Title:
Providing a method for incremental mining of frequent patterns using batch frequent pattern trees

Abstract:
There are a lot of data in the information industry which should be turned into useful information so that it become accessible; therefore, the analysis of these large amounts of data and mining of their useful information is essential which this process is called data mining. Here, frequent patterns can be called as one of the most important research topics in data mining. Frequent patterns are patterns which continuously occur in the transaction data. In many applications, the database must be updated due to the addition of new data. Since the incremental mining of frequent pattern is the best method for updating the database, a lot of researches have been performed in relation to offer of effective algorithms based on the pattern of growth for incremental mining of frequent patterns. These algorithms divided into two categories of incremental algorithms and batch incremental algorithms. BIT and GM algorithms can be cited for batch incremental mining algorithms. Batch incremental algorithm batch of BIT requires two scans of the database and also it needs nodes of tree is sorted by frequency; therefore, it requires rescanning the database and calculating new frequency for keeping this order when it merges two trees. Also, there may be lost frequent patterns in this batch incremental algorithm. Therefore, GM algorithm is presented to solve two problems of BIT algorithm, but this algorithm is uncompressed and it causes high memory and time consumption. In order to solve the above mentioned problems, offering a method for incremental mining of frequent patterns using batch incremental tree is introduced in this study. In this way, when we enter the main and local database and producing main FP-tree, a list is used to hold data and non-frequent items are not removed too. Using the proposed method, only one scan of the database is required while compactness of the database is maintained and there will not be lost frequent pattern too. The results of performed tests on different sets of dense and non-dense data confirm that the proposed method is more efficient than BIT algorithm and GM, so that 70% of time is improved and 100% of frequent pattern mining is performed.

Keywords:
Frequent patterns, Incremental mining, Batch incremental mining, FP-tree