Novel Algorithm for Mining High Utility

S. Shanker

Abstract - Literature review revealed that one of the important issues in data mining is the interestingness problem . Typically in a data mining process , the number of discovered can easily exceed the capabilities of a human user to identify interesting results . To address this problem utility measures have been used o reduce the patterns prior to presenting them to the user . The fundamental idea behind mining frequent item sets is that only item sets with high frequency are of interest to users . However , the practical usefulness of frequent item sets is limited by the significance of the discovered item sets . A frequent item sets only reflects the statistical correlation between items , and it does not reflect the semantic significance of the items. In this paper we are using a utility based item set mining approach to overcome this limitation . Utility based data mining is a new research area interested in all types of utility factors in data mining processes and targeted at in comporting utility considerations in data mining tasks . high utility item set mining is a research area of utility based data mining , aimed at finding item sets that contribute high utility . This paper presents a novel algorithm Fast Utility Mining (FUM) which finds all high utility item sets within the given utility constraint threshold . It is faster and simpler than the original mining algorithm. The experimental evaluation on artificial datasets show that our algorithm executes faster than mining algorithm, When more item consider their most valuable sets are identified as high utility item sets and when the number of distinct items in the database increases. The proposed FUM algorithm scales well as the size of the transaction database increases with regard to the number of distinct items

available .