Intelligent Data Analysis 20 (2016) 477–479
DOI 10.3233/IDA-160815
IOS Press

Editorial

Dear Colleague:

Welcome to volume 20(3) of Intelligent Data Analysis (IDA) Journal.

With this issue of the IDA journal, the third issue of 2016, we continue to celebrate the 20^{th} anniversary of our journal and our continuing success. This issue consists of twelve articles, all covering a wide range of topics related to the theoretical and applied research in the field of Intelligent Data Analysis.

The first four articles are on various forms of learning. Chao and Sun in the first article discuss maximum entropy discrimination (MED) and propose an approach named multi-kernel MED which is suitable for multi-view learning (MVL). Based on the kernel matrix obtained through multiple kernel combination, the authors proceed MVL within the MED framework. Their experimental results on multiple datasets demonstrate the effectiveness of their proposed approach which outperforms the single-view MEDs. Chen *et al.* in the second article of this issue discuss the side effects of sanitization in large data sets which is normally applied in order to ensure data privacy and concealment of sensitive information. The authors focus on privacy preserving in association rule mining and since there is a trade-off within different side effects, their approach is to minimize them from the view of multi-objective optimization. They propose a rule hiding approach based on evolutionary multi-objective optimization which is based on hiding sensitive rules through removing identified items. Their experimental results on real datasets illustrate that the proposed approach can achieve satisfactory results with fewer side effects. Bellodi *et al.* in the third article of this issue argue that management of business processes can support efficiency improvements in organizations where one of the most interesting problems could be mining and representation of process models in a declarative language. The authors present a Statistical Relational Learning approach that is suitable for work-flow mining and it takes into account both flexibility and uncertainty in real environments. The authors apply this approach to three datasets and compare it with DPML alone and a number of other algorithms. Their technique is shown to be effective in classification of new execution traces, showing higher accuracy and areas under the PR/ROC curves in most cases. In the last article of this group Chakraborty and Tripathy discuss privacy preserving anonymization of social networks using an Eigen Vector Centrality approach and argue that attackers can very easily exploit the sensitive information as well as the identity of the users from the raw data. The authors propose an approach based on the eigenvector centrality value of the individual nodes to achieve k-anonymity as well as l-diversity by adding noise nodes in the raw data. Through various measures and experiments, the authors establish the effectiveness of their proposed algorithm over a number of existing social network anonymization techniques.

The next group of articles are about various aspects of intelligent data analysis methods. Xu *et al.* in the first article of this group investigate the cluster ensemble problem and propose a cluster ensemble approach which is based on subspace similarity (CEASS). From a subspace similarity perspective, the authors then look for the optimal subspace which is most similar to the given subspaces corresponding to the cluster solutions to be combined. Among the techniques applied, K-means algorithm is applied with the minimum-maximum principle to cluster instances according to their coordinates in the embedding space. Their experimental results demonstrate that CEASS generally outperforms other algorithms

in terms of normalized mutual information and F1 measure and is also extremely efficient compared to hierarchical clustering algorithms. Moses and Deisy in the next article of this issue propose a mobile-Cardiovascular Abnormality Detection Engine (m-CADE) which encompasses a low computational feature extraction method, feature selection for identifying highly discriminative feature groups and an integrated classifier for abnormality detection. In this approach, novel features are extracted from time, frequency and statistical domains. The approach involves three feature combinations which are morphological descriptors from time domain, Daubechies Wavelet coefficients from the frequency domain and a set of statistical domain features. The authors incorporated these methods in an Android Application (App size 547 KB) which detects abnormality with an average accuracy of 98.9%, sensitivity of 99.2% and specificity of 97.6% in 1.3 sec. Saho *et al.* in the seventh article of this issue argue that the maximal information coefficient (MIC), a measure of dependence for two-variable relationships, can be used to discover the relationships between pair of variables in big data. The authors propose a new mathematical program model for calculating the value of MIC, design an algorithm to solve the model in big data environment, and apply the entire system to the analysis of railway accidents data. The experimental results presented in this article show that the proposed algorithm could find important relationships between two variables from big data. Ben Ishak *et al.* in the next article of this issue argue that the validation of any database mining methodology goes through an evaluation process where benchmarks availability is essential. In this article, the authors aim to randomly generate relational database bench-marks that allow to check probabilistic dependencies among the attributes. Focusing on probabilistic relational models (RPM), the authors extend Bayesian networks (BNs) to a relational data mining context that enables effective and robust reasoning about relational data structures. The proposed method allows to generate PRMs as well as synthetic relational data from a randomly generated relational schema and a random set of probabilistic dependencies which can be of interest for machine learning researchers to evaluate their proposals and for database designers to evaluate the effectiveness of the components of a database management system. Hu and Ferguson in the last article of this group present an algorithm, called global positioning graph matching (GPGM), to perform global network alignments between pairs of undirected graphs by minimizing a dissimilarity score over matched vertices. The approach is based on incorporating measures of vertex-vertex dissimilarity in a convex combination. The authors present evaluation of their approach in pairwise alignments of protein-protein interaction networks of *Xenopus laevis* (frog), *Rattus norvegicus* (rat), *Caenorhabditis elegans* (worm), *Mus musculus* (mouse), and *Drosophila melanogaster* (fly).

The third group of articles in this issue are on applied IDA research. Lin *et al.* in the tenth article of this issue discuss that most sequential pattern mining algorithms can only handle static databases, which is not practical in real-life situations. They propose an algorithm to maintain and update the discovered association rules for transaction deletion. The proposed algorithm first partitions the discovered sequential patterns into three parts with nine cases. The discovered sequential patterns of each case are then maintained and updated by the designed procedure. Their experiments show that the proposed algorithm has good performance when compared to other batch-mode algorithms or maintenance algorithms. Pazouki and Rahmati in the next article explain that a multi-object tracking problem in a network of cameras consists of two steps which are: (i) tracking objects in each camera and tacklets and (ii) associating tracklets and obtaining trace of the objects. In this article, the authors focus on the second step in which they propose a multiphase varational method and compare it against single-phase variational model applied for the association. Their experimental results on the real and synthetic datasets demonstrate that the multiphase model yields more complete tracking results, while the single-phase model yields more accurate tracking results. And finally Wang *et al.* in the last article of this issue propose an algorithm

for human activity recognition which is based on Gaussian Process Classifier (GPC). They first apply a hierarchical strategy to classify dynamic and static behaviors. Then, in each layer, three kinds of classification approaches are validated and evaluated for promoting recognition accuracy. The authors perform extensive experiments where they observe: (i) GPC achieves comparable classification accuracy with other classifiers under the same experimental condition, (ii) in case of less training samples, GPC outperforms the prominent Support Vector Machine (SVM) classifier, and (iii) unlike SVM, GPC is more robust to the high dimensional features.

In conclusion, with this issue of the IDA journal, which is Volume 20(3), we continue our celebration of the 20^{th} anniversary of the journal. The IOS Press, the publisher of the IDA journal, has several celebration events planned for 2016. These plans would be for two related conferences: AI-Stat 2016 in Cadiz-Spain and ECML/PKDD 2016 in Verona-Italy.

During 2016, in addition to our six regular issues, we will also publish a special issue related to CIARP series of conferences. 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