



A Systematic Literature Review of Frequent Pattern Mining Techniques

Rajendra Chouhan

M. Tech Scholar, Dept. of Computer Science and Engineering, LNCT, Indore, Madhya Pradesh, India

Er. Khushboo Sawant

Assistant Professor, Dept. of Computer Science and Engineering, LNCT, Indore, Madhya Pradesh, India

Dr. Harish Patidar

HOD, Associate Professor, Dept. of Computer Science and Engineering, LNCT, Indore, Madhya Pradesh, India

ABSTRACT

Mining of frequent items from a voluminous storage of data is the most favorite topic over the years. Frequent pattern mining has a wide range of real world applications; market basket analysis is one of them. In this paper, we present an overview of modern frequent pattern mining techniques using data mining algorithms. Frequent pattern mining in data mining takes a lot of data base scans. Therefore it is a computationally expensive task. So still there is a need to update and enhance the existing frequent pattern mining techniques so that we can get the more efficient methods for the same task. In this paper, a study of all the modern and most popular frequent pattern mining technique is also performed.

Keywords: Data Mining, Frequent Item Mining, Support, Confidence, Market Basket Analysis, Parallel Execution

I. INTRODUCTION

The use of data mining [1,2] is placed in various decisions making task, using the analysis of the different properties and similarity in the different properties can help to make decisions for the different applications. Among them the prediction is one of the most essential applications of the data mining and machine learning. This work is dedicated to investigate about the decision making task using the data mining algorithms. Data mining is associated with extraction of non trivial data from a large and voluminous data set. Figure 1 shows the general working of data mining.

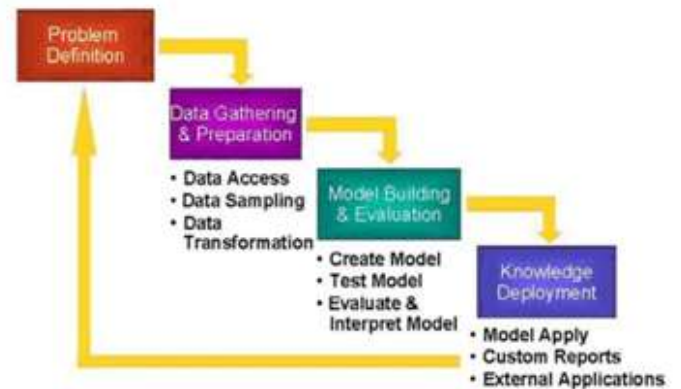


Figure 1: Data Mining

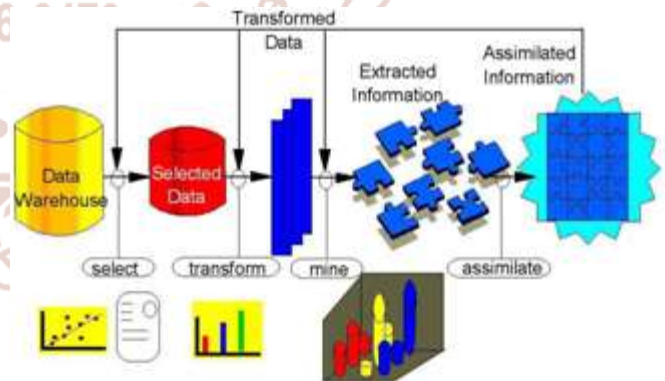


Figure 2: key steps in data mining

Figure 2 shows, key steps performed during the process of data mining. The data mining is a process of analysis of the data and extraction of the essential patterns from the data. These patterns are used with the different applications for making decision making and prediction related task. The decision making and prediction is performed on the basis of the learning of

algorithms. The data mining algorithms supports both kinds of learning supervised and unsupervised. In unsupervised learning only the data is used for performing the learning and in supervised technique the data and the class labels both are required to perform the accurate training. In supervised learning the accuracy [3,4] is maintained by creating the feedbacks form the class labels and enhance the classification performance by reducing the error factors from the learning model.

Frequent pattern mining is the concept which is used to extract most frequently occurring item from a data set. It is associated with a terminology known as support. The support of an item is obtained by first calculating the number of transactions in data set containing that item and then dividing this by total number of transactions. Also if the support of an item is more than a user defined quantity (minimum support threshold) then that item is known to be frequent. Otherwise, item is known as infrequent item.

II. LITERATURE SURVEY

In this step, all patterns which have support no less than the user-specified minsup value are mined as frequent patterns. Frequent pattern mining is used to prune the search space and limit the number of association rules being generated. Many algorithms discussed in the literature use “single minsup framework” to discover the complete set of frequent patterns. The reason for the popular usage of “single minsup framework” is that frequent patterns discovered with this framework satisfy downward closure property, i.e., all non-empty subsets of a frequent pattern must also be frequent.

The downward closure property makes association rule mining practical in real-world applications [5],[6]. The two popular algorithms to discover frequent patterns are: Apriori and Frequent Pattern-growth (FP-growth) [7] algorithms. The Apriori algorithm employs breadth-first search (or candidate-generate-and-test) technique to discover the complete set of frequent patterns. The FP-growth algorithm employs depth-first search (or pattern-growth) technique to discover the complete set of frequent patterns. It has been shown in the literature that FP-growth algorithm is relatively efficient than the Apriori algorithm [7].

In many cases it is useful to use low minimum support thresholds. But, unfortunately, the number of extracted patterns grows exponentially as we decrease. It thus happens that the collection of discovered patterns is so large to require an additional mining process that should filter the really interesting patterns. The Apriori property [8] does not provide an effective pruning of candidates: every subset of a candidate is likely to be frequent. In conclusion, the complexity of the mining task becomes rapidly intractable by using conventional algorithms.

The first of these kind of algorithms was Pascal [9,10,11,12], and now any FIM algorithm uses a similar expedient. More importantly, association rules extracted from closed itemsets have been proved to be more meaningful for analysts, because many redundancies are discarded.

Guo et al [13] proposed a vertical variant of the a priori algorithm. In apriori, several scans of the data base are required. The author proposed a version of the improved a priori algorithm. In this version lesser scans of the data base are required.

Recently some authors [14][15][16] have developed some frequent pattern mining algorithms. These algorithms (BSO-ARM, PGARM, PeARM) take less number of data base scans to find frequent patterns. But there is a short fall in all such algorithms, they only find a part of frequent items and donot find all the possible frequent patterns from a data set. Author in [17] introduced a new data structure to storing patterns, which ultimately resulted in improved efficiency of a priori algorithm. Also [18] proposed an efficient technique to mine fuzzy periodic association rules. This proposed technique scans the database almost two times.

III. TECHNICAL REVIEW

The brief history of the research algorithms of the Frequent Pattern Mining in Horizontal and Vertical Data Layouts has been discussed in this section. The following table provides the key information on the literature survey.

Table -1 Comparative Analysis of the different existing algorithms

S.No.	Algorithms	Data Structure/Layout	Search Direction
1	APRIORI	Hash tree (Horizontal)	BFS
2	VIPER	Vertical	BFS
3	ECLAT	Vertical	DFS
4	FP-Growth	Prefix tree (Horizontal)	DFS
5	MAFIA	Vertical	BFS
6	PP-Mine	Prefix tree (Horizontal)	DFS
7	COFI	Prefix tree (Horizontal)	DFS
8	DIFFSET	Vertical	BFS
9	TM	Vertical	DFS
10	TFP	Prefix tree (Horizontal)	Hybrid
11	SSR	Horizontal	DFS

IV. CONCLUSIONS

The basic objective of frequent item set mining cum association rule mining is to find strong correlation among the items in the transaction data set. All the researchers are aware of the fact that they are required to deal with the voluminous data while performing mining on the data. So the goal is to devise such algorithms which are time and memory efficient. This paper elaborates the frequent item set mining and the work done by various authors to perform mining on the transaction data set.

REFERENCES

1. Y. Bastide, R. Taouil, N. Pasquier, G. Stumme, and L. Lakhal. Mining frequent patterns with counting inference. SIGKDD Explorations Newsletter, 2(2):66–75, December 2000.
2. Y. Chi, Y. Yang, Y. Xia, and R. R. Muntz. CMTreMiner: Mining both closed and maximal frequent subtrees. In PAKDD '04: Proceeding of the eighth Pacific Asia Conference on Knowledge Discovery and Data Mining, pages 63–73, May 2004.
3. R. Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. In I. Rival, editor, Ordered sets, pages 445–470, Dordrecht–Boston, 1982. Reidel.
4. X. Yan and J. Han. Closegraph: mining closed frequent graph patterns. In KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 286–295, August 2003
5. Chen Chun-Hao , Hong Tzung-Pei and Tseng Vincent S. , “Genetic- fuzzy mining with multiple minimum supports based on fuzzy clustering”, 2319-2333, Springer 2011.

6. Weimin Ouyang , Qinhuang Huang ,“Mining direct and indirect fuzzy association rules with multiple minimum supports in large transaction databases”, 947 – 951, IEEE 2011.
7. Han J., Pei J., Yin Y., and Mao R., "Mining frequent patterns without candidate generation: A frequent-pattern tree approach", *Data Min. Knowledge. Discovery.* 8(1):53–87, 2004.
8. X. Yan, J. Han, and R. Afshar. Clospan: Mining closed sequential patterns in large datasets. In *SDM'03: Proceedings of the third SIAM International Conference on Data Mining*, pages 166–177, May 2003.
9. N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Discovering frequent closed itemsets for association rules. In *ICDT '99: Proceeding of the 7th International Conference on Database Theory*, pages 398–416, January 1999.
10. J. Pei, J. Han, and R. Mao. Closet: An efficient algorithm for mining frequent closed itemsets. In *DMKD '00: ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, pages 21–30, May 2000.
11. M. J. Zaki and C.-J. Hsiao. Charm: An efficient algorithm for closed itemset mining. In *SDM '02: Proceedings of the second SIAM International Conference on Data Mining*, April 2002.
12. K. Gouda and M. J. Zaki. Genmax: An efficient algorithm for mining maximal frequent itemsets. *Data Mining and Knowledge Discovery*, 11(3):223–242, 2005.
13. Guo Yi-ming and Wang Zhi-jun, “A vertical format algorithm for mining frequent item sets,” *2nd International Conference on Advanced Computer Control (ICACC)*, Vol. 4, pp. 11 – 13, 2010.
14. Djenouri, Y., Drias, H., Habbas, Z.: Bees swarm optimisation using multiple strategies for association rule mining. *Int. J. Bio-Inspired Comput.* 6(4), 239– 249 (2014)
15. Gheraibia, Y., Moussaoui, A., Djenouri, Y., Kabir, S., Yin, P.Y.: Penguins search optimisation algorithm for association rules mining. *CIT J. Comput. Inf. Technol.* 24(2), 165–179 (2016)
16. Luna, J.M., Pechenizkiy, M., Ventura, S.: Mining exceptional relationships with grammar-guided genetic programming. *Knowl. Inf. Syst.* 47(3), 571–594 (2016)
17. J. Adhikari and P. Rao, Identifying calendar-based periodic patterns, in *Emerging Paradigms in Machine Learning*. New York, NY, USA: Springer, 2013, pp. 329—357.
18. W.-J. Lee, J.-Y. Jiang, and S.-J. Lee, Mining fuzzy periodic association rules, *Data Knowl. Eng.*, vol. 65, no. 3, pp. 442—462, 2008