieval systems seek to retrieve video contents concerning their impact on viewers' emotions. These systems typically apply a multimodal approach that fuses information from different modalities to specify the affect category. They emphasize that the main drawback of existing information fusion methods exploited in affective video retrieval systems is that they consider all modalities equally important. In order to address this drawback, the authors propose a new information fusion method that is based on the Dempster-Shafer theory of evidence. Their experiments on the video clips of DEAP dataset indicate that the proposed method outperforms existing evidential information fusion methods significantly. In the last article of this group Xu *et al.* argue that topic models have been widely applied in discovering topics that underlie a collection of documents and incorporating human knowledge can guide conventional topic models to produce topics which are easily interpreted and semantically coherent. The authors further discuss that several knowledge-based topic models have been proposed in the past, but these models just leverage lexical knowledge of words that are often not in accordance with topics. To solve the problem, the authors investigate utilizing entity knowledge, concepts and categories in Wikipedia, as prior knowledge into topic models to discover more coherent topics. They propose a novel knowledge-based topic model,

which not only models the relationship between words and topics, but also utilizes concept and category knowledge of entities to model the semantic relation of entities and topics. They present the results of their approach with the state-of-the-art knowledge-based topic models, on three datasets where the experimental results show that their approach outperforms the existing base-line methods on all three datasets.

In conclusion, we would like to let you know that this year, as what we have done since 2014, in addition to our six regular issues, we will also publish a special issue related to a highly relevant conference (The Fifth ASE International Conference on Big Data) that was held in Taiwan. This issue will be published between our second and third issue of this year. 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