MANTECH Publications

Journal of Data Mining and Knowledge Engineering
Volume 7, Issue 2, July-December 2022

A Systematic Review of the Literature on Techniques for Mining Frequent Patterns

Dr. Poorvi Saxsena[1], Ravindra Pathak[2]

Assistant Professor[1], Student[2]

Department of Computer Science and Engineering,

Vasantidevi Patil Institute of Engineering and Technology

Corresponding Author's Email: -*ravindrapathak2@yahoo.com[1]*

Abstract

Over the years, the most popular issue has been the mining of common objects from large amounts of data storage. Market basket analysis is one of the many real-world applications of frequent pattern mining. We offer an overview of recent frequent pattern mining strategies employing data mining algorithms in this study. Frequent pattern mining in data mining necessitates a large number of database scans. As a result, it is a computationally costly process. As a result, there is still a need to update and improve existing frequent pattern mining approaches in order to obtain more efficient solutions for the same activity. A review of all the contemporary and most popular frequent pattern mining techniques is also undertaken in this work.

Keywords*: Data Mining, Frequent Item Mining, Support, Confidence, Market BasketAnalysis, Parallel Execution*

INTRODUCTION

The usage of data mining [1,2] is put to use in many decision-making tasks, where the study of the different qualities and similarity in the different properties may assist in creating judgments for the various applications. Prediction is one of the most important applications of data mining and machine learning. This study is devoted to researching the decision-making task utilizing the Figure 2 depicts the essential processes taken throughout the data mining process. Data mining is the process of analyzing data and extracting the necessary data mining algorithms.

Page 42- 48 © MANTECH PUBLICATIONS 2022. All Rights Reserved

The extraction of non-trivial data from a vast and voluminous data source is connected with data mining. Figure 1 depicts the general operation of data mining.

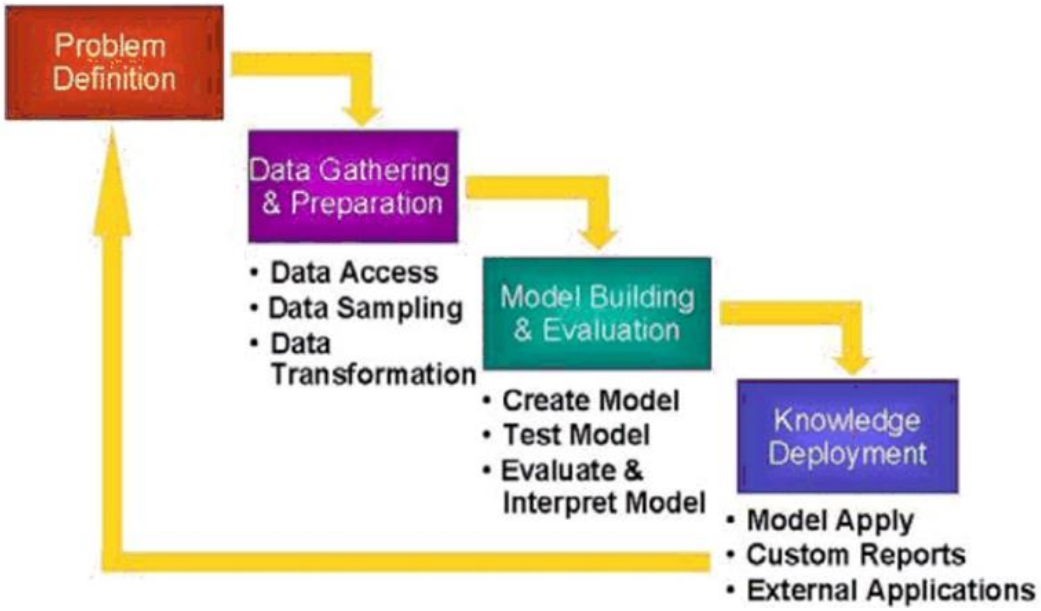


Figure 1: Data Mining

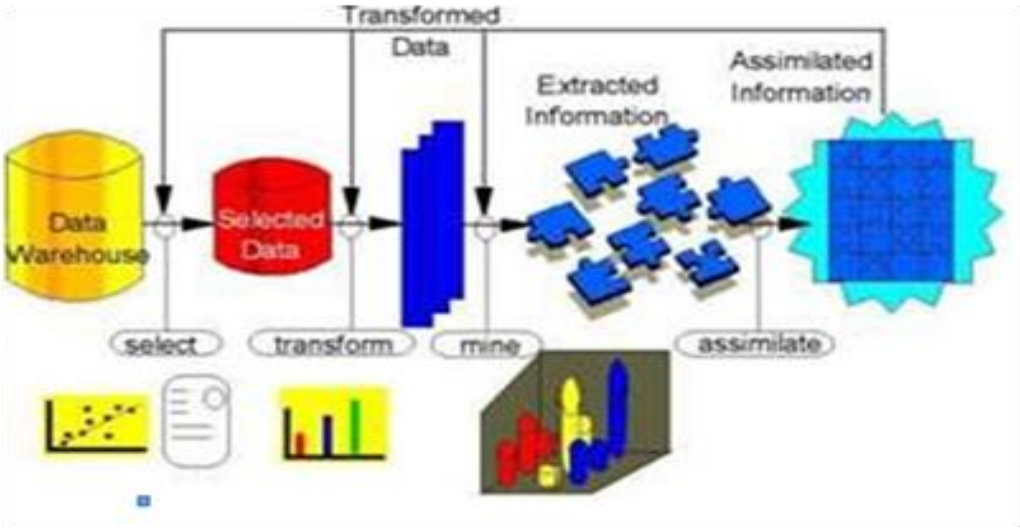


Figure 2: key steps in data mining

These patterns are utilised by many applications to perform decision-making and prediction tasks. The learning of algorithms is used to make decisions and anticipate outcomes. The data mining techniques enable both supervised and unsupervised learning. In unsupervised learning, just data is used to do the learning, but in supervised learning, both data and class labels are required to

Page 42- 48 © MANTECH PUBLICATIONS 2022. All Rights Reserved

perform accurate training. The accuracy [3,4] of supervised learning is maintained by producing feedback from the class labels and improving classification performance by minimizing the error components of the learning model.

The notion of frequent pattern mining is used to extract the most frequently occurring item from a data set. It is related to the concept of support. The number of transactions in the data set containing that item is calculated and then divided by the total number of transactions to determine an item's support. Also, if an item's support exceeds a user-defined number (support threshold), that item is said to be frequent. Otherwise, the object is referred to as an "uncommon item."

LITERATURE SURVEY

All patterns with support greater than the user-specified minsup value are mined as common patterns in this stage. Frequent pattern mining is used to reduce the search space and restrict the number of generated association rules. To uncover the whole collection of common patterns, several algorithms reported in the literature employ a "single minsup framework." The term "single minsup framework" is famous because frequent patterns discovered using it meet the "downward closure property," which states that any non-empty subsets of a frequent pattern must likewise be frequent.

Because of the downward closure characteristic, association rule mining is applicable in real-world applications [5], [6]. Apriori and Frequent Pattern-Growth (FP-growth) [7] are two common methods for discovering frequent patterns. To uncover the whole collection of common patterns, the Apriori method applies a breadth-first search (or candidate-generate-and-test) technique. To uncover the whole collection of common patterns, the FP-growth method applies the depth-first search (or pattern-growth) technique. The FP-growth method has been demonstrated in the literature to be more efficient than the Apriori approach [7].

In many circumstances, using low minimum support levels is beneficial. Unfortunately, as we go smaller, the number of retrieved patterns rises exponentially. As a result, the collection of identified patterns is so huge that an extra mining procedure is required to filter out the truly interesting patterns. The Apriori property [8] does not allow for effective candidate trimming since every subset of a candidate is likely to be common. In conclusion, using traditional algorithms

Page 42- 48 © MANTECH PUBLICATIONS 2022. All Rights Reserved

quickly renders the mining operation intractable.

Pascal was the first of these algorithms [9,10,11,12], and any FIM algorithm currently employs a similar method. More significantly, because many redundancies are eliminated, association rules generated from closed itemsets have been shown to be more useful for analysts.

A vertical form of the a priori approach was presented by Guo et al. [13]. Several scans of the database are necessary ahead of time. The author developed an enhanced a priori algorithm version. This version requires fewer scans of the database.

Certain writers [14][15][16]have recently created some frequent pattern mining methods. These methods (BSO-ARM, PGARM, and PeARM) require fewer database searches to identify frequent patterns. However, all such algorithms have a shortcoming in that they only locate a subset of frequent items and do not uncover all possible frequent patterns from a data collection. The author of [17] created a novel data structure for storing patterns, which resulted in increased performance of the a priori method. [18] also presented an effective method for mining fuzzy periodic association rules. This proposed method checks the database nearly twice.

TECHNICAL REVIEW

This section provides a brief history of the Frequent Pattern Mining research methods in Horizontal and Vertical Data Layouts. The following table summarises the significant findings from the literature review.

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

Page 42- 48 © MANTECH PUBLICATIONS 2022. All Rights Reserved

8	DIFFSET	Vertical	BFS
9	TM	Vertical	DFS
10	TFP	Prefix tree (Horizonta)	Hybrid
11	SSR	Horizontal	DFS

CONCLUSIONS

The primary goal of frequent item set mining or association rule mining is to discover strong correlations between the items in a transaction data set. All of the researchers are aware that they must deal with large amounts of data while undertaking data mining. So the objective is to create algorithms that are both time and memory efficient. This article discusses frequent item set mining as well as the work done by many authors to execute mining on the transaction data set.

REFERENCES

1. International Conference on Data Mining, pages 166–177, May 2003.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.

2. 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.

3. Y. Chi, Y. Yang, Y. Xia, and R. R. Muntz. CMTreeMiner: 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.

4. 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.

5. 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

Page 42- 48 © MANTECH PUBLICATIONS 2022. All Rights Reserved

6. 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.

7. Weimin Ouyang , Qinhua Huang ,"Mining direct and indirect fuzzy association rules with multiple minimum supports in large transaction databases", 947 – 951, IEEE 2011.

8. 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.

9. X. Yan, J. Han, and R. Afshar. Clospan: Mining closed sequential patterns in large datasets. In SDM'03: Proceedings of the third SIAM

10. J. Pei, J. Han, and R. Mao. Closet: An efficient algorithm for mining frequent closed itemsets. In DMKD '00: ACMSIGMOD 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

Page 42- 48 © MANTECH PUBLICATIONS 2022. All Rights Reserved

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

Page 42- 48 © MANTECH PUBLICATIONS 2022. All Rights Reserved