

Editorial

Dear Colleague:

Welcome to volume 26(1) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the first issue for our 26th 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. Al-Dayani et al. in the first article provide a systematic literature review on the improvements of bat algorithm (BA) for optimal feature selection. The authors argue that many researchers have proposed different techniques to handle feature selection problems. However, there is a lack of systematic review on BA which could shed light on its variants. The authors explain that a range of improvements made to the BA varies, where in their study the authors have employed a standard systematic literature review method on four scientific databases. Several critical and significant findings based on their investigation are reported in this article which can be used as a guideline for scientists in the future to do further research. The second article by Shu et al. is about information granularity-based incremental feature selection for partially labeled hybrid data. The authors argue that feature selection can reduce the dimensionality of data effectively. Most of the existing feature selection approaches using rough sets focus on the static single type data. However, in many real-world applications, data sets are the hybrid data including symbolic, numerical and missing features. In this article, incremental feature selection algorithms based on information granularity are developed for dynamic partially labeled hybrid data with the variation of an object set. Their extensive experimental results on different data sets demonstrate that compared with the non-incremental feature selection algorithms, incremental feature selection algorithms can select a subset of features in shorter time without losing the classification accuracy, especially when the group of objects changes dynamically, the group incremental feature selection algorithm is more efficient. The third article of this issue by Sun and Yu is about a novel multi-class feature selection approach in which the authors propose a novel method to solve the regularized least squares problem directly based on iterative hard thresholding, which can produce exact row-sparsity solution for weights matrix, and features can be selected more precisely. Their proposed method is verified on eight biological datasets where their experimental results show that the proposed method can achieve higher classification accuracy with fewer number of selected features than the approximate convex counterparts. Aggoune in the last article of this group propose an intelligent data integration approach for heterogeneous relational databases containing incomplete and uncertain information. The author proposes a novel approach for a mediation-based integration for integrating these types of information from heterogeneous relational databases. The integration process of this approach is based on the use of fuzzy logic and semantic similarity measures for more effective integration of incomplete and uncertain information. A thorough experiment presented in this article verifies the approach improves the performance of data integration under various configurations.

The second group of articles in this issue are about unsupervised and supervised learning methods in IDA. Wang et al. in the first article of this group present a clustering approach that is based on adaptive local density (ALD) with evidential assigning strategy. The authors argue that in density peaks clustering,

many other density metrics fail to detect cluster centres on multi-density datasets, where the ALD deals with the tasks very well since it can better utilize the local information. Simulations on both synthetic and real-world datasets demonstrate the outstanding performance of the proposed approach compared with density peak clustering and some of its successors. In the second article of this group Angel Deborah et al. present a contextual emotion detection approach for text using Gaussian process and tree based classifiers. The authors argue that as the use of digital agents have increased in text messaging applications, it is essential for these agents to provide sensible responses to its users. The results of their experiment obtained are compared with decision tree and ensemble models. Out of the five models built on a small dataset with class imbalance, it has been found that Gaussian Process classifier predicts emotions better than the other classifiers. The seventh article of this issue by Anandaraj and Alphonse is about a tree based ensemble for enhanced prediction of Epileptic Seizures (TEEP). This approach presents a seizure prediction model that performs data extraction and feature engineering to enable effective demarcation of preictal signals from interictal signals. The proposed Tree based Ensemble for Enhanced Prediction (TEEP) model is composed of three major phases; the feature extraction phase, feature selection phase and the prediction phase. The resultant features are passed to the boosted ensemble model for training and prediction. The TEEP model is analyzed using the Epileptic Seizure Recognition Data from several sources. Results from these datasets indicate effective performances. In the next article of this group, Jia et al. redefine community based on theoretical analysis and propose an overlapping community discovery algorithm that is based on the local interaction model. By fusing node attributes and structural information, the authors first propose an improved density peak fast search method to obtain multiple core nodes in the community. Then, according to the interaction range and interaction mode of the core node, they establish a local interaction model of the core node. Their experimental results show that compared with other similar community discovery algorithms, the proposed method outperforms the state-of-the-art approaches for community detections. Li et al. in the ninth article of this issue present a temporal link prediction in directed networks that is based on self-attention mechanism. The authors argue that development of graph neural networks (GCN) makes it possible to learn structural features from evolving complex networks. The authors propose a deep learning model based on GCN and self-attention mechanism. The proposed model adopts an autoencoder architecture, which utilizes graph attentional layers to capture the structural features of neighbourhood nodes, as well as a set of graph convolutional layers to capture motif features. The authors present a series of comparative experiments on four realistic networks to validate the effectiveness of their proposed approach. In the last article of this group Zhang et al. present a particle swarm optimization pattern recognition neural network approach that is applied for transmission lines faults classification. They compare the performance of their approach with the ones of K-Nearest Neighbor, Decision Tree, PSO-KNN and PSO-DT classifiers. Their results show that the classification accuracy of PSO-PRNN outperforms most traditional classifiers when being applied to the data collected from all three bureaus.

The last group of articles in this issue are about enabling techniques and applied methods in IDA. In the first article of this group Qiu present a community-based algorithm for influence maximization under the weighted cascade model on social networks. The authors explain that influence maximization is a problem of selecting k nodes from social networks to make the expected number of the active nodes maximum. The authors argue that the existing influence maximization algorithms with high accuracy are usually difficult to be applied to the large-scale social networks. To solve this problem, the authors propose a new algorithm that utilizes the Louvain algorithm to divide the large-scale networks into some small-scale networks. Their experimental results on six real-world datasets demonstrate that their proposed algorithm outperforms all comparison algorithms when comprehensively considering the accuracy and efficiency.

Allahgholi et al. in the twelfth article of this issue argue that analyzing the relationships among various drugs is an essential issue in the field of computational biology and present an approach suitable to analyse data from drug-drug relationships to drug repurposing. They introduce a general-purpose method that applies deep learning on scientific literature to automatically extract the graph of syntactic and semantic relationships among drugs. The authors demonstrate that the proposed method justifies the predictive power of textual data in PubMed abstracts and can be used to predict repurposing drugs with high accuracy. The next article of this issue by Yuan et al. is an adaptive simulated annealing and artificial fish swarm algorithm for the optimization of multi-depot express delivery vehicle routing. The authors propose an adaptive simulated annealing and artificial fish swarm algorithm in which the basic ideas are to use a “certainty” probability to accept the worst solution through the Metropolis criterion. Their experimental results show that the proposed algorithm outperforms four well-known algorithms. And finally, Takhanov and Kolmogorov in the last article of this issue present an approach for combining pattern-based conditional random field (CRF) and weighted context-free grammars. The authors propose a grammatical pattern-based CRF model that combines the two in a natural way. The focus of this article is to analyze the complexity of inference tasks such as computing MAP. They present a polynomial-time algorithm that is a faster version for a special case called Interaction Grammars.

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 is the start of our 26th years of publication for which we are very grateful first to all colleagues and secondly to 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 2022. 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