Available online at www.sciencedirect.com
ScienceDirect
International Journal of Approximate Reasoning
48 (2008) 1–3
INTERNATIONAL JOURNAL OF APPROXIMATE REASONING
www.elsevier.com/locate/ijar

Editorial

Perception based data mining and decision support systems

The massive amount of data currently being accumulated and stored has necessitated the development of algorithmic techniques to analyze large databases in order to synthesize useful information for decision analysis and forecasting. Over the past fifteen years the field of data mining has produced techniques for extracting information that have been successfully employed in a number of application areas including economics, finance, business, telecommunications, and e-business.

When humans are involved in the process, either by providing the input data or utilizing the results of the data analysis, perception plays a role in the representation and the interpretation of the information. The methodology for computing with words and perceptions developed by Zadeh provides the basis for a fuzzy semantics of words and for reasoning with fuzzy perceptions [1–3]. The formal analysis of perceptions utilizes a linguistic term set consisting of words and modifiers. The underlying interpretation of the terms is given as fuzzy membership functions and the modifiers are modeled by fuzzy set operations. Because of the ability to represent information in linguistic terms, the incorporation of fuzzy methodologies has been identified as one of the research areas with the potential to have a significant impact on the next generation of machine learning and data mining systems [4].

Fuzzy sets have been used in data mining since the mid 1990s (see [5] for an overview of early applications of fuzzy set theory in data mining). The primary motivation for fuzzy sets in this work was to avoid unnatural boundaries in partitioning attribute domains in quantitative association rules and to facilitate the interpretation of the resulting rules. In representing perceptions linguistically and computing with words, the focus is more on the term set itself rather than on the underlying membership functions. The papers in this special issue are concerned with the generation and refinement of linguistic terms from data and the assessment of data represented as linguistic terms.

This volume contains the five papers that consider different aspects of perception based data mining and decision analysis. In all of the papers, fuzzy sets provide the underlying representation of linguistic and perceptual information. The first four papers consider techniques for identifying linguistically represented associations in numeric data, temporal data, and web browsing data. The final paper focuses on the analysis of linguistically represented assessment information in a decision support system.

The first two papers describe general methods for the construction of linguistically describable associations and use logical properties to reduce the size of the resulting fuzzy rules. The paper "Mining Pure Linguistic Associations from Numerical Data" of Vilém Novák, Irina Perfilieva, Antonín Dvořák, Guoqing Chen, Qiang Wei and Peng Yan presents a method for a direct search for linguistic associations from numerical data. The support for the associations is determined using the theory of evaluating linguistic expressions and predications. Linguistic association rules discovered from data are composed of linguistic expressions like "small" and "big" and modified by fuzzy hedges "extremely", "significantly", "very", "quite roughly", etc. Real-valued data is evaluated using the corresponding linguistic expressions and standard data-mining techniques

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2007.06.002

are used to generate the linguistic associations. The logical formulation of the rules provides the potential of reducing the size of discovered linguistic associations. Two methods of such reduction are considered in the paper: the rules of logical entailment and semantic reduction rules. The method is demonstrated using the GUHA data-mining method [6].

In "Refinement of Temporal Constraints in Fuzzy Associations" Thomas Sudkamp introduces two refinement strategies for association rules with linguistic temporal constraints. Disjunctive generalization produces more general rules by merging adjacent constraints within a partition of the window of temporal relevance. Temporal specification uses linguistic hedges to reduce the duration of a constraint to better model the distribution of examples. Both types of refinement maintain the integrity of the original linguistic term set by producing rules expressible using the terms of the original rules. The acquisition of the information needed to perform the refinements is incorporated into a general algorithm for determining the number of examples and counterexamples of rules with fuzzy temporal constraints.

The fuzzy transform used in the paper "Fuzzy Transform in the Analysis of Data" of Irina Perfilieva, Vilém Novák and Antonín Dvořák transforms linguistic expressions represented by a fuzzy partition into a numeric vector representation [7]. After reviewing the properties of the transformation, the authors demonstrate the use of the transformation for detecting dependencies among attributes when the objects in the domain are described by multiple real-valued attributes. The ability to identify functional dependencies is then used for mining linguistic associations from numerical data. The ability of the technique to identify linguistical association rules is demonstrated using an example of the relationship between traffic congestion, environmental conditions, and pollution.

The paper "Linguistic Object-Oriented Web Usage Mining" of Tzung-Pei Hong, Cheng-Ming Huang and Shi-Jinn Horng proposes a fuzzy object-oriented web mining algorithm to derive fuzzy knowledge from log data on web servers. The approach uses a two-step process that analyzes browsing patterns within a single page and inter-page linkage. The first step, the intra-page mining, uses attributes of user visits to a single page to derive sets of frequently occurring attributes. The large itemsets produced from the intra-page analysis provide the underlying elements for inter-page browsing analysis. Two Apriori type algorithms are used to produce the intra-page linguistic associations and the inter-page linguistic browsing patterns simultaneously.

In "Ontology-based Intelligent Decision Support Agent for CMMI Project Monitoring and Control", Chang-Shing Lee, Mei-Hui Wang and Jui-Jen Chen propose an ontology-based intelligent decision support agent (OIDSA) applied to project monitoring and control of Capability Maturity Model Integration (CMMI). The work of the agent is based on the analysis and extraction of information from linguistic data sets. The OIDSA is composed of a natural language processing agent, a fuzzy inference agent, and a performance decision support agent. The fuzzy inference agent computes the similarity of the planned progress report and actual progress report, based on the CMMI ontology, the project personal ontology, and natural language processing results. The results provided by the OIDSA support an evaluation of the performance of project members by project manager.

As guest editors, we thank the authors for their contributions. We are also grateful to T. Denoeux, Editor-in-Chief of the International Journal of Approximate Reasoning, for his support of this special issue.

References

[1] L.A. Zadeh, From computing with numbers to computing with words from manipulation of measurments to manipulation of perceptions, IEEE Transactions on Circuits and Systems 45 (1) (1999) 105–119.

[2] L.A. Zadeh, A new direction in AI—toward a computational theory of perceptions, AI Magazine 22 (1) (2001) 73–84.

[3] I. Batyrshin, J. Kacprzyk, L. Sheremetov, L.A. Zadeh (Eds.), Perception-based Data Mining and Decision Making in Economics and Finance, Studies in Computational Intelligence, vol. 36, Springer, 2007.

[4] K.P. Adlassnig, C. Combi, A.K. Das, E.T. Keravnou, G. Pozzi, Temporal representation and reasoning in medicine: research directions and challenges, Artificial Intelligence in Medicine 38 (2006) 101–113.

[5] M. Delgado, N. Marín, D. Sánchez, M.A. Vila, Fuzzy association rules: general model and applications, IEEE Transactions on Fuzzy Systems (2003) 214–225.

[6] P. Hájek, T. Havránek (Eds.), Mechanizing Hypothesis Formation: Mathematical Foundations for a General Theory, Spring-Verlag, Berlin, 1978.
[7] I. Perfilieva, Fuzzy transforms: theory and applications, Fuzzy Sets and Systems 157 (2006) 993–1023.

Ildar Batyrshin
Research Program of Applied Mathematics and Computations,
Mexican Petroleum Institute 07730, D.F., Mexico
E-mail address: batyr@imp.mx

Thomas Sudkamp
Department of Computer Science and Engineering,
Wright State University,
Dayton, OH 45435, USA
E-mail address: thomas.sudkamp@wright.edu

Available online 1 July 2007