

Editorial

Dear Colleague:

Welcome to volume 24(2) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the second issue for our 24th year of publication. It contains twelve articles representing a wide range of topics related to the theoretical and applied research in the field of Intelligent Data Analysis.

The first two articles are about sentiment analysis and learning in IDA. Chen *et al.* in the first article argue that it is difficult to use the popular supervised learning methods to complete a sentiment classification task because marking data manually is time-consuming and laborious and unsupervised sentiment classification methods are mostly based on sentiment lexicons. The authors propose a novel framework that is based on multi-source information fusion for learning. They extract four kinds of emotional information, which are lexicon emotional, word co-occurrence, word polarity and polarity relationship information of emotional word pair. Their experimental results on five Amazon product review datasets show that the sentiment dictionary constructed by their proposed method can significantly improve the performance of review sentiment classification compared with the state-of-the-art methods. The second article by Musa *et al.* is about cross-lingual sentiment topic model evolution where they argue that most of the existing models have been designed to work well in a language with more abundant sources. The authors propose a cross-lingual sentiment topic model evolution over time by jointly modeling time with topic and sentiment. The topic-specific sentiment is extracted from the entire data at once and not for single document. The authors conduct an inference algorithm of their approach based on Gibbs Sampling where their experimental results on Chinese and English news-reader dataset show that their approach achieves significant improvements over the state-of-the-art methods.

The second group of articles in this issue are about optimization and advanced data pre-processing in IDA. Kempfert *et al.* in the first article of this group present a study on nonlinear dimension reduction methods. The authors provide a brief review of popular techniques for this class of problems and conduct a comparative performance study based on three simulated datasets, after which the performance of the techniques are evaluated through application to a pattern recognition problem in face image analysis. Adak and Demiriz in the next article of this group propose a hybrid application of Population Based Ant Colony optimization that uses a data mining procedure to wisely initialize the pheromone entries. The authors argue that in this area of research, frequent patterns in a number of initial high-quality solutions are extracted to guide the subsequent iterations of an algorithm, which results in an improvement in solution quality and computational time. They propose to carry out independent runs and collect elite sets over these trials. Their computational experiments, conducted both on symmetric travelling salesman problem and symmetric/asymmetric quadratic assignment problem instances, showed that their proposal produces significantly better results, and is more robust than pure applications of population-based ant colony optimization. Abedi and Gharehchopogh in the next article of this group argue that Firefly Algorithm is a successful metaheuristic algorithm for solving continuous optimization problems and although it performs very well in local search, it has weaknesses and disadvantages in finding solution in global search. The authors propose three different approaches that are based on the Dragonfly Algorithm (DA) processes and the OBL method to improve exploration, performance, efficiency and

information-sharing. Their evaluation shows better converge with the higher dimensional optimization functions toward the target in comparison with other metaheuristic algorithms. Rosales-Salas *et al.* argue that by providing a complete record of time use for a given population, time use studies enable investigators to test various hypotheses concerning that behavior where measuring or even fully identifying all time uses would be impossible without the proper data mining tools. The authors propose a framework for mining sequences of activities to capture more complex patterns than those currently available on how individuals organize their days. The proposed framework has been applied to the American Time Use Surveys dataset to explore individual time allocation behavior, identifying sequences of activities that are frequent.

And finally the third group of articles are about advanced learning methods in IDA. The first article of this group by Li *et al.* is about an intrusion detection method based on active transfer learning where the techniques plays a very important role in the field of network security. The authors propose an intrusion detection algorithm based on active transfer learning that takes advantage of transfer learning and does not need to satisfy the basic assumptions of the traditional machine learning. Their experimental results show that the intrusion detection rate of the new algorithm is greater than benchmark algorithms, and the training time efficiency improves at the same time. Duan *et al.* in the next article argue that although Bayesian network classifiers (BNCs) are powerful tools to mine statistical knowledge from data and infer under conditions of uncertainty, most traditional BNCs focus on mining the dependency relationships existed in labelled data while neglecting the information hidden in unlabelled data, which may result in the biased decision boundaries. The authors introduce a new order-based greedy search heuristic based on mutual information for building efficient structures in tree-augmented naive Bayes, which is a highly accurate learner while maintaining simplicity and efficiency. Their extensive experimental results on UCI machine learning repository demonstrate that their proposed algorithm is a competitive alternative to state-of-the-art classifiers. Wen *et al.* in the ninth article of this issue argue that although domain adaptation is an important branch of transfer learning, previous studies have always attempted to minimize the optimization goal while neglecting the relative quality of features or instances. To reduce interference between instances in the process of domain adaptation, the authors introduce a novel method that uses the overlapping degree to measure every feature or instance's relative quality and implement feature or instance reweighting. Their experiments verify that their proposed method outperforms others where finding optimal parameters can yield more accurate results than the original method. Gui *et al.* in the next article of this group explain distant supervision which has become the leading method for training large-scale information extractors. However, most previous works use only simple labelling functions, resulting in too much noise in the training data, and the knowledge bases are far from well-explored. The authors propose to make use of existing knowledge bases to effectively learn labelling functions that are represented as Markov Logic. Their experimental results show that the training data produced by the learned labelling functions is significantly improved in quality. Deng *et al.* in the next article of this issue explain that knowledge bases provide a large amount of structured information for entities and relations while distantly supervised relation extraction only utilizes knowledge bases to automatically generate datasets by ignoring the background information in KBs during the relation extraction process. The authors propose knowledge-embodied attention that leverages knowledge information in knowledge bases to reduce the impact of noisy data for distantly supervised relation extraction. Their experimental results demonstrate that their approach outperforms all baselines. And finally, El Mousawi *et al.* in the last article of this issue describe a new semi-supervised clustering algorithm as part of a more general framework of interactive exploratory clustering, that favors the exploration of possible clustering solutions so that an expert can tailor the best clustering according to his/her domain

knowledge and preferences. Their experiments show that the best results may be achieved only with the addition of preferences to traditional metric learning algorithms and that their approach performs better than state-of-the-art algorithms.

In conclusion, we would like to thank all the authors who have submitted their manuscripts reporting the results of their excellent research that is evaluated by our hard working referees and 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