

Editorial

Dear Colleague:

Welcome to volume 14(3) of *Intelligent Data Analysis* – An international Journal.

This issue of the IDA journal consists of nine articles. These articles are mainly divided into 3 groups, all within the theoretical and applied research. These are data preprocessing problems, learning/classification and two application papers.

In the first article of this issue, Čehovin and Bosnić present an empirical evaluation of five feature selection methods among which is the random forest which is not extensively evaluated in the literature. The main idea behind this evaluation process is to identify how different feature selection methods can affect the accuracy of different classifiers. Their results showed that Relief and random forest enable classifiers to achieve the best accuracy. Dash and Ng in the second article of this issue discuss the importance of outlier detection in transactional data and propose a novel and efficient method for outlier detection in this domain. Their proposed method uses a unique approach to create a representative sample and subsequently determine the degree to which each transaction could be considered an outlier. This approach that does not require any parameter settings, is evaluated using a number of benchmark data sets. In the third article of this issue, Twala and Cartwright discuss the problem of building prediction models in software engineering and the importance of having quality data. They propose a method for improving software effort prediction accuracy that is based on decision tree learning and generation of ensembles using two imputation methods. The paper includes results of their benchmarking on ten data sets where the results of their proposed ensemble strategy are presented.

LaTorre *et al.* in the next group of articles in this issue propose to use reinforcement learning as a mechanism to control how different evolutionary algorithms can contribute to a typical search process. Their approach is based on three reinforcement learning methods in which they also use six large-scale continuous optimization functions. Their experimentations have proved that reinforcement learning methods can successfully learn optimal patterns for the combination of evolutionary algorithms. Peterson and Martinez, discuss the possibility of learning in different tasks, advocate that for most effective learning, it would be valuable to understand the nature of regularity in learning. They propose a distance metric that is suitable for finding similarity in learning tasks. It is demonstrated that the distance matrix generated through this process can be used for visualizing the relationship between learning tasks and searching through task space.

In the next article, which is an application one, Min *et al.* discuss fingerprint classification and emphasize the inter-class and intra-class properties of fingerprints. Their proposed approach is based on the analysis of fingerprint sub-classes with multiple decision templates which generate clusters of soft outputs of support vector machines grouped into several subclasses using self-organized maps. Their experimental results demonstrate the effectiveness of their approach. Similarly, Gago-Alonso *et al.* in the next article of this issue emphasize on the weaknesses of the frequent connected sub-graph mining algorithms and introduce a novel property for the canonical adjacency matrices to reduce the number of canonical form tests in this type of data mining. They introduce a new algorithm and perform a set of experiments on a number of real data sets where the results of evaluation of their algorithm are presented.

Zhu *et al.* in the eighth article of this issue, highlight the needs for extending the traditional snapshot-based privacy checking to be multi-dimensional by adding a time dimension. This is due to the lack of effectiveness of current techniques in privacy data mining. They present an approach to privacy in time-series data mining which is based on blind source separation techniques in statistical signal processing. Their experiments on real data show the effectiveness of their approach. And finally, Hernández-León *et al.* propose two algorithms that are suitable for mining frequent itemsets in large sparse data sets. The first of these algorithms allows analyzing time-domain data sets that do not change overtime while the second one is intended for data sets that their contents change throughout the time. Both of these algorithms introduce a novel way to use equivalence classes of itemsets by performing a breadth first search and storing the information in compressed arrays. Results of their evaluation are included in their article.

In conclusion, in Volume 14 of the IDA journal the submission rate has grown very rapidly. That is the main reason that the number of articles included in each issue have increased over the last year. For example, as seen here, this issue consists of nine articles. This demonstrates that IDA researchers believe in our journal and its value. In Volume 14, we already know that we will have three special issues related to three events that were held over the last 1–1.5 years. We look forward to receiving your feedback along with more and more quality articles in both applied and theoretical research.

With our best wishes,

Dr. A. Famili
Editor-in-Chief