

Editorial

Dear Colleague:

Welcome to volume 11(2) of *Intelligent Data Analysis* – An international Journal.

This issue consists of 6 articles that are mainly related to data pre-processing. They cover areas such as performance measures, clustering, and data dimensionality reduction.

The first two articles are on the use of performance measures in data preprocessing stage of data analysis. Kontos et al. propose a statistical approach for dimensionality reduction that is a combination of several methods including Bootstrapping, Bayesian Inference and Markov Chain Monte Carlo. They also introduce a weighted Euclidean distance to properly classify the regions of interest in the entire multidimensional space. The article includes some results about the effectiveness of their approach on classifying experiments and in similarity searches using relatively large data sets. Keogh-Brown and Bogacka present a method to extract a time-series representation of the www log traffic. This method would be used as a measure to reflect the performance or quality of service in web applications. The paper includes results obtained for various cache data extracted from web applications and how they are well presented. The approach also allows quality service to be maintained for a web access application.

The next three articles are about clustering techniques and their application to various types of data analysis problems. Fonseca and Cardo, referring to the problem of selecting specific criteria for clustering analysis, investigate the relationship between the performance of information criteria and the type of measurement of clustering variables. Experimenting with various mixture model cluster analysis approaches, they conclude that in clustering with mixture models, there is a relationship between information criteria and clustering variables. Similarly, Garatti et al. examine the use of clustering as part of the entire data analysis process of some biological (e.g. high throughout microarray) data. They look at various steps, where clustering can play a key role not only as an unsupervised method to identify the most informative attributes (e.g. genes) but also as a data preprocessing utility. Interesting models are presented in this paper.

Clustering has also been considered for dealing with incomplete data and data imputation. Huang et al., in the last article of this group, look at the most popular missing data imputation algorithms and introduce a new algorithm that is based on clustering. While benchmarking all algorithms using a number of data sets and including the newly introduced approach, the authors propose a set of guidelines to be used in the selection of an appropriate imputation algorithm for specific applications. The results in this paper consist of evaluation of the guidelines for a process modeling case study.

In the last article of this issue, Kumar et al. discuss ways for meaningful representation of high dimensional data, focusing on nonlinear dimensional reduction approaches. They present a new scheme, called sequential-Isomap, that is based on incremental singular value decomposition. Using simulated data sets, they demonstrate the validity and significant potential of the newly introduced approach for real-time applications in autonomous systems.

And finally, this year IDA 2007, a biannual symposium on Intelligent Data Analysis, will be held in Ljubljana, Slovenia from September 6–8. Registration and submission information are available at

www.ida2007.org. As always, we will try to have one or two special issues this year, one of which will be based on the best papers from this year's symposium. We look forward to receiving your ideas in producing special issues in the future.

With our best wishes,

Dr. A. Famili
Editor-in-Chief