Editorial

Dear Colleague:

Welcome to volume 20(2) of Intelligent Data Analysis (IDA) Journal.

With this issue of the IDA journal, the 2nd issue of 2016, we celebrate the 20th anniversary of our journal. 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 five articles are on various forms of pattern mining. Plantevit *et al.* in the first article discuss the topic of pattern mining over data streams and argue that most of the current techniques attempt to discover relationships between time-point events but are not practical for discovering dependencies over interval-based events. The authors present a new approach to mine dependencies between streams of interval-based events that link two events if they occur in a similar manner. The two case studies presented in this article are outdoor surveillance of a building via video camera and motion sensors, and assistance for road deicing operations based on the humidity and temperature measurements at the urban scale. These applications demonstrate the efficiency and effectiveness of their proposed approach. Borchani *et al.* in the second article of this issue argue that despite a plethora of approaches that have been proposed to deal with mining concept-drifting data streams, most of these approaches can only be applied to uni-dimensional classification problems where each input instance has to be assigned to a single output class variable. The authors propose a novel adaptive method, named Locally Adaptive-MB-MBC (LA-MB-MBC), for mining streaming multi-dimensional data. Their experimental study carried out using synthetic multi-dimensional data streams shows the merits of their proposed method in terms of concept drift detection as well as classification performance. Rashedi *et al.* in the next article discuss the issue of choosing aggregates in hierarchical clustering where it depends on the aggregation operator (aggregator) used in the combination. To facilitate this task, the authors introduce new aggregator types, triangular norms and averaging operators, and then compile a list of main properties for these aggregators. An extra property which is needed in hierarchical clustering combination is also defined. A set of experiments are presented to select the optimized hierarchy aggregator from the variety of these aggregators. Alatrista-Salas *et al.* in the fourth article of this issue discuss the topic of spatio-temporal data in which spatial information (e.g., the location of an event) and temporal information (e.g., the date of the event) are included. The authors primarily focus on spatio-temporal patterns extraction from this kind of databases. They propose two algorithms where the first one uses a priori strategy and the second one is based on pattern-growth approach. Their experiments involve two different real datasets and some synthetic data where they evaluate the performance of the proposed algorithms. In the last article of this group Lazo-Cortes *et al.* revisit the topic of binary discernibility matrix and reformulate some basic concepts that allow them to build an algorithm for computing reducts. Their proposed algorithm takes advantage of the binary representation used for the discernibility matrix and depending on the fulfillment of certain pruning properties several candidates can be jumped (pruned). The proposed algorithm is able to compute all the reducts of an information table, and can also be used to find a reduct, or a certain number of them. Experiments over standard datasets show that their algorithm has a good performance.

The next group of articles are about various aspects of applied research in the field of intelligent data analysis. Saarenpaa *et al.* in the sixth article of this issue discuss the vast amount of information in society that has a huge and largely untapped potential for modelling societal and environmental phenomena. To discover correlations and relationships from this class of data, the authors propose to use data mining to identify relationships from public sector data related to hybrid vehicle adoption and small area level socio-demography. Based on some of their discovered models, they assess how favourable each area type is for the adoption. Their proposed approach is based on a combination of Self-Organizing Maps, a feature selection using Genetic Algorithm and a linear regression model. Tufekci in the next article of this issue present four predictive models that have been developed for a classification task in predicting the direction of movement in the sessional, daily, weekly, and monthly Istanbul Stock Exchange National (ISEN) using five years of data. The proposed approach is based on a multilayer perceptron (MLP), which comprises artificial neural networks (ANN), Logistic Regression (LR), and Bagging of Logistic Regression (BLR) classification techniques. The experimental results presented in the paper shows that prediction performances of the models, which are 64.13%, 63.09%, 81.54%, and 100% for the sessional, daily, weekly, and monthly datasets respectively, were found by MLP significantly better than the other classifiers compared in this study. The eighth article of this issue by Yu *et al.* is on the topic of eye fixation identification considers eye movements as spatial-temporal trajectories and present a spatial-temporal trajectory clustering algorithm for eye fixations identification. The main idea of their algorithm is based on a Density-Based Spatial Clustering Algorithm with Noise, which is commonly applied to spatial clustering data. The authors compared their classification results obtained from their algorithm with four other algorithms for eye fixations identification and showed that their proposed algorithm demonstrated an equal or better performance. Ryang *et al.* in the last article of this group, which is about high utility pattern mining, argue that in real-world databases items not only have relative importance but also are represented as non-binary values in transactions. In this article, high utility pattern mining is considered as one of the most essential issues in the pattern mining field, which recently emerged to address the limitation of frequent pattern mining. The authors propose a tree structure, which captures database information through a single-pass. They also suggest a restructuring method that contains strategies for reducing overestimated utilities. Their experimental results show that their proposed algorithm outperforms a state-of-the-art one in terms of runtime and the number of generated candidates with a similar memory usage.

The third group of articles in this issue are on graph mining. Lee and On discuss graph visualization techniques and argue that proper understanding of large public data sets requires efficient visualization techniques. In this application paper, the authors propose a novel graph visualization technique based on friend matching and hierarchical matching and in their empirical study, they apply their visualization algorithms to the voting data from Korea, Senate bills in U.S., and U.N. voting data. The authors showed the effectiveness of their approach through these experiments. Consoli *et al.* in the eleventh article of this issue argue that in certain application fields, such as linguistic and computer vision, there is an increasing need of reference data for the empirical analysis of new methods and the assessment of different algorithms. They emphasize that current evaluations are based on few real-life collections or on artificial data generators built on models that are too simplistic to cover real scenarios and to allow researchers to identify crucial limitations of their algorithms. The authors propose a flexible approach to generate high-dimensional vectors, with directional properties controlled by the distribution of their pair-wise cosine distances. Their empirical study shows that their proposed approach can create large high-dimensional data collections with desired properties in reasonable time. And finally in the last article of this issue Rahmani *et al.* discuss the topic of entity resolution in disjoint graphs and argue

that in new applications of ER, such as in the genealogical domain, the very limited linkage information among references results in a disjoint graph in which the existing content/context-based ER techniques have very limited applicability. The authors propose first to use the homophily principle for augmentation of the original input graph by connecting the potential similar references, and second, to use a Random Walk based approach to consider contextual information available for each reference in the augmented graph. They evaluate the proposed method by applying it to a large genealogical dataset and succeed to predict 420,000 reference matches with a precision of 92% and discovery of six novel and informative patterns some of which cannot be detected in the original disjoint graph.

In conclusion, with this issue of the IDA journal, which is Volume 20(2), we are celebrating the 20th anniversary of our journal. The IOS press office, 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.

In addition to our six regular issues that each contain 11–12 articles, our plan starting 2014 has been to publish one special issue per year, which is normally related to a scientific conference for which organizers have submitted an interesting proposal.

During 2016, we will 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.

<div style="text-align: right">

With our best wishes,
Dr. A. Famili
Editor-in-Chief

</div>