

Editorial

Dear Colleague

Welcome to volume 19(1) of Intelligent Data Analysis Journal.

This issue of the IDA journal is the first issue of 2015 and consists of eleven articles, all covering various topics of the theoretical and applied research in the field of Intelligent Data Analysis.

The first article of this issue by Bouzar-Benlabiod *et al.* is on first order logic and its application on alert correlation in intrusion detection. The authors propose a new algorithm for a logic-based alerts correlation approach that integrates domain knowledge into the entire filtering process. The representation and reasoning of the domain knowledge are done using a newly introduced logic. The authors provide experimental results which are based on data from a real-world alert monitoring system. Colla *et al.* in the second article discuss the problems with the estimation of probability function of an unknown stochastic process and argue that most traditional approaches perform a preliminary choice of a parametric mathematical model that is based on the estimation and subsequent fitting of its parameters. They propose an alternative approach which does not require any assumption on the available data as it extracts the probability density function from the output of a neural network. The results of their experiments performed on a set of synthetic and industrial data sets are included in their article. In the third article of this issue Lin *et al.* discuss association rule mining and argue that association rule mining may not be suitable in all applications such as product purchasing. The authors propose utility mining as an extension to frequent itemset mining in order to consider various factors involved in this class of applications. Their proposed algorithm can handle transaction deletion to efficiently update discovered high utility itemsets. Their experiments show that their proposed approach outperforms the batch two-phase approach.

The next three articles are on graph theory and graph mining. The fourth article of this issue by Eberle and Holder is on scalable anomaly detection in graphs. They argue that in graph-based anomaly detection one can identify and analyze relationships between elements for structural oddities that could be related to problems such as fraud and network intrusion. They also argue that most traditional methods in anomaly detection have ignored relational information in the data and propose graph representation to solve this problem. Their proposed approach consists of methods for scalable graph-based anomaly detection that involves graph partitioning and windowing where they demonstrate its efficiency to detect anomalies in data represented as graph. Similarly Mejdoub *et al.* in the next article argue that bag of words describe an image as a histogram of visual words and therefore relation between words may get lost. They propose an image classification framework which may include benefits from the efficiency of the graph in modeling structural information and the classification performances given by the bag of words approach. Their approach involves selecting the most frequent subgraphs to construct a bag of subgraphs where they associate to each image a subgraph histogram that describes its visual contents. Their experiments reported in the article includes two data sets in which they demonstrate how the proposed method outperforms the bag of words and the spatial pyramid models in terms of recognition rate. Polajnar and Demsar article discuss the topic of missing or potential links and edges in network analysis. They define the problem of small network completion which deals with sets of small networks,

possibly with no recorded temporal dynamics. They present a method that extracts frequent patterns from small networks and uses them to predict missing vertices and edges in new networks. Their empirical evaluation on real and synthetic data shows that their approach performs reasonably well.

Movahedian and Khayyambashi in the seventh article of this issue discuss the topic social tagging, all the information that is produced in this domain and how this information could be used for recommendation purposes. They propose a recommender system that is based on the similarities between users and item profiles. Their proposed approach is based on generating user and item profiles by discovering frequent user-generated tag patterns. To evaluate the performance of their proposed approach the authors use real data sets from a US website. Their experimental results demonstrate that their proposed approach provides a better representation of user interests and achieves better recommendation results in terms of precision and ranking accuracy comparing to existing methods. The next article by Chen *et al.* is a comparative study of ensemble methods. The authors focus on the benefit of ensemble classifiers in a number of studies that involves credit-risk applications and particularly when cost of misclassification errors in asymmetric. The authors build ensemble models on the basis of a set of component classifiers that are derived from different subsets of instances or features by a single learning algorithm. They evaluate the performance of their algorithm in terms of expected misclassification costs. Their experimental results demonstrate the functionality of ensembles for boosting the performance of individual classifiers.

Outlier detection is the focus of the ninth article of this issue by Wan *et al.* The authors argue that local difference, which is commonly measured in outlier detection, cannot properly represent outlyingness. The authors propose a new measure that reflects the connectivity of any object to the main body of a data set. From the reports of their experiments, the proposed measure is apparently applicable to arbitrary-density and arbitrarily-shaped data. Rizoiu, Velcin and Lallich article is on the topic of semantic-enriched visual vocabulary construction. The authors discuss the challenges related to content-based image classification. They propose enriching the semantic content of the image representation by using external knowledge, expecting that it would yield higher machine learning performances. Their experiments validate their hypothesis in which they demonstrate higher classification performances than the base-line representation. And finally, Tahernezhad *et al.* in the last article of this issue discuss the topic of multi-objective evolutionary algorithms and present a new approach, which is motivated by hierarchical clustering, to enhance the population diversity among the non-dominated vectors in the solution space. The proposed idea is applied to a wide range of test problems. Their results validate the motivation on the basis of diversity and performance measures in comparison with the state of the art algorithms.

In conclusion, with this issue of the IDA journal, which is Volume 19(1), we are glad to start out 19th year of success with an excellent submission rate. Among the special issues that our colleagues have been working on and will be published soon, there is one that is from CIARP-2013 (Ibero American Congress on Pattern Recognition), targeted for publication in early 2015 and another one related to BAFI-2014 (Business Analytics in Finance and Industry), targeted for publication in early summer 2015. 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