B

Efficient Ordering Policy for Imperfect Quality Items Using Association Rule Mining

Mandeep Mittal
Department of Computer Science Engineering, Amity School of Engineering and Technology, India

Sarla Pareek
Apaji Institute of Mathematics and Applied Computer Technology, Banasthali University, India

Reshu Agarwal
Apaji Institute of Mathematics and Applied Computer Technology, Banasthali University, India

INTRODUCTION

Data mining is the process of finding interesting patterns from huge databases. It is the main part of the knowledge discovery in database (KDD) process. The KDD process may consist of following steps: data cleaning, data transformation, pattern searching, finding presentation, finding interpretation and finding evaluation. Emergence of data mining and knowledge discovery results in fast development of information and database technologies.

Further, (Mannila & Raiha, 1987) have done work in the database community on inferring functional dependencies, and efficient inference algorithms. However, association rule mining received a great attention (Agrawal et al., 1993). It is used to find fascinating rules from large collections of data which express an association or relation between items or sets of items. The most important algorithm for generating association rules is apriori-algorithm. This algorithm is designed to operate on databases containing transactions. It works to find the items for frequent item-set based on minimum support and generate association rules based on threshold confidence. Figure 1, explains the flowchart of apriori-algorithm. Now, we will explain the meaning of the terms minimum support and confidence. Given the item-set I containing items $\{i_1, i_2, i_3 \ldots, i_n\}$, support for item i_1 is defined as the frequency of its occurrence in total number of transactions. It is given by formula:

$$\text{Support}\left(i_1\right) = \frac{\text{Frequency of } i_1}{\text{Total number of Transactions}}$$

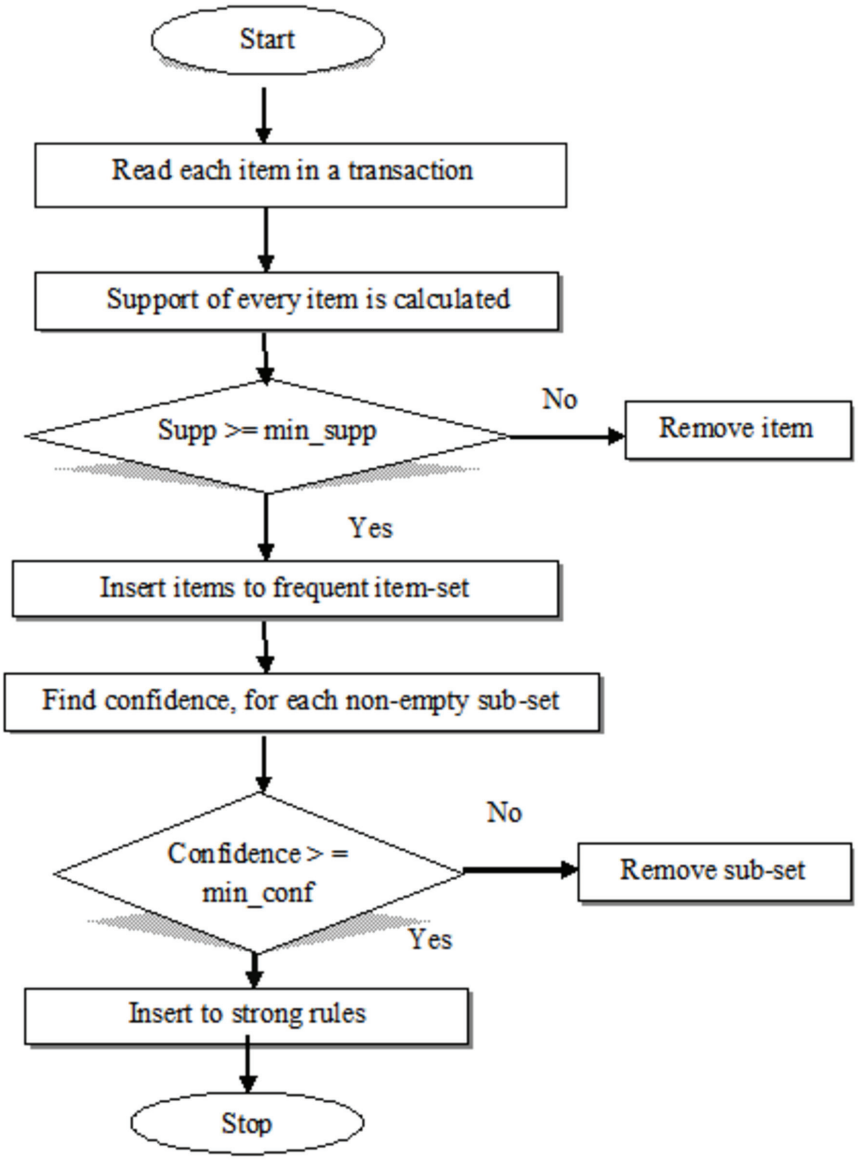
The relationship between items is expressed in terms of confidence. Confidence is defined as conditional probability as confidence $(i_1 \rightarrow i_2)$ refers to frequency of purchasing i_2 when i_1 is purchased. It is defined by the formula:

$$\text{Confidence}\left(i_1 \rightarrow i_2\right) = \frac{\text{Support of } i_1 \bigcup i_2}{\text{Support of } i_1}$$

Further, many variants of mining association rules are studied to explore more mining capabilities, such as incremental updating (Lee et al., 2001), mining of generalized and multi-level rules (Srikant & Agrawal, 1995), mining of multi-dimensional rules (Ng & Han, 1994) and temporal association rule discovery (Ale & Rossi, 2000). Temporal association rule mining leads to discovery of association rules that may hold during some time intervals but not during others as explained by (Li et al., 2003). The general temporal association rule is termed to be frequent within its maximum common exhibition period (MCEP) if and only if its support is not smaller than the minimum support (min_supp), and its confidence is not smaller

DOI: 10.4018/978-1-4666-5888-2.ch074

Figure 1. Flow chart of Apriori-algorithm



than the minimum confidence (min_conf) as defined by (Lee et al., 2003). Generally the exhibition period of an item-set is expressed in terms of maximal common exhibition period. MCEP (P) denotes the MCEP value of P. The MCEP value of an item-set P is the shortest MCEP among the items in item-set P. For example, value of MCEP (P) = (1, 3) in Table 1.

Consequently, the general temporal association rule mining algorithm can be explained in three steps:

1. Find every frequent maximal temporal item-set (TIs) with their support values.
2. Find the support values of every corresponding temporal sub-item-set (SIs) of frequent TIs.