Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017)
IEEE Xplore Compliant - Part Number: CFP17L34-ART, ISBN: 978-1-5386-4031-9

2017-18
IEEE

A Survey on Frequent Itemset Mining Algorithms and their Impact

V. B. More
Assistant Professor, Department of Computer Engineering, MET's Institute of Engineering, Bhujbal Knowledge City, Nashik, MS, India

Dr. M. U. Kharat
Professor and Head, Department of Computer Engineering, MET's Institute of Engineering, Bhujbal Knowledge City, Nashik, MS, India

Abstract— **Frequent itemset mining is one of the important data mining techniques used to extract useful and meaningful itemsets or patterns from large data sets. Since this technique helps many entrepreneurs who manage large transaction data use data mining techniques to explore important information among large transactions to take decisions for increase in sale. This is also an important area of research. Various techniques have been proposed to improve the performance of frequent itemset mining algorithms by various researchers. In this paper, we highlight basic concepts of frequent itemset mining algorithms presented by various authors. Here we concentrate on recent advancements in new techniques other than Apriori. Each technique is discussed briefly. This review is based on the intension of how researcher can select particular line of research more specific.**

Keywords*— *Data mining technique, frequent itemset, pattern, transaction.

I. INTRODUCTION

A. Apriori Algorithm

The Apriori algorithm was proposed by Agrawal and Srikant in 1994. The Apriori algorithm for frequent itemset[1] mining is regarded as one of the most influential and successful techniques in data mining over transactional databases. Apriori uses a bottom up approach, where frequent subsets are extended one item at a time. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by *Apriori* can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis. It forms the basis of several data mining tasks such as mining association rules.

B. Frequent Itemset Mining (FIM)

Frequent itemsets are patterns that appear in a large data set most frequently. A transaction supports a pattern if it contains the pattern. The frequency of a pattern is the proportion of transactions in the data that support it. The goal in FIM is to discover and report the patterns that occur most frequently in the data.

If itemsets P and Q are both frequent and mostly occur together, i.e. $\sup(P \cup Q) \leq \sup(P) \times \sup(Q)$, then P and Q are frequent items in the dataset.

Apriori's efficiency can be improved by various methods. Some of them are:

- *Hash-based itemset counting:* A k-itemset whose corresponding hashing bucket count is below the threshold cannot be frequent;
- *Transaction reduction:* A transaction that does not possess any frequent k-itemset is inadequate in subsequent scans;
- *Partitioning:* Any itemset that is potentially frequent in DB must be frequent in at least one of the partitions of DB;
- *Sampling:* mining on a subset of given data, lower support threshold + a method to determine the completeness;
- *Dynamic itemset counting:* add new candidate itemsets if all of their subsets are estimated as frequent.

II. LITERATURE SURVEY

Jaideep Vaidya and Chris Clifton [1], discussed about privacy considerations often constrain data mining projects. This paper addresses the problem of association rule mining where transactions are distributed across sources. Each site holds some attributes of each transaction, and the sites wish to collaborate to identify globally valid association rules. However, the sites must not reveal individual transaction data. For this, they presented a two-party algorithm for efficiently discovering frequent itemsets with minimum support levels, without either site revealing individual transaction values.

Shariq Bashir, Muhammad Shuaib, Yasir Sultan, and Dr. A. Rauf Baig [2], proposed an improved frequent itemset mining algorithms performance by using efficient implementation technique. Mining frequent itemset in transactional datasets is considered to be a very challenging research oriented task in data mining due to its large applicability in real world problems. Due to the NP-Complete nature of problem, the efficiency of frequent itemset mining highly depends on the efficiency of algorithm implementation. They proposed number of different implementation techniques (other than itemset mining), that can improve the running time of any frequent itemset algorithm implementation. To check the efficiency of these implementation techniques they integrated those techniques into the original implementations of current best itemset mining implementations on those days.

[1] Itemset and patterns are interchangeable.