

Editorial

Dear Colleague:

Welcome to volume 27(3) of the Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the third issue for our 27th year of publication. It consists of 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 in this issue are about state of the art supervised and unsupervised learning methods in IDA. In the first article of this group Hernández-Gómez et al. present an agglomerative hierarchical clustering approach to identify coexisting bacteria in groups of bacterial vaginosis (BV) patients. The main aim of this study is to explore a dataset of patients with BV to determine a more informed number of groups to create a path for further analysis of bacteria's coexistence. The authors introduce an Agglomerative Hierarchical Clustering (AHC) algorithm to a BV dataset from an urban population in South-Eastern Mexico and three classes (BV-positive, BV-negative, BV-indeterminate). They provide evidence that via the AHC algorithm, it was possible to identify an optimal number of clusters with high intra-similarity and inter-dissimilarity. He, in the second article of this issue, presents an active learning approach for ordinal classification on incomplete data. The author argues that although cutting-edge imputation methods can be used to impute the missing values before commencing active learning, inaccurately imputed instances are unavoidable and may degrade the ordinal classifier's performance once labelled. The author proposes penalizing the query selection with a novel imputation uncertainty measure that combines a feature-level imputation uncertainty and a knowledge-level imputation uncertainty. Extensive experiments on ten public ordinal classification datasets with varying value missing rates show that the proposed approach outperforms several baseline methods. Jin et al. in the next article of this issue present a clustering-based adaptive synthetic minority oversampling technique for imbalanced data classification. The approach is based on decomposing minority instances into sub-clusters according to their connectivity in the feature space and then selecting minority sub-clusters which are relatively close to the decision boundary as the candidate regions to oversample. The experimental evaluation shows that the proposed approach improves the occurrence of overlaps between the minority and majority instances. In the fourth article of this issue Hao et al. present an approach to identify local morphological patterns for time series classification, arguing that the key problem of time series classification is the similarity measure between time series. The authors propose a novel similarity measurement method named dynamic time warping based on the local morphological pattern (MPDTW), which first decomposes the local sub-sequences of time series using discrete wavelet transforms for extracting the local structure information. Their experiments have been carried out on the classification tasks of the large datasets and the results show that the proposed method outperforms the existing baselines. Xiang et al. in the next article of this group present a selection technique that is based on relative density and density peaks clustering for k nearest neighbour classification. They also present a new edition method based on relative density and a new condensing method based on relative density and density peaks clustering. Their extensive experiments prove that their proposed approach outperforms 6 popular hybrid prototype selection techniques on extensive real data sets in weighing accuracy and reduction of the running time

is also acceptable. The sixth article of this issue by Hu et al. is entitled: modeling multi-attribute and implicit relationship factors with self-supervised learning for recommender systems. The authors have developed a novel recommendation model which explicitly proposes and models multi-attribute and implicit relationship factors for collaborative filtering recommendation. Their extensive experimental results on three public datasets not only show the significant performance and robustness gain of the proposed model but also prove the effectiveness and interpretability of fine-grained implicit factors modeling. In the next article of this issue Hu et al. present a cross-community shortcut detection method that is based on network representation learning and structural features. The authors propose an edge PageRank to mine shortcuts that are located among communities and play an important role in the spread of public opinion. Their experimental results on five different sizes of social networks verify whether the inferred shortcut is indeed more useful for information dissemination, and the utility of three sets of edges inferred by different methods. The last article of this issue Wan et al. present an approach entitled: attribute-guided and attribute-manipulated similarity learning network for fashion image retrieval. The proposed approach is based on learning multiple attribute-specific embedding spaces that measures fine-grained similarity in the corresponding spaces. It can accurately map different attributes to the corresponding areas of the image, thereby facilitating the feature fusion of two different modalities of text and image, including enhancement and replacement. With comprehensive validation on two real-world fashion datasets, the authors demonstrate that their method can effectively leverage semantic knowledge to improve image retrieval performance, including attribute-guided and attribute manipulated retrieval tasks.

The second group of articles are about enabling techniques and innovative case studies in IDA. Phan and Do in the first article of this group present a novel framework to enhance the performance of training distributed deep neural networks (DNN). The proposed approach helps to develop applications that train DNN in a distributed environment in a simple, natural, and flexible way. The authors present experiments by training a convolutional neural network model with large datasets to demonstrate the flexibility and effectiveness of their proposed approach. You et al. in the tenth article of this issue present a quantum-behaved particle swarm optimization (PSO) with dynamic grouping search strategy. The approach is based on the grouping particles dynamically into two subpopulations, which are assigned to implement the exploration and exploitation search. Their experimental comparison is conducted between their proposed approach and seven state-of-the-art PSO variants which show that the proposed approach has a promising performance in terms of the solution accuracy and the convergence speed on the majority of these test functions, and especially on multimodal problems. In the eleventh article of this issue, Chen et al. explain their research on community evolution based on node influence and multi-attribute fusion and present a novel approach to tackle this problem. The model combines the characteristics of community stability and community difference, and redefined evolutionary events. As part of their experiments, the authors evaluate the effectiveness of their proposed model in identifying community evolution events on different data sets. Hong et al. in the next article of this group present an approach for learning binary codes for fast image retrieval with sparse discriminant analysis and deep auto-encoders. Their proposed approach utilizes sparse discriminant analysis to select the most important original feature set, and solve the small class problem in the relevance feedback. The approach also takes advantage of both labelled and unlabelled samples to improve the retrieval precision. Their experimental results on three databases demonstrate that the proposed method obtains competitive precision compared with other state-of-the-art image retrieval methods. The thirteenth article of this issue by Lyu et al. is about positive public opinion guidance (PPOG) model based on dual learning in social network. The approach uses use a fast unfolding algorithm to divide social networks into the public opinion guidance communities.

Their experiments demonstrate beneficial effects of their proposed model to capture useful information in the network topology. And finally, Hassan et al in the last article of this issue discuss the efficacy of confinement policies on the COVID-19 spread dynamics in the early period of the pandemic. The authors use a clustering-based approach, namely Hierarchical Cluster Analysis (HCA), on time-series data to capture the spread patterns at various countries. They particularly investigate the confinement policies adopted by different countries and their impact on the spread patterns of COVID-19. The authors found that lockdowns become less effective in densely populated regions, and lack of testing, contact tracing, and social awareness in some countries forestall people from self-isolation and maintaining social distance.

In conclusion for the third issue of our volume 27, I would like remind you that as the founding Editor-in-Chief of the IDA journal, I am gradually wrapping up my duties and I have transferred the responsibility to my colleague, Dr. Jose Maria Pena (from Oxford, UK), whom I have known since 1997. Please join me in welcoming Dr. Pena to the position of the Editor-in-Chief of the IDA Journal. We are also glad to announce that our impact factor has increased by over 50% since last year (from 0.860 to 1.321). 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 ***Dr. J.M. Pena Founding***
Editor ***Editor-in-Chief***