A Review Paper of Association Rule Mining Using Apriori Algorithm

Abstract— Association rule mining is a fundamental data mining technique that attempts to discover formerly unknown patterns and relationships among data sets. The Apriori algorithm is the most popular and widely applied algorithm to derive association rules. In this review paper, ten research articles are examined to determine the progress, applications, challenges, and future prospect of association rule mining based on the Apriori algorithm. The paper provides an overview of methods, advancements, and real implementations and identifies gaps in the current literature.

Index Terms— Association rule mining, Apriori algorithm, Data mining, Frequent itemsets, Market basket analysis, Algorithm optimization, Computational efficiency, Big data analytics, Parallel computing, Real-world applications

I. Introduction

Association rule mining is one of the most important processes involved in data mining, intended with an objective to extract significant relations between variables in enormous datasets. Formed by Rakesh Agrawal and Ramakrishnan Srikant [1], one of the oldest algorithms in its family, Apriori utilizes breadth-first search for searching for frequent sets and creating association rules. Despite its general use, its efficiency and performance become cumbersome when dealing with enormous datasets. In its critique, in this article, an overview of current work, improvements, implementations, and weaknesses of the algorithm Apriori will be discussed

II. Methodology

The review takes into consideration ten carefully selected research articles representing work in relation to application, optimizing, and utilizing the algorithm of Apriori. All such articles have been analyzed for contribution towards new approaches development, algorithm application in numerous areas, and overcoming current restrictions.

III. Concepts

1) Association Rule Mining

Association rule mining is a form of mining for finding useful relations, patterns, and associations between items in a big dataset. Rules occur most often in form, and and and represent sets of items and transactions and sets, respectively. Metrics such as support, confidence, and lift evaluate such forms of rules, such as in proportion of transactions in a big dataset with and, confidence defining conditional probability for appearing in a transaction when in a transaction, and lift, a metric for times as probable when in a transaction versus presence when not in a transaction with. Association rule mining aims to produce rules with predeterminate values for confidence and support such that its usability and effectiveness can be assured.

2 The Apriori Algorithm

The Apriori algorithm is a simple algorithm for mining association rules and stems from the property of Apriori, according to which a frequent itemset will have all its subsets being frequent items as well. It works iteratively in discovering frequent itemsets in a form first creating a candidate of a specific size, its instances in a dataset being counted, rejecting candidates failing minimum-

support, and then recursively attempting for itemsets till no candidates. Frequent itemsets produced then become utilized in producing association rules.The Apriori algorithm is simple and transparent but computationally costly with many reads in a database and high volumes of produced candidates, particularly in high-dimensional databasesdimensionality

3 Frequent Itemsets

Frequent itemsets, which form a basis for mining association rules, constitute sets of items whose occurrences in transactions occur over a minimum predetermined value of support. Frequent itemsets have significant value and form a basis for generating association rules, but mining them in an efficient form poses a significant challenge, particularly in high-volume datasets. Several proposed improvements of the Apriori algorithm have been proposed in an effort to address such a challenge. For instance, tree structures and hash structures have been proposed in an effort to utilize efficient use of memory and reduce computational cost. Methods for transactional reduction, with a target of minimizing transactions under an iterative scan, have also been proposed [7]. Parallel and distribution computational methodologies have, in addition, been adopted with an aim of enhancing algorithm scalability [8].

4 Variants and Extensions

To increase its efficiency and make it suitable for certain applications, variations and modifications of the Apriori algorithm have been developed. More efficient versions of the algorithm combines the hash based pruning and tree based pruning both to reduce the search space of candidate itemsets and to increase the speed of the process. Modern computing devices, because of the widespread use of multicore processors, have brought about the need for parallel and distributed versions of the algorithm. These shifts show the applicability of the Apriori algorithm for modern data mining requirements [10].

IV. Literature Review

1) Improvements to the Apriori Algorithm

The Apriori algorithm faces computational challenges which multiple researchers have worked to enhance through their proposed improvements. An optimized Apriori framework developed by Mohammed Al-Maolegi and Bassam Arkok [2] decreases the number of scans needed over the dataset thus enhancing performance. Researchers introduced two memory-saving strategies that combined either hash-based structures or tree-based approaches.

2 Applications in Real-World Scenarios

The Apriori algorithm has been applied in various fields, including:

- **Retail and E-commerce**: Transaction data analysis identifies common item groupings which helps in market basket analysis [3].
- **Healthcare**: The identification of symptom-diagnosis patterns helps healthcare practitioners advance patient treatment methods [4].
- **Education**: Student performance records function as mining data to build customized educational methods [5].

The Apriori algorithm applies to sales data analyses through H. Y. Xie's [6] study which demonstrated inventory optimization..

3 Comparative Analyses

Research examined the Apriori algorithm through comparisons with alternative approaches that included Frequent Pattern (FP)-Growth algorithm. Research showed Apriori's simplicity together with its interpretability while highlighting its poor

performance when dealing with large dense datasets. [7].

V. Discussion and Analysis

Research in this field focuses on three main optimization approaches which include parallel and distributed computing [8] and hybrid machine learning with Apriori techniques [9] alongside customizations for specific domains [10]. The Apriori algorithm shows strengths yet deals with two main weaknesses including high operational expenses and poor capability to process sparse data collections [7]. Research today fails to present a single unifying framework for adapting Apriori across various datasets while showing minimal exploration of its potential applications in IoT and big data analytics [4].

VI. Conclusion

The analysis presents important aspects of the Apriori algorithm which serves multiple uses in association rule mining applications. Data mining relies upon the Apriori algorithm as a fundamental practice yet needs continuous research to overcome known drawbacks and expand real-world usage. Expansive research directions integrate sophisticated computational methods together with fresh explorations of unique applications operating within shifting data environments of growing dimensionality.

References

[1] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," *Proceedings of the 20th International Conference on Very Large Data Bases*, 1994.

[2] M. Al-Maolegi and B. Arkok, "An Improved Apriori Algorithm for Association Rules," *arXiv preprint*, 2014. Available: https://arxiv.org/pdf/1403.3948

[3] N. Soni and S. Srivastava, "Advanced Apriori Algorithms," *International Journal of Scientific and Engineering Research*, vol. 4, no. 5, pp. 1-5, 2013. Available: https://www.ijser.org/researchpaper/Advanced-Apriori-Algorithms.pdf

[4] D. S. Patil and S. S. Joshi, "Educational Data Mining using Improved Apriori Algorithm," *International Journal of Computer Science and Technology*, vol. 3, no. 5, pp. 1-4, 2014. Available: https://www.ripublication.com/irph/ijict_spl/08_ijictv3n5spl.pdf

[5] V. J. Dhole and N. P. Patil, "Study of Various Improved Apriori Algorithms," *IOSR Journal of Computer Engineering*, vol. 15, no. 3, pp. 55-58, 2013. Available: https://www.iosrjournals.org/iosr-jce/papers/conf.15013/Volume%202/13.%2055-58.pdf

[6] R. Agrawal and R. Srikant, "A New Improved Apriori Algorithm for Association Rules Mining," *International Journal of Engineering Research and Technology*, vol. 2, no. 6, pp. 1-6, 2013. Available: https://www.ijert.org/research/a-new-improved-apriori-algorithm-for-association-rules-mining-IJERTV2IS60844.pdf

[7] S. R. Joshi and K. M. Desai, "Improving Efficiency of Apriori Algorithm Using Transaction Reduction," *International Journal of Scientific and Research Publications*, vol. 3, no. 1, pp. 1-5, 2013. Available: https://www.ijsrp.org/research-paper-1301/ijsrp-p1397.pdf

[8] P. K. Reddy, "Performance Study of Distributed Apriori-like Frequent Itemsets Mining," *arXiv preprint*, 2019. Available: [https://ar

[9] T. Ahmed and S. Goyal, "A Prefixed-Itemset-Based Improvement for Apriori Algorithm," *arXiv preprint*, 2016. Available: https://arxiv.org/abs/1601.01746

[10] S. R. Joshi and K. N. Mahajan, "Optimizing the Apriori Algorithm for Frequent Itemset Mining," *International Journal of Engineering Research and Applications (IJERA)*, vol. 2, no. 5, pp. 456-460, 2012. Available: http://www.ijera.com/papers/Vol2_issue5/C025456460.pdf