

Abstract

The evolution of the data mining field started with the discovery of relevant and significant data generated from organizations. The data may contain specific associations among the various items that can be used for the decision-making process. This type of association is called association rule mining. It is one of the subset fields in data mining that identifies the correlations and the associations. Association rule mining is applied in diverse fields and applications such as social networks, stock markets, market basket analysis, e-commerce, etc. The data in such applications are increasing at a rampant growth. Processing and analyzing an enormous quantity of data may require I/O operations, computational cost, and time. To adhere to this problem, an efficient algorithm is needed for mining such a dataset keeping in view the three factors. The first contribution of the thesis is the proposed representation of the itemsets. It is observed that the representation of the itemsets greatly influences the way the mining process works. If the representation of itemsets is condensed then the memory consumption, I/O operation, and computational cost are reduced since the search space is minimized. The proposed itemset representation is incorporated in the association rule mining algorithm, Apriori Algorithm. To increase the efficiency of the Apriori algorithm, the searching technique is also incorporated into the Apriori algorithm that uses the proposed itemset representation. The second contribution is graph-based itemset mining. The dataset is represented using Graph $G=V, E$ where V denotes the vertex and E is the edge. The vertex V consists of the set of vertices $v_1, v_2, v_3, \dots, v_n$ and Edge E consist of edges $e_1, e_2, e_3, \dots, e_n$. Each itemset denotes a vertex that is represented using the proposed itemset representation. The evaluation of the graph-based algorithm is compared with other algorithms in terms of memory and time consumption. The third contribution is the generation of the rules from the already mined frequent itemsets. These frequent itemsets have been generated using the Apriori The algorithm and the representation of these itemsets are represented using the proposed itemset representation. The performance of the different rule generation iii algorithms are compared with time and memory

requirements.

Keywords: Association rule mining, Apriori algorithm, itemset representation, graph-based data algorithm, rule generation algorithm, memory consumption, time consumption, condensed representation.