

Editorial

Dear Colleague:

Welcome to volume 27(4) of the Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the fourth issue for our 27th year of publication. It consists of sixteen 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 Kim and Jun present a dynamic mutual information-based feature selection model for multi-label learning. The idea is to reduce the dimensionality of the input space while improving or maintaining classification performance. This approach can handle redundancy among features controlling the input space. The authors compare the proposed method with some existing problem transformation and algorithm adaptation methods applied to real multi-label datasets using the metrics of multi-label accuracy and hamming loss. Their results show that the proposed method demonstrates more stable and better performance for nearly all multi-label datasets. In the second article of this issue Li *et al.* explain filter pruning via feature map clustering. Filter pruning is a sub-direction of network compression research, which reduces memory and computational consumption by reducing the number of parameters of model filters. The authors have discovered that feature maps are not fully positively correlated with the sparsity of filter weights by observing the visualization of feature maps and the corresponding filters. The authors have defined a criterion called redundancy index to rank the filters and introduced it into their filter pruning strategy. Their extensive experiments demonstrate the effectiveness of their approach on different model architectures. Yu *et al.* in the third article of this issue present a trajectory personalization privacy preservation method based on multi-sensitivity attribute generalization and local suppression. To meet the users' personalized privacy requirements and ensure the utility of trajectory location and sensitive information, the authors define different security levels for each trajectory by calculating the correlation between sensitive attributes to establish a sensitive attribute classification tree. The authors then generalize sensitive attributes based on privacy preservation levels for each trajectory, the trajectory data still at risk of privacy leakage after generalization was locally suppressed. Their experimental results on real datasets demonstrate that their method could improve data availability while preserving privacy. In the fourth article of this issue, Liu *et al.* discuss the idea of safe co-training for semi-supervised regression for which one of the key issues being quality of pseudo-labels. The authors propose a co-training algorithm for regression with two new characteristics, (i) a safe labeling technique and (ii) a label dynamic adjustment strategy. Their results show that their proposed method is superior to other co-training style regression algorithms and state-of-the-art semi-supervised regression algorithms. In the next article of this issue He presents an active ordinal classification by querying relative information. The author argues that collecting and learning with auxiliary information is a way to further reduce the labeling cost of active learning. The idea is to convert the absolute and relative information into the class interval-labeled training instances form by introducing a class interval concept and two reasoning rules. Following that the approach specifies that each query pair consists of an unlabeled instance and a labeled instance. Extensive experiments on twelve public datasets validate that the proposed method is

superior to the state-of-the-art methods. In the sixth article of this issue, Yao *et al.* present an approach for learning a regularized reinforcement agent for keyphrase generation which is intended for condensing the content from the source text to the target concise phrases. The authors present an approach that includes actor-critic based-reinforcement learning control and policy regularization under the first principle of minimizing the maximum likelihood estimation criterion by a sequence-to-sequence (Seq2Seq) deep learning model, for efficient keyphrase generation. Their extensive experiments show that the proposed method brings improvement in terms of the evaluation metrics on scientific article benchmark datasets. The seventh article by Radovanovic *et al.* introduce a postprocessing technique, called fair additive weighting for achieving group and individual fairness in multi-criteria decision-making methods. The proposed methodology is based on changing the score of an alternative by imposing fair criteria weights which can be successfully used in multi-criteria decision-making methods where the additive weighting is used to evaluate scores of individuals. Their results show their proposed approach achieves group fairness in terms of statistical parity, while also retaining individual fairness. Zheng *et al.* in the last article of this group present an approach called feature evolvable learning with image streams. The authors present a novel ensemble residual network, in which the prediction is weighted combination of classifiers learnt by the feature representations from several residual blocks, such that the learning is able to start with a shallow network that enjoys fast convergence, and then gradually switch to a deeper model when more data has been received to learn more complex hypotheses. Their experiments on both virtual and real scenarios over large-scale images, and the experimental results demonstrate the effectiveness of the proposed method.

The second group of articles are about enabling techniques and innovative case studies in IDA. Li *et al.* in the first article of this group present a parallel and balanced SVM algorithm targeting data-intensive computing. The approach is optimized on the basis of the traditional Cascade SVM algorithm, which solves the problems of data skew and the large difference between local support vector and global support vector. Their experimental results show that the two-classification tests extensively improve efficiency and the accuracy of classifications. In the tenth article of this issue Meira *et al.* present a unique data-driven predictive maintenance framework for railway systems. The proposed method assists in detecting failures and errors in machinery before they reach critical stages. The authors present an anomaly detection model following an unsupervised approach, combining the Half-Space-trees method with One Class K Nearest Neighbor. Their experiments show the proposed model produced few type-I errors, significantly increasing the value of precision when compared to other models. Gao *et al.* in the next article of this group present an improved hybrid structure learning strategy for Bayesian networks based on ensemble learning. The authors introduce the idea of parallel ensemble learning where they adopt the elite-based structure learner using genetic algorithm as the base learner. Their comparative experiments on the standard Bayesian network data sets show that the accuracy and reliability of the proposed algorithm are significantly better than other algorithms. In the twelfth article of this issue Wu and Zhang present an efficient intrusion detection method that is based on federated transfer learning and support vector machine with privacy preserving. In this approach the authors first build a transfer support vector machine model to solve the problem of data distribution differences among various organizations; then, under the mechanism of federated learning, the model is used for learning without sharing training data on each organization to protect data privacy. Their experimental results verify that the proposed method can achieve better results for detection and robust performance, especially for small samples and emerging intrusion behaviors. In the next article of this group Wang *et al.* discuss the process of exploiting the implicit independence assumption for learning directed graphical models. The authors argue that explicit independence assumptions can effectively and efficiently reduce the size of the search

space for solving the NP-complete problem of structure learning. The authors propose an extension to the k-dependence Bayesian classifier that achieves the bias/variance trade-off by verifying the rationality of implicit independence assumption. Their comprehensive experimental results on UCI datasets show that their proposed algorithm achieves competitive classification performance when compared to the state-of-the-art methods. The fourteenth article of this issue by Azoulay *et al.* is a presentation of machine learning techniques for received signal strength indicator prediction. The authors compare different machine learning techniques that can be used to predict the received signal strength values of a device, given the received signal strength values of other devices in the region. They consider various ML methods, such as multi-layer ANN, K nearest neighbors, decision trees, random forest, etc. for the prediction challenge and conclude that in environments where the size of data is relatively small, and close geographical points are available, a method that predicts the coverage of a point using the coverage near geographical points can be more successful and more accurate compared with other ML methods. Hichem et al in the fifteenth article of this issue present a discrete equilibrium optimization algorithm for breast cancer diagnosis. This article aims to study breast cancer diagnosis using the associative classification technique where it generates interpretable diagnosis models. Their comparison results show that the proposed approach can generate more accurate and interpretable diagnosis models for breast cancer than other algorithms. And finally, Zicai *et al.* in the last article of this issue present a Chinese text clustering algorithm that combine word rotator's distance with the K-means algorithm, and propose the WRDK-means algorithm, which uses word rotator's distance to calculate the similarity between texts and preserve more text features. The authors selected three suitable datasets and five evaluation indicators to verify the feasibility of the proposed algorithm. Their results seem to be very encouraging.

In conclusion for the fourth issue of our volume 27, I would like remind you that as the founding Editor-in-Chief of the IDA journal, this is the last editorial that I am writing as I have transferred the responsibility to my colleague, Dr. Jose Maria Pena. I founded the IDA journal during 1995-96 where it was officially launched in July 1996. The first issue was published in January 1997 when we started with less than 30 submissions per year. At the beginning of our production we only had four accepted articles per issue and four issues each year. Now that I am stepping down from the position of the Editor-in-Chief, I am glad that the IDA journal has become well known in the scientific community. For over 20 years we have published six issue per year and over the last few years our submission rate has exceeded 800 articles per year and this issue contains sixteen accepted article. 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**
Editor Founding **Editor-in-Chief**