

Editorial

Dear Colleague:

Welcome to volume 25(3) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the third issue for our 25th year of publication. This issue contains thirteen articles representing a wide range of topics related to the theoretical and applied research in the field of Intelligent Data Analysis.

The first group of articles are about advanced data preprocessing and data engineering. Garcia *et al.* in the first article of this group present a novel approach for n-gram generation in text classification which is adapted to prune word sequences by combining three feature selection techniques. This approach reduces the number of branches to be explored, speeding up the discovery process and making the construction of all the word sequences tractable. Their Experiments on text classification datasets for sentiment analysis demonstrate that their proposed approach yields the best predictive performance when compared with other feature selection approaches, while also facilitating a better understanding of the words and phrases that explain a given task. Masoudi *et al.* in the second article of this group propose an approach for the fusion of structural and functional brain data with a deep learning-based model to take advantage of data fusion and increase the accuracy of schizophrenia disorder diagnosis. The proposed method consists of an architecture of 3D convolutional neural networks that are applied to magnetic resonance imaging, functional magnetic resonance imaging, and diffusion tensor imaging extracted features. The proposed method is validated on the COBRE dataset, where an average accuracy of 99.35% is reported. The proposed method demonstrates promising classification performance and can be applied to real data. The next article by Wang and Sun is also about handling imbalanced data where they present a SMOTE hyperparameter optimization approach for dealing with this class of data. The model proposed in their approach presents an outstanding performance and strong generalization ability, thus effectively solving imbalanced classification and significantly improving the classification accuracy of the minority class. The fourth article of this issue by Wang *et al.* is about sentiment classification based on weak tagging information and imbalanced data. The authors argue that in the sentiment classification of text data, such as online reviews, traditional deep learning models are dedicated to algorithm optimization but ignore the characteristics of imbalanced distribution of the number of classified samples and the inclusion of weak tagging information such as ratings and tags. The authors used the traditional deep learning model, which is an approach for random oversampling and cost sensitivity to increase the contribution of a minority of samples to the model loss function and avoid the model biasing to the majority of samples. Their experimental results show that the method is considerably better than traditional deep learning models in the sentiment classification. In the last article of this group, Frnda *et al.* present a novel perceptual evaluation method of video quality based on neural network. The proposed approach allows one to interconnect results obtained from video objective and subjective assessment methods in the form of a neural network. The authors argue that no unified interpretation scale has been standardized for both approaches, therefore it is difficult to determine the level of end-user satisfaction obtained from the objective assessment. Thus, contribution of the proposed method lies in description of the way to create a hybrid metric that delivers fast and reliable subjective score of perceived video quality for internet television broadcasting companies.

The second group of articles in this issue are about supervised and unsupervised techniques in IDA. Si *et al.* in the first article of this group argue that storyline extraction aims to generate concise summaries of related events unfolding over time from a collection of temporally-ordered news articles. The authors explain that some existing approaches to storyline extraction are typically built on probabilistic graphical models that jointly model the extraction of events and the storylines from news published in different periods. However, their parameter inference procedures are often complex and require a long time to converge. The authors propose a novel unsupervised neural network-based approach to extract latent events and link patterns of storylines jointly from documents over time. The authors present the evaluation of their proposed model on three news corpora where their experimental results show that it outperforms state-of-the-art approaches with significant improvements. Zhao *et al.* in the seventh article of this issue present a novel krill herd algorithm with orthogonality applied to data clustering where they incorporate orthogonal learning to enhance the performance of this class of algorithms. Their proposed algorithm is applied to solve numerical problems, and its robustness is verified based on the simulation results in handling data clustering problems selected from the UCI Machine Learning Repository. Their experimental results illustrate that the proposed approach is superior to or at least competitive with other representative clustering techniques. The next article by Meng *et al.* is also about classification where the authors present an attention network based on feature sequences for cross-domain sentiment classification. Their approach uses a three-layer convolutional neural network to perform deep feature extraction on the text, and then uses a bidirectional long short-term memory to capture the long-term dependency relationship among the text feature sequences. Their experimental results on cross-domain sentiment analysis tasks demonstrate that the proposed approach can significantly outperform the state-of-the-art methods for cross-domain sentiment classification problems. The ninth article of this issue by Wang *et al.* is about bagging k-dependence Bayesian network classifiers where the authors presents an efficient and effective bagging approach, that learns a set of independent Bayesian network classifiers from disjoint data subspaces. In this approach, the number of bits needed to describe the data is measured in terms of log likelihood, and redundant edges are identified to optimize the topologies of the learned BNCs. Their experimental evaluation on a number of publicly available datasets from the UCI machine learning repository reveals that the proposed algorithm achieves a competitive classification performance compared with state-of-the-art methods. The next article by Qia *et al.* is about cross-modal variable-length hashing based on hierarchy where hash learning is widely used frequently in many cross-modal retrieval scenarios. The authors propose a novel method which can learn the hash functions more accurately to improve retrieval performance, and also reduce the computational costs and training time. Their extensive experimental results show the superior performance of this approach compared with recent state-of-the-art cross-modal methods on several well-known data sets. The last article of this group by Boskabadi *et al.* is about regression trees with splitting based on changes of dependencies among covariates. The authors propose a new approach for regression trees that considers the dependency structures among covariates for splitting the observations. To assess the accuracy of the proposed model, various criteria are defined in their experiments. The performance of the new approach is assessed by conducting a Monte-Carlo simulation study on two real data sets on classification and regression problems.

The last two articles in the third group are about link prediction. Pham and Du in the first article of this group argue that link prediction on heterogeneous information network is a challenging problem due to the complexity and diversity in types of nodes and links and present a topic-driven network embedding model for link prediction in content-based heterogeneous information network. The authors propose a novel approach of topic-driven multiple meta-path-based heterogeneous information network representation learning framework which leverages the quality of node representations by combining

multiple meta-paths as well as calculating the topic similarity weight for each meta-path. To validate their approach, the authors apply their model in solving several link prediction tasks in both content-based and non-content-based applications. Their experimental outputs demonstrate the effectiveness of their proposed model which outperforms recent state-of-the-art representation learning models. And finally, Wu *et al.* in the last article of this issue present a directed link prediction method that is based on graph convolutional network and social ranking theory. Their main motivation is from the reciprocated and unreciprocated nature of social ties, where they separate nodes in the neighbor subgraph of the missing link into a higher-ranked and a lower-ranked set. Their empirical study on eight real-world datasets shows that their proposed method is capable of preserving rich information related to directed link direction and consistently performs well on graphs from numerous domains.

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. Over the last few years, our submission rate has exceeded 600 manuscripts per year, with an acceptance rate of around 12–15%. 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