Intelligent Data Analysis 25 (2021) 1345–1347
DOI 10.3233/IDA-210004
IOS Press

Editorial

Dear Colleague:
Welcome to volume 25(6) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the sixth and the last issue for our 25th year of publication. It contains fourteen 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 in IDA. Hsu et al. in the first article of this group discuss how to analyze mixed-type data by using word embedding for handling categorical features. The authors argue that the limitation of most existing methods is lack of appropriate numeric representations of categorical values. They address this deficiency by transforming categorical values to their numeric representation so as to facilitate various analyses of mixed-type data. They further propose a transformation method that preserves semantics of categorical values with respect to the other values in the dataset. The proposed method is verified and compared with other methods on extensive real-world datasets. Saifan and Lataifeh in the next article present an approach for privacy preserving defect prediction using generalization and entropy-based data reduction. They use Tomek link and AllNN data reduction approaches to discard noisy records that may affect the usefulness of the shared data. The proposed approach considers diversity of sensitive attributes as an important factor to avoid inference and background knowledge attacks on the anonymized data. Their experiments conducted on several benchmark software defect datasets, using both data quality and privacy measures, evaluate the proposed approach. Li and Yu in the third article of this issue present an approach to detecting a multigranularity event in an unequal interval time series that is based on self-adaptive segmenting. In view of the trend features of a time series, a self-adaptive segmenting algorithm is proposed to divide a time series into unfixed-length segmentations based on the trends. Then, by clustering the segmentations and mapping the clusters to different identical symbols, a symbol sequence is built. The experimental results reported in this article demonstrates how the proposed approach can achieve higher efficiency and accuracy than existing algorithms. Wang et al. in the last article of this group discuss how to alleviate the independence assumptions of averaged one-dependence estimators by model weighting. The authors argue that the topology of each one-dependence estimator can be divided into a set of local directed acyclic graph (DAG) that is based on the independence assumption, for which a multivariate mutual information is introduced to measure the extent to which the DAGs fit the data. The proposed approach is validated on several benchmark datasets from UCI machine learning repository.

The second group of articles in this issue are about unsupervised and supervised leaning methods in IDA. Tang et al. in the first article of this group explain that most density-based clustering algorithms have the problems of difficult parameter setting, high time complexity, poor noise recognition, and weak clustering for datasets with uneven density. To overcome this problem, they propose an approach that finds the demarcation point from the Augmented Cluster-Ordering and uses the reachability-distance as the radius of neighborhood of its corresponding cluster. Their experimental results show that the proposed approach has the lowest time complexity, and outperforms other algorithms in parameter setting and noise recognition. Bai et al. in the next article of this group present a differential evolution algorithm-based multiple-factor optimization method for data assimilation. They argue that the search methods for

optimized parameters have substantial effects on the forecast accuracy of ensemble data assimilation systems. The authors claim that by combining with fast-searching differential algorithms, they can retrieve the most ideal parameter combinations. It is reported that the new method is capable of outperforming previous search algorithms under both perfect and imperfect model scenarios. Chen et al. in the seventh article of this issue present a time-series data dynamic density clustering approach as in many clustering problems, the whole data is not always static and over time, part of it is likely to be changed, such as updated or erased. The authors analyze the transition rules of data set and cluster structure when the time slice shifts to next. They find there is a distinct correlation of data set and succession of cluster structure between two adjacent ones, which means one can use it to reduce the cost of whole clustering process. Their results show that their approach can get high accuracy while reducing the overall cost markedly. Chunying et al. in the next article of this group present a set pair k-modes clustering algorithm for incomplete categorical matrix data. Based on this approach, first the correlation theory of set pair information granule is introduced into k-modes clustering. Secondly, based on multiple criteria, it is decided whether the sample belongs to multiple clusters or not. Finally, the experimental results show that the set pair k-modes clustering algorithm can effectively handle incomplete categorical matrix data sets, and has good clustering performance. Kang and Jun in the ninth article of this issue introduce a mutual information-based multi-output tree learning algorithm. The performance of the proposed tree learning algorithm is similar to or better than that a multi-output version of CART algorithm and the time complexity of the proposed algorithm is significantly reduced compared to CART. The last article of this group by Golshanrad et al. is about the best classifier combination using meta-learning and a genetic algorithm. The proposed approach has three main components: Training, Model Interpretation and Testing. The authors present extensive experimental results that demonstrate the performance of their proposed method which achieves superior results in a comparison with three other methods and, most importantly, is able to find novel interpretable rules that can be used to select the best combination of classifiers for an unseen dataset.

The last group of articles in this issue are about enabling techniques and applied methods in IDA. In the first article of this group Wang et al. propose a novel end-to-end network ovarian cancer model, which can extract the characteristics of ovarian cancer more efficiently. In this method, deformable convolution is used to enhance the model's ability to learn geometric deformation in space. Their experiments are conducted on datasets collected from several hospitals in China that show good results on both accuracy and efficiency. The twelfth article of this issue by Rocha and Rodrigues is about a method for forecasting emergency department admissions. The approach describes a solution developed for the presentation of hourly, four-hour, eight-hour and daily number of admissions to a hospital emergency department. The models generated the most accurate hourly time predictions were the recurrent neural network with one-layer and three layers and also XGBoost. The authors report that in terms of efficiency, the XGBoost method has by far outperformed all others. Yao et al. in the thirteenth article of this issue present a system to improve the robustness of speech recognition systems through classifying stressed speech caused by the psychological stress under multitasking workloads. The authors propose a multi-feature fusion model based on the attention mechanism to measure the importance of segments for stress classification. The proposed model is compared with traditional methods on a Chinese emotion corpus and Fujitsu stressed speech corpus, where the results show that the proposed model has better performance in speaker-independent stress classification. And finally Safaei and Habibi-Asl present a multidimensional indexing normalization technique for medical image retrieval that is used to reduce redundancy in relational database design. Data structure of the proposed multidimensional index and also different required operations are designed to create and handle such a multidimensional index. Their

results show that the proposed indexing technique has a good performance in terms of memory usage, as well as execution time for the usual operations and can improve the information retrieval process for healthcare search engines.

In conclusion, we would like to thank all the authors who have submitted their manuscripts with the results of their excellent applied and theoretical research to be evaluated by our referees and published in the IDA journal. This issue concludes our 25 years of publication for which we are very grateful to all colleagues and our publisher, the IOS Press. Over the last few years, our submission rate has exceeded 650 manuscripts per year, with an acceptance rate of around 12–15%. We are also glad to announce that our impact factor has increased by 32% since last year (from 0.651 to 0.860). In addition, there is a special issue of the IDA journal under preparation entitled: "**Cloud/Fog/Edge Computing for Urban Big Data**" that is scheduled for publication for next year. Interested authors can contact: Prof. Jerry Chun-Wei Lin (jerrylin@ieee.org) to submit their manuscript. 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