

Theoretical principles and implementation issues of fuzzy GUHA association rules

Martin Ralbovský

Department Of Information and Knowledge Engineering
Faculty of informatics and statistics
University of Economics, Prague

There are two approaches to association rules mining: the mainstream approach deals with items and itemsets and is carried out by the known apriori algorithm. The second and much older approach is based on the ASSOC procedure of the GUHA method, which has profound logics and statistics theory background. Implementation of the latter approach works with bits of strings, which offers construction of more complex forms of association rules while maintaining fast executions. Fuzzy extensions of apriori-based association rules and their quality measures have been examined by numerous authors and remain hot scientific topic of recent years. There have been also several fuzzy extensions to the apriori algorithm.

Also for the GUHA method, there have been works dealing with special properties of fuzzy observational calculi, fuzzy generalized and statistical quantifiers. However, until now the method lacked a decent fuzzy implementation capable of showing the advantages of the GUHA method in general and ASSOC association rules mining in particular. This paper presents novel implementation of fuzzy GUHA in the Ferda system from two standpoints. First standpoint is the theoretical principles of fuzzy GUHA association rules. Theoretical models describing fuzzy association rules are explained, together with their evolution from logical calculi of crisp association rules. The aim is to cover not only fuzzy set theory aspect of the model, but also to include fuzzy logic. These new models are compared to existing models of apriori-based fuzzy association rules.

One of the characteristics of the bit string implementations of the crisp ASSOC procedure was its efficient storage and realization of bit string operations by processor instructions. In order to preserve this characteristic important implementation issues had to be dealt and decisions had to be made. The paper introduces reader to the issues of choosing proper data structures and efficient algorithms implementing fuzzy bit string operations. It also explains how different processor instruction sets and different platforms affected the decisions. Extensive testing of the new fuzzy bit string engine is included.