Intelligent Data Analysis 23 (2019) 255–257
DOI 10.3233/IDA-190001
IOS Press

Editorial

Dear Colleague:

Welcome to volume 23(2) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal, is the second issue for our 23rd year of publication. It contains 12 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 text and sentiment analysis in IDA. Liu *et al.* in the first article of this issue argue that since volume of short text data increases rapidly it is essential to organize and summarize these data automatically where topic model is one of the effective approaches. The authors introduce a model which makes use of Recurrent Neural Networks to learn relationships. By using the learned relationships, the authors introduce Bigrams, which can display topics of interest. Through experiments on two open-source and real-world datasets, the authors demonstrate better coherence in topic discovery. Along the same line of research, Gravoc *et al.* in the second article of this issue argue that Sentiment Polarity Detection (SPD) (classifying texts by "positive" or "negative" orientation) has become more important and challenging task in recent years. The authors explore different n-gram based text representation models in order to determine the most valuable model for the representation of text documents in various languages, which can be used successfully by ML classification techniques for solving SPD tasks. Their proposed n-gram models were used in conjunction with k-Nearest Neighbourhood (kNN), Support Vector Machine (SVM) and Maximum Entropy (MaxEnt) algorithms to determine opinion polarity of the proposed movie reviews. Significant improvements over the baselines have been demonstrated in this article.

The second group of articles in this issue are about computational methods in IDA. Johnson and Giraud-Carrier in the third article of this issue argue that based on empirical evidence, ensembles with adequate levels of pairwise diversity among a set of accurate member algorithms can significantly outperform any of the individual algorithms. The authors show that there is natural tension between the pairwise diversity of ensemble members and their individual accuracy. While efficient ensembles can be built with stronger forms of diversity, they also suffer in overall accuracy. On the other hand, ensembles built with weaker forms of diversity can be very accurate, but tend to be significantly more computationally expensive. Along the same line of research, Zhou *et al.* in the next article argue that ensembles with adequate levels of pairwise diversity among a set of accurate member algorithms can significantly outperform any of the individual algorithms. The authors show that there is a natural tension between the pairwise diversity of ensemble members and their individual accuracy. The authors also demonstrate that while efficient ensembles can be built with stronger forms of diversity, they also suffer on overall accuracy. Yoo and Bow in the next article of this group discuss that spatial co-location mining is a useful tool for discovering spatial association patterns of feature sets which are frequently observed together in nearby geographic space. The authors introduce the problem of mining reduced sets of co-location patterns in order to concisely represent interesting spatial relationship patterns. They propose an algorithmic framework to discover maximal co-location patterns and closed co-location patterns as well as all prevalent co-location patterns, and present the algorithm details for each pattern discovery.

Their experimental results show that the framework reduces candidate feature sets effectively and finds co-location patterns efficiently.

The third group of articles are about clustering and imbalanced data. Geng and Luo in the first article of this group discuss that time-series classification and class imbalance problem are two common issues in a multitude of real-life scenarios. The authors explore both issues with deep convolution neural networks (CNNs). They propose an adaptive cost-sensitive learning strategy to address this problem where they modify a standard CNN to a cost-sensitive network (CS-CNN), which is able to punish the misclassified samples using a class-dependent cost matrix. Their experiments show that their modified networks are superior in all metrics. Peng *et al.* in the next article of this group introduce a general framework for multi-label learning targeting class correlations and class imbalance. The proposed framework incorporates topic modeling to seamlessly address both problems of multi-label learning and class imbalance. The authors show that these frameworks can allow even the most naıve methods, such as Binary Relevance, to perform similarly to state-of-the-art methods. Vovan in the last article of this group demonstrates some interesting results for Cluster Width of probability Density (CWD) functions. The authors propose a measure called similar coefficient to evaluate the quality of the established clusters. Furthermore, the authors use CWD as a criterion to build two algorithms: to determine the suitable number of clusters and to analyse the fuzzy clusters. The numerical examples given illustrate the proposed algorithms and prove their advantages over existing methods.

And finally, the last group of articles in this issue are about enabling techniques and novel methods. In the first article of this group, Tuomchomtam and Soonthornphisaj argue that because of the vast and growing number of subreddits, users in the Reddit social media website need to discover and familiarize themselves with all existing communities before submission. The authors propose new feature sets for an online community which are text posts ratio, the average length of text in the post and the domain-specific features. The proposed framework successfully identifies and collects textual communities by finding their representatives using a combination of clustering and logistic regression algorithms. Their comprehensive experimental evaluations on Reddit dataset reveal high precision. Mirzaei *et al.* in the next article of this group discuss the topic of resource assignment in cooperative energy heterogeneous systems with non-orthogonal multiple access. They propose an architecture in which resource allocation and user association frameworks would be reconfigured because conventional schemes use orthogonal multiple access. They suggest a novel approach for optimal power allocation and user association techniques to achieve the maximum degree of energy efficiency in which the quality of experience parameters are assumed to be bounded during multi-cell multicast sessions. The authors demonstrate the effectiveness of the suggested approach through their experiments and numerical results. In the eleventh article of this issue Alani and Osunmakinde discuss the topic of electricity consumption prediction in smart homes and its effective management. They argue that effective planning through intelligent data analysis of the electricity load is needed to enable a sustainable distribution among consumers. The authors envision that such an intelligent analysis and planning approach is activated by the need to visualize the data and predict future electricity consumption within a short period, considering how weather variables affect predictions. Their research proposal includes a near-zero cooperative probabilistic scenario analysis and decision tree (PSA-DT) model to address the predictive errors facing state-of-the-art models, and analyses the effect each weather profile has on the cooperative model. Their experiments show that the PSA-DT model outperforms state-of-the-art models in terms of accuracy to a near-zero error rate. The last article of this issue by Wang *et al.* is about traffic flow prediction that plays a crucial component in today's transportation management. The authors propose a novel regression framework for short-term traffic flow prediction with automatic parameter tuning in which Support Vector Regression is the primary regression model for traffic flow prediction and the Bayesian Optimization being the

major method for parameter selection. In their proposed approach, the optimal short-term traffic flow regression model is constructed through repeated Gaussian Process update and iterative multiple training of the model. Their experimental results show that the accuracy of proposed method is superior to several existing techniques listed in the literature.

In conclusion, we would like to thank all the authors who have submitted the results of their excellent research to be evaluated by our 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