International Journal of Uncertainty,
Fuzziness and Knowledge-Based Systems
Vol. 15, No. 5 (2007) v–vii
© World Scientific Publishing Company

World Scientific
www.worldscientific.com

EDITORIAL

Special Issue on Fuzzy Methods in Machine Learning and Data Mining

EYKE HÜLLERMEIER
Marburg University, Germany

FRANK KLAWONN
University of Applied Sciences Braunschweig/Wolfenbuettel, Germany

ANDREAS NÜRNBERGER
University of Magdeburg, Germany

Machine learning, data mining, and related research fields have received a great deal of attention in recent years and, beyond doubt, have established themselves as core elements of intelligent and knowledge-based systems design. In these fields, a multitude of efficient algorithmic methods for inducing models from data in an automated way and for finding "interesting" patterns and relationships in large data sets have been devised. Such methods exploit the capability of computers to search huge amounts of data in a fast and effective manner. More often than not, however, the data to be analyzed is imprecise and afflicted with uncertainty. In the case of heterogeneous data sources such as text and video, the data might moreover be ambiguous and partly conflicting. Last but not least, patterns, relationships, and models of interest are usually vague and match with the data at best approximately. Thus, in order to make the learning and data mining process more robust or, say, "human-like", methods for searching and learning are needed that are tolerant toward imprecision, uncertainty, and exceptions, have approximate reasoning capabilities and are able to handle partial truth.

Here is where soft computing methods in general and fuzzy set theory in particular come into play. The capability of fuzzy sets to interface quantitative patterns with qualitative knowledge structures expressed in terms of natural language can improve the comprehensibility of extracted patterns considerably, which is a point of major importance in data mining. Fuzzy information granulation further allows for trading off accuracy against efficiency and understandability of models.

Among other things, fuzzy sets can also be useful in data reduction, in dealing with incomplete and heterogeneous data, in modeling prior knowledge, or in interactive data mining, where the mining process is under partial control of the analyst. In the context of knowledge discovery, it can thus be hoped that fuzzy set theory contributes to methods that combine the complementary searching, reasoning and pattern recognition capabilities of both computers and humans in an optimal manner.

This is one of the main motivations for devoting a special issue of this journal to the application of fuzzy methods in machine learning and data mining. The special issue grew out of the "Fuzzy Systems in Computer Science 2006" symposium that was jointly organized by the the North German Softcomputing Association (AFN), the European Society for Fuzzy Logic and Technology (EUSFLAT), and several fuzzy-oriented research groups within the Otto-von-Guericke University of Magdeburg on the occasion of the tenth anniversary of the university's Neuro-Fuzzy research group, which is headed by Professor Rudolf Kruse. Subsequent to the symposium, a small number of handpicked contributors have been invited to submit extended versions of their papers. These submissions have furthermore undergone a regular reviewing process. The papers included in this special issue, addressing classical topics like rule learning and clustering but also more recent research trends such as visualization, are the result of this two-step selection procedure.

As mentioned previously, the ability to represent patterns in a comprehensible way is often mentioned as one of the key advantages of fuzzy methods in the context of data mining. The paper by Mencar, Castellano, and Fanelli addresses this issue in more detail. The authors argue that the use of fuzzy methods can indeed improve the understandability of extracted patterns in a considerable way, provided that several prerequisites are respected, such as the definition of appropriate constraints, structural issues of the representation itself as well as user- and context-related issues. Furthermore, they provide a very critical and to some extent philosophical discussion of the terms understandability and interpretability. Therefore, this contribution is highly recommended for people who like to get a better understanding of interpretability issues in general.

Alcalá, Alcalá-Fdez, Gacto, and Herrera present a genetic algorithm to obtain a fuzzy rule-based system that is supposed to show a better trade-off between interpretability and accuracy than existing approaches. This is achieved by a post processing step that considers the number of rules and the system error as objectives of an optimization problem. The approach performs rule selection and membership function tuning in order to ensure that a rule base with the smallest number of rules that still provides a high accuracy is obtained. Besides, the authors provide a thorough motivation and critical discussion of this optimization problem on a more general level and thus provide a nice example of an application framework that can help to tackle the interpretability issues discussed in the contribution of Mencar, Castellano, and Fanelli.

In the paper by Berzal, Cubero, Sánchez, and Vila, the authors reconsider a special type of association rule, so-called *fuzzy gradual dependencies.* They propose a new definition of such dependencies and discuss related evaluation criteria. Moreover, a mining algorithm is developed which is an adaptation of existing association rule mining algorithms.

Kolodyazhniy, Klawonn, and Tschumitschew introduce a technique for dimension reduction of data sets with a larger number of attributes. The technique is based on a neural network with bottleneck hidden layer through that the data have to pass before reproducing themselves as output. The bottleneck layer provides the data representation in the lower dimensional space. The classical multilayer perceptron is replaced by a special type of neuro-fuzzy model which makes the initialisation easier and outperforms other aproaches.

Clustering techniques play an important role in exploratory data analysis. A main problem in cluster analysis is the evaluation of the clustering results. The evaluation of clustering results is also crucial to determnine the number of clusters. Especially in fuzzy cluster analysis, validity measures are very popular for this purposes. In supervised classification and classical cluster analysis, resampling is more common than validity measures. Resampling in the context of cluster analysis refers to taking subsamples of the data, cluster the subsamples and compare the results. Borgelt extends resampling techniques to fuzzy and probabilistic clustering to achieve more reliable evaluations of clustering results than with validity measured.

Rehm, Klawonn, and Kruse propose a visualisation technique for fuzzy classifiers. This graph- and scatterplot-based visualisation of fuzzy classifier and their training data enables a better understanding of the structure of the rule base, how rules interact, where conflicts between rules occur and which data are not properly covered by the fuzzy classifier. This kind if visualisation can be especially helpful for fuzzy classifiers with a larger number of rules.

Van Eck and Waltman provide an overview on the structure and evolution of publications in the computational intelligence field over the years 1996 to 2005. To this end, they apply methods from the CI field itself, combined with standard bibliometric approaches for data preprocessing and visualization. Thus, the reader gets information about the field itself and, at the same time, on a very inspiring application of these methods.

In closing this editorial, we would like to express our gratitude to the authors who contributed to a special issue that will definitely be of great interest for the readership of this journal. Moreover, we would like to recognize the many reviewers who guaranteed the high quality of the papers that have finally been accepted.