Fuzzy methods and Association Rule Mining

M.D. Ruiz[1], J.M. Serrano[2]

[1] Computer Engineering Department, University of Cádiz, Spain
[2]Dept. of Computer Science, University of Jaén (Spain)
mariadolores.ruiz@uca.es,jschica@ujaen.es
**

Abstract. Fuzzy Set Theory has not only enabled a generalization from crisp to fuzzy association rules, but also the development of new approaches for mining useful knowledge from quantitative data. Examples of this are fuzzy and approximate dependencies.

Keywords: Fuzzy Association Rules, Data Mining, Approximate Dependencies

1 Introduction

Association rule mining (ARM) is an exploratory technique in Data Mining field, widely used in many application domains. Association rules were first proposed in [1] to find interesting, previously unknown, correlations between items present in a set of transactions (or, in general, in a database). Although the main idea behind ARM can be seen as nothing new in areas such as Statistics, particularly in Correlation Analysis, an association rule not only allows to discover a correlation between two or more items, it also contributes with a way to measure this correlation [3], as well as informs about the direction of the implication (not to be confused with causality) between them.

Fuzzy set theory has also burst into this field, and many approaches have been developed in order to improve the mining process over quantitative data, as well as the fuzziness of data itself. In the former, some approaches consider fuzzy rules as a tool to obtain information in quantitative data [6]. Instead of dividing the quantitative attributes domain into intervals, fuzzy partitions are considered. The motivation is to allow *"soft"* transitions instead of crisp boundaries between intervals, avoiding the undesirable behavior of assessment measures in these boundaries. Moreover, fuzzy sets may represent a friendly interpretation, nearer to human perception, of linguistic terms often used in daily communication. For instance, the height of a person may be expressed by a range which it belongs, as for example, $[1.70, 1.90]$. However, depending on how these intervals are defined, the obtained results may vary significantly. Instead, the use of linguistic labels such as *"high"*, represented by a fuzzy set,

can be a good option to depict the height of a person, having at the same time a meaningful semantic to the user.

On the other hand, the nature of data can be diverse and come described in different ways: numerical, categorical, imprecise, etc. Beside this, we may also have a dataset with imprecise knowledge where ordinary crisp methods cannot be directly applied. Several proposals can be found in order to deal with this kind of data, for example, extending crisp assessment measures into fuzzy ones, as well as replacing operators in the crisp measure by their fuzzy counterparts. Definitions for fuzzy transaction and fuzzy association rule were introduced in [4]. Subsequent approaches use aggregation procedures to obtain the assessment measures by representing a fuzzy rule as a set of crisp rules [5]. In both cases, suitable assessment measures were proposed, as well as a methodology to aggregate obtained measurements in order to summarize and ease the decision making stage.

Even more, inherent from the use of fuzzy sets is their application to soften operators or relations. Prove of this are the extensions of functional dependencies by relaxing some of the elements involved in the definition of the dependence between attributes. Of special interest are the notions of approximate dependencies, where the universal quantifier is relaxed [7], or fuzzy approximate dependencies, which in addition replace the *strong* imposition of equality among attribute values by a resemblance relation [2].

To conclude, this short paper puts into manifest the role of Fuzzy Sets Theory in Association rule mining, not only as a generalization method to obtain fuzzy association rules, but as an interesting methodology to obtain information from data by smoothing some of the operators or relations present in the definition of association rule. It is also important to highlight how fuzzy association rules can be useful in the discovery of new types of rules as the mentioned approximate or fuzzy dependencies.

References

1. R. Agrawal, T. Imielinski, and A. Swami. Mining associations between sets of items in large databases. In *ACM-SIGMOD Int. Conf. on Data*, pages 207–216, 1993.
2. F. Berzal, I. Blanco, D. Sánchez, J.M. Serrano, and M.A. Vila. A definition for fuzzy approximate dependencies. *Fuzzy Sets and Systems*, 149(1):105 – 129, 2005.
3. F. Berzal, M. Delgado, D. Sánchez, and M.A. Vila. Measuring accuracy and interest of association rules: A new framework. *Intelligent Data Analysis*, 6(3):221–235, 2002.
4. M. Delgado, N. Marín, D. Sánchez, and M.A. Vila. Fuzzy association rules: General model and applications. *IEEE Transactions on Fuzzy Systems*, 11(2):214–225, 2003.
5. M. Delgado, M.D. Ruiz, D. Sánchez, and J.M. Serrano. A formal model for mining fuzzy rules using the RL representation theory. *Information Sciences*, 181(23):5194–5213, 2011.
6. C.M. Kuok, A. Fu, and M.H.Wong. Mining fuzzy association rules in databases. *SIGMOD Rec.*, 27(1):41–46, March 1998.
7. D. Sánchez, J.M. Serrano, I. Blanco, M.J. Martin-Bautista, and M.A. Vila. Using association rules to mine for strong approximate dependencies. *Data Mining and Knowledge Discovery*, 16(3):313–348, 2008.