ABSTRACT

Data Mining is the process of extracting interesting and previously unknown patterns and correlations from data stored in huge data repositories. Association Rule Mining, a descriptive mining technique of Data Mining is the process of discovering items, which tend to occur together in transactions. Association rules are interesting correlations among attributes in a database.

A key issue that needs more concern when using Association Rule Mining is the soundness of the rules external to the dataset from which they are generated. Rules are usually the derivative of the patterns in a specific data set. When a different phenomenon occurs, the changes in the set of rules obtained from the new setting could be significant. This thesis provides a Group based Mining of Association Rules (G-MAR) model by paying special attention as how the variation between two different settings affects the changes of the rules, based on the notion of fine partitioned groups termed as factions. Using this G-MAR model, a simple technique called Coalescent Dataset, is proposed to get a fine approximation of the set of rules for a new situation.

The approach proposed, works independently of the core mining process and can be easily implemented with all variations of the rule mining techniques.
Also, a fuzzy clustering based Association Rule Mining system is proposed in targeting customers to improve sales which improvises the G-MAR model by predicting sales based on customer needs and functional features. For a potential customer arriving the store, which customer group one should belong to according to customer needs, what are the preferred functional features or products that the customer focuses on and what kind of offers will satisfy the customer etc., finds to be the key factor in targeting customers to improve sales. Generally, a transactional database is created to record all the products purchased by the customer. To focus on the market segment that each customer falls into, the transaction database can be grouped into different clusters based on the customer needs.

From the above outline, it can be seen that the main point of the proposed model is to formulate and sample the Coalescent Dataset, which is independent of any core mining algorithm. Existing algorithms as how to discover association rules proposed by various researchers can be used for association rule generation. For different cases, different algorithms can be used.