

An Efficient Count Based Transaction Reduction Approach For Mining Frequent Patterns

V.Vijayalakshmi^a, Dr.A.Pethalakshmi^b

^aResearch Scholar, M.S University, Tirunelveli, T.N

^bHead & Associate Professor of Comp.Science , M.V.M Government Arts College, Dindigul, T.N

Abstract

Apriori algorithm is a classical algorithm of association rule mining and widely used for generating frequent item sets. This classical algorithm is inefficient due to so many scans of database. And if the database is large, it will take too much time to scan the database. To overcome these limitations, researchers have made a lot of improvements to the Apriori. This paper analyses the classical algorithm as well as some disadvantages of the improved Apriori and also proposed two new transaction reduction techniques for mining frequent patterns in large databases. In this approach, the whole database is scanned only once and the data is compressed in the form of a Bit Array Matrix. The frequent patterns are then mined directly from this Matrix. It also adopts a new count-based transaction reduction and support count method for candidates. Appropriate operations are designed and performed on matrices to achieve efficiency. All the algorithms are executed in 5% to 25% support level and the results are compared. Efficiency is proved through performance analysis.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of organizing committee of the Graph Algorithms, High Performance Implementations and Applications (ICGHIA2014)

Keywords: Association Rule, Frequent Item Set, Support Count, Apriori, Transaction Reduction Technique

1. Introduction

With the progress of the technology of information and the need for extracting useful information of business people from dataset, data mining and its techniques is appeared to achieve the above goal. Data mining is the essential process of discovering hidden and interesting patterns from massive amount of data where data is stored in data warehouse, OLAP (on line analytical process), databases and other repositories of information.

* Corresponding author. Tel.: 9788170037 ; fax: +0-000-000-0000

E-mail address: v.viji08@yahoo.in

This data may reach to more than terabytes. Data mining is also called (KDD) knowledge discovery in databases, and it includes an integration of techniques from many disciplines such as statistics, neural networks, database technology, machine learning and information retrieval, etc [6].

Interesting patterns are extracted at reasonable time by KDD's techniques. KDD process has several steps, which are performed to extract patterns to user, such as data cleaning, data selection, data transformation, data pre-processing, data mining and pattern evaluation.

The architecture of data mining system has the following main components [6]: data warehouse, database or other repositories of information, a server that fetches the relevant data from repositories based on the user's request, knowledge base is used as guide of search according to defined constraint, data mining engine include set of essential modules, such as characterization, classification, clustering, association, regression and analysis of evolution. Pattern evaluation module that interacts with the modules of data mining to strive towards interested patterns. Finally, graphical user interfaces from through it the user can communicate with the data mining system and allow the user to interact.

Association Mining is one of the most important data mining's functionalities and it is the most popular technique has been studied by researchers. Extracting association rules is the core of data mining. It is mining for association rules in database of sales transactions between items which is important field of the research in dataset. The benefits of these rules are detecting unknown relationships, producing results which can perform basis for decision making and prediction.

In this regard, the first algorithm Apriori was proposed by Agarwal and Srikanth in the year 1994 to mine the frequent item set. Time constraint and efficiency of algorithms leads to lot of research in the area of algorithm to build efficient algorithm which takes less time and few number of database scans to mine frequent item set. The normally followed scheme for mining association rules consists of two stages [1]:

1. the discovery of frequent itemsets
2. the generation of association rules.

As the second step is rather straightforward and as the first step dominates the processing time, we explicitly focus the paper on the first step: the discovery of frequent itemsets.

The remaining part of this paper is organized as follows: Section 2 contains related works. In section 3, elaborates the proposed transaction reduction techniques. Section 4 discusses about the performance analysis of proposed algorithm compared with Apriori algorithm. Section 5 contains conclusion.

2. Related Works

Agrawal proposed an algorithm, called AIS algorithm [1], for generating frequent itemsets. In the AIS algorithm, frequent itemsets are generated through iterations on scanning the database. The iteration terminates when no new frequent item-set is derived. After reading a transaction in the k^{th} iteration, the AIS algorithm computes the candidate k – itemsets by first deriving a set of $(k-1)$ –itemsets which contains itemsets that are both in the frequent $(k-1)$ –itemsets and in the transaction. However, Apriori algorithm has the limitation of producing a large number of candidate itemsets and scanning the database too many times.

Many researchers have given different approaches for improving the performance of Apriori algorithm.

Ayres.J [2] introduced an effective pruning mechanism called depth first strategy to mine the sequential pattern in large database,. This strategy defines the database in vertical bitmap format with effective support counting. For each item in the dataset, a vertical bitmap is constructed by which each data set transaction is represented as a bit. The value for items is set based on the item present in the transaction. The efficient support counting and candidate generation is obtained by partitioning the bitmap.

Changsheng Zhang and Jing Ruan [3] have worked on the improvement of Apriori algorithm by applying dataset reduction method and by reducing the I/O spending. Changsheng and Jing Ruan have applied the modified algorithm for instituting cross selling strategies of the retail industry and to improve the sales performance.

One of the novel algorithm presented by Chen.J [4] is a BISC (Binary Item set Support Counting), which is

Download English Version:

<https://daneshyari.com/en/article/490031>

Download Persian Version:

<https://daneshyari.com/article/490031>

[Daneshyari.com](https://daneshyari.com)