Editorial

Dear Colleague:

Welcome to volume 19(4) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the fourth issue of 2015 that consists of twelve articles, all covering a range of topics related to the theoretical and applied research in the field of Intelligent Data Analysis.

The first three articles are on various forms of SVM classification, a supervised form of learning. Xie and Sun in the first article discuss the idea of twin support vector machines and propose a multi-view support vector machines that is intended to solve a pair of quadratic programming problems. The authors provide a detailed derivation of the Lagrange dual optimization formulation and further generalize a linear multi-view twin support vector machines into a nonlinear case by the kernel trick. Experiments reported in this article demonstrate the effectiveness of their proposed method. Kang and Cho in the next article of this issue discuss the drawbacks of multi-class SVM based classifications and propose a multi-class classification approach that is based on a one class SVM. The distinguishing point of their proposed approach is that the algorithm solves a one-class classification problem into several smaller sub-problems. Their experiments reported in the article show their approach outperforms other SVM-based multi-class classification methods. The last article of this group by Fu and Yang contains a new stopping criterion for active learning SVM. The proposed idea is based on low density separation concept which is extensively used in semi-supervised and unsupervised learning methods. Their proposed algorithm measures whether SVM₄rs separating hyperplane lies in the low density region. Their experiments show that the proposed stopping criterion strikes a good balance generalization performance and label cost.

The next four articles are on various forms of supervised and unsupervised learning. Lu *et al.* in the fourth article of this issue discuss some of the drawbacks in algorithms applied in extreme learning applications and propose ideas for an effective computations methods for extreme learning class of problems. They compare two of their proposed methods with Moore-Penrose inverse matrices for several extreme learning applications and demonstrate that their approach is much faster than the common used methods. Warintarawej *et al.* in the next article of this group discuss the problem of understanding software identifiers through various forms of learning and propose using domain concept categories to support automatic software understanding. The approach is based on character patterns of identifier names. The authors list their main challenges as being automatic splitting of identifier names into relevant constituent subnames and building a model associating such a set of subnames to predefined domain concepts. The authors propose an approach to this problem and report experiments performed on a number of real software source code. Zhi *et al.* in the next article of this issue discuss the topic of class imbalance from the viewpoint of ensemble pruning and propose a novel approach to improve classifiers performance. Their proposed approach is based on considering imbalance problem at the prediction stage, rather than the training stage that is more common. They also combine some existing methods that involve under-sampling. Their experiments on several data sets demonstrate the superior performance of their approach. And in the last article of this group, Setyohadi *et al.* view overlap in a dataset as an uncertainty problem and propose an overlap clustering algorithm, which involves the use of the discernibility

concept to improve the overlap clusters as an existing variant of the overlap clustering algorithm. Their experiment performed on a number of UCI data sets and described in the article demonstrate that their proposed method improves the performance and increases the accuracy of clustering while avoiding the time complexity problem.

The last five articles of this issue are about various forms of optimization and intelligent decision support. Dallaki *et al.* discuss the advantages of hybrid methods in optimization and classification, highlight the drawbacks of these methods with their high computational complexity in large data sets and propose an approach to solve this problem. Their proposed approach consist of a K-mode for sampling and a new fitness measure for the learning algorithm which result in speeding up the learning process. Ghazanfar in the ninth article of this issue discusses prediction capabilities of recommender systems and argue that these systems suffer from scalability, data sparsity, and over-specialization, among some of their disadvantages. The author proposes a unique generalized switching hybrid recommendation algorithm that combines machine learning with collaborative filtering recommender system. The experimental results reported in this article demonstrate the scalability and better performance of this system in terms of accuracy and coverage. Kim *et al.* in the next article of this issue propose a stock market instability index with a corresponding p-value that relies on model fitting for a stable period. The approach is based on a random walk model that combines a nonparametric model through a Bayesian model averaging. Their experimentation is based on using an Artificial Neural Network as a nonparametric model.

The last two articles on network learning are related to two domains, one in life sciences and the other in social networks. Rahmani *et al.* explain that understanding life science networks (such as protein-protein interaction) could be either individual (based on one disease) or network (based on multiple diseases). Their proposed approach is based on a network of 20 different human diseases where they show that their discovered network is biologically meaningful and capable of augmenting the initial prior knowledge of different diseases through sharing information across a set of highly related diseases. And finally in the last article of this issue Bhat and Abulaish discuss some of the challenges in discovering network patterns in community social networks where a node could belong to multiple communities or multiple smaller communities could be embedded within a large community. The authors propose a density-based overlapping community detection method that is intended for identify overlapping community structures in social networks. This approach does not require the neighborhood threshold parameters to be set by the user and utilizes a distance function which takes into account the weights of the edges in weighted networks. The article reports the efficacy of the proposed method through a number of experiments on real world and synthetic data sets.

In conclusion, with this issue of the IDA journal, which is Volume 19(4), we are pleased to see a consistent increase in the submission of high quality manuscripts to our journal. In addition to our six regular issues that now contain 11–12 articles, we expect to publish one special issue each year, which is normally related to a scientific conference for which organizers have submitted an interesting proposal. 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