

Editorial

Dear Colleague:

Welcome to Volume 4(1) of the journal *Intelligent Data Analysis*!

The beginning of the twenty first century and the new millennium offer an opportunity to think of what has been done related to the field of Intelligent Data Analysis in the last decade and what will be done in the future. Over the last decade, a number of techniques, tools, methodologies and architectures were developed by researchers. Some were commercialized, a number of them quite successful. Various applications were also developed, data analysis consulting firms started and the real value of “pot of gold”, that the data is referred to, was recognized. We believe the trend will continue, at least for a number of years. More techniques will be developed. Larger and larger data sets will be analyzed and a major part of the effort will be focused on developing methods to analyze on-line data, whether it is aerospace, financial, process industries or else. And more importantly, there will be a lot of emphasis on automating the entire data analysis process.

Volume 4(1) of IDA consists of five articles that represent some of the best theoretical and applied research work in the field of Intelligent Data Analysis. In the first article, Quafafou and Boussouf address the problem of feature subset selection in the context of rough sets. They extend the notion of attribute dependency and introduce the idea of strong feature subsets which they use to evaluate the performance improvements on algorithms such as C4.5. A new method for dealing with feature subset selection is proposed and compared with similar techniques. Similarly, Talavera in the second paper, introduces an unsupervised feature selection approach which is based on the assumption that in the absence of class labels, one can eliminate those features that exhibit low dependencies with the rest of features. This paper includes results of experiments using several data sets that show that the proposed approach is able to detect completely irrelevant features. The results also include removal of other attributes without significantly effecting the performance of a clustering algorithm.

The third paper by Hong and Wang addresses the problem of determining appropriate membership functions for fuzzy induction and proposes a new learning method to automatically derive membership functions and fuzzy rules from a set of training examples. The results from evaluation of this approach show that it can significantly reduce the time and effort needed to develop a rule-based expert system. Zhang, in the next paper of this issue, introduces a probabilistic model to represent probabilistic data. This model is based on the nearest neighbor algebra which is normally used to measure the quality of probabilistic data. The paper also includes a probabilistic relational algebra so that queries on probabilistic databases can be performed. The last paper by Andre-Jonsson and Badal is about indexing time series data using signatures. In real world, indexing time series data is more complex than indexing normal text. The authors propose to use signature files to index the time series and a method to index signature files to speed up the search of large time series data. The paper includes results from implementation of this approach which show good precision even from small signatures and small time windows.

And finally, this is the first issue of IDA that is published by IOS Press. Our publication plan remains the same. We will publish both an on-line and a paper version of the journal. We are also planning to

have two special issues this year. The rate of submission of papers has been overwhelming. We are extremely pleased with the level of participation and feedback that we are receiving from our colleagues. Both the Publisher and the Editorial Board would like to receive your suggestions and new ideas. Thank you.

Best wishes,

A. Famili
Editor-in-Chief