International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Review on Optimization of Apriori Algorithm for Finding the Association Rules in Different Business and Other Datasets for Retrieval of Relations Between Different Entities

Prof. Pradeep N. Fale[1] , Narendra Moundekar[2], RiteshSaudagar[3], Prajwal Kamdi[4], Mrunali Rode[5], Janvi Borkar[6]
[1]Assistant Professor, Department of Information Technology, Priyadarshini College of Engineering, Nagpur, Maharashtra, India
[2-6]Department of Information Technology, Priyadarshini College of Engineering, Nagpur, Maharashtra, India

Article Info
Volume 9, Issue 2
Page Number : 271-276
Publication Issue :
March-April-2022
Article History
Accepted : 01 April 2022
Published: 06 April 2022

ABSTRACT

The secret of doing successful business lies in the accuracy of the decisions taken for the inventory management, production plans, being customer centric and being agile for the market developments. The business data processing for any business is huge one and may contain many hidden things, which must be revealed out intelligently and with optimization with respect to the time and other source constraints. Many times, it is beyond the scope of the human mind to figure out and relate the interdependencies of the multiple factors embedded in the business data and hence the machines could help in this context to make the task easy. When it comes to find the Association rules between different products of any shop or store, the Apriori algorithm tops the choice. The current review work depicts the attempts to use the Apriori algorithm in an optimized way and implementing the same according to the prevailing conditions.

Keywords: Customer Centric, Apriori Algorithm, Association Rules

I. INTRODUCTION

Apriori algorithm refers to the algorithm, which is used to calculate the association rules between objects. It means how two or more objects are related to one another. In other words, we can say that the apriori algorithm is an association rule leaning that analyzes that people who bought product A also bought product B. The primary objective of the apriori algorithm is to create the association rule between different objects. The association rule describes how two or more objects are related to one another. Apriori algorithm is also called frequent pattern mining. Generally, you operate the Apriori algorithm on a database that consists of a huge number of transactions. Let's understand the apriori algorithm with the help of an example; suppose you go to Big Bazar and buy different products. It helps the customers buy their products with ease and increases the sales performance of the Big Bazar. In this tutorial, we will discuss the apriori algorithm with examples. We take an example to understand the concept better. You must have noticed that the Pizza shop seller makes a pizza, soft drink, and breadstick combo together. He also offers a discount to their customers who buy these combos. Do you ever think why does he do so? He thinks that customers who buy

pizza also buy soft drinks and breadsticks. However, by making combos, he makes it easy for the customers. At the same time, he also increases his sales performance. Similarly, you go to Big Bazaar, and you will find biscuits, chips, and Chocolate bundled together. It shows that the shopkeeper makes it comfortable for the customers to buy these products in the same place. The above two examples are the best examples of Association Rules in Data Mining. It helps to learn the concept of apriori algorithms.

Apriori algorithm refers to an algorithm that is used in mining frequent products sets and relevant association rules. Generally, the apriori algorithm operates on a database containing a huge number of transactions. For example, the items customers but at a Big Bazar. Apriori algorithm helps the customers to buy their products with ease and increases the sales performance of the particular store.

II. RELATED WORKS

In [2], Ashish Shah et al proposed a method, which tries to find the frequent item sets in a bottom-up manner as well as a top down manner. It maintains a list of al item sets. While passing through the database, it counts the support of these large candidate item sets to see check if they are actually frequent. In such an event, it is known that all the subsets of these frequent sets are going to be frequent and so they can be remove from the list to be scanned. This will increase performance. If lucky, it may discover a very large maximal frequent item set very early in the algorithm. Consider a pass k, which signifies the length of item sets to be searched. If some item set that is a maximal candidate item set, say X, if frequent then all its subsets and constituents must be frequent as well. Hence, all of its subsets should be pruned from the set to be scanned. In this procedure, all candidate sets considered in the bottom-up direction in the pass are removed. If this set subsumes all the candidate sets of level k, then there is no need to proceed further and thus save many database passes. Clearly, the proposed technique has an advantage over apriori algorithm when the largest frequent item set is long. In every step, author prune all items whose support is less than the min sup count from the maximal candidate item set. Let's call the maximal candidate item set M. After generating candidate item sets, author divides them into 2 categories. One whose support is above the minimum support threshold go into category L and ones whose support is below the threshold into category S.

In [3] Yi cong et. al states that Apriori algorithm generates large number of candidate item sets while discovering frequent item sets and with repeated scanning I/O takes large time. Author states that after scanning the database for the first time, new database is obtained. At the same time, judgment dataset is also added and the irrelevant and meaningless candidate item sets are removed which reduces the time consumption cost. When this approach was compared with the classic Apriori algorithm, following results obtained.

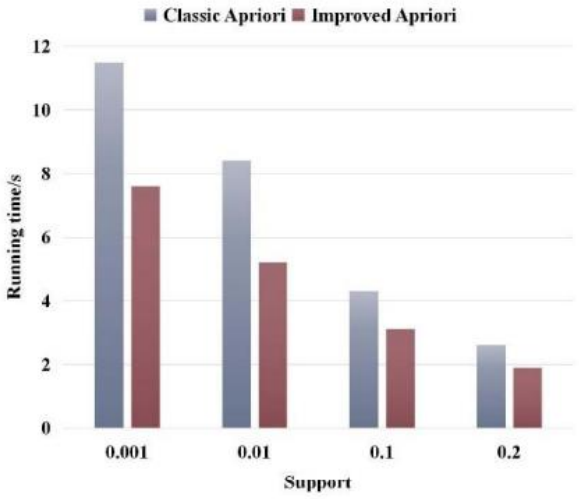

Fig. 1 Comparison of two algorithms with same support [3]

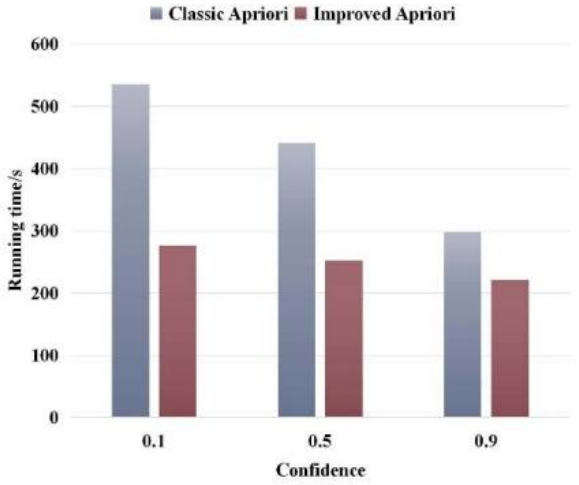

Fig. 2 Comparison of two algorithms with same confidence [3]

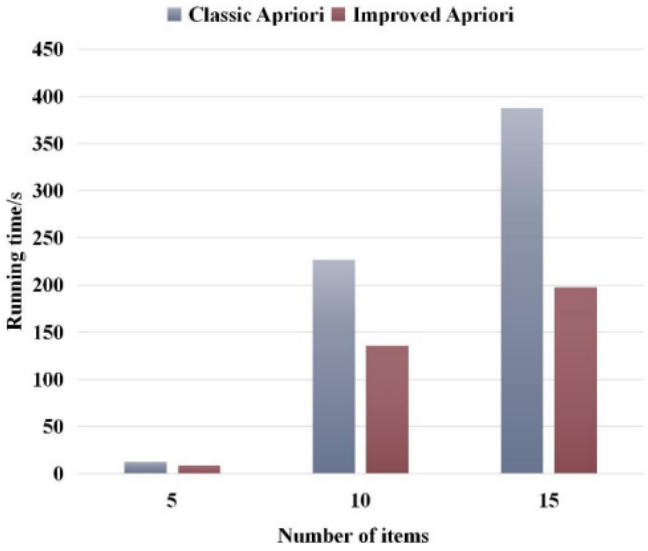

Fig. 3 Comparison of two algorithms with same number of items [3]

In [4], Hanxiao Zhang, Wei Song, Lizhen Liu, Hanshi Wang et. al. mainly discuss the application of Matrix Apriori algorithm in Web log mining based on matrix storage. First, authors analyze the improvements of Matrix Apriori and describe the process of the algorithm. Authors make some comparisons of several association rules algorithms. Then, Matrix Apriori algorithm is applied to Sogou search log and shoes website search log. Finally, according to the results of the Web log mining, authors can make personalized recommendation and optimize site settings. In order to make some comparison, Apriori algorithm, Apriori Tid algorithm and Apriori_LB algorithm are applied to Web log mining. As the increasing of the data, the Matrix Apriori's efficiency has obvious improvement as shown in fig 4 as shown figure.

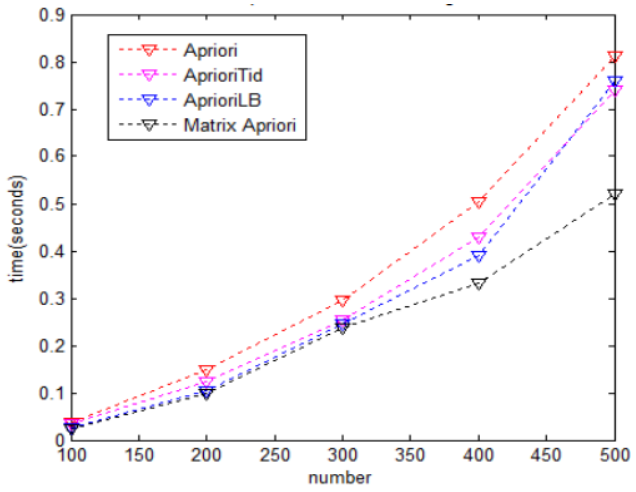

Fig.4 The comparison between Apriori algorithm, AprioriTid algorithm, Apriori_LB algorithm and Matrix Apriori [4]

In [5] XU Hongfei, LIANG Xuesong, CUI Wei, LIU Wei et. al. proposed Apriori algorithm of riddling compression and has carried on the simulation, the result demonstrated the Apriori algorithm of riddling compression can improve the efficiency greatly. It can greatly reduce the candidate frequent item sets, keeps the completion of frequent item sets, reduces the cost of computing, and improve the efficiency of algorithm.

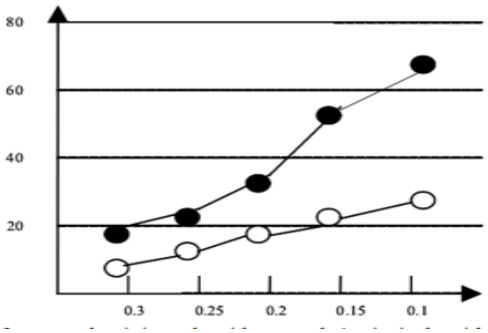

Fig. 5: Improved mining algorithms and Apriori algorithm running time comparison [5]

In [6] Mohammad Javad Shayegan Fard and Parsa Asgari Namin et. al. carried review which aims to present an insight into the works done in the

intersection of two matters: big data and the Apriori algorithm. It is concerned with Apriori based algorithms presented in the recent decade with a focus on the three popular big data platforms: Apache Hadoop, Spark, and Flink. Also, a major point of each approach and solution is presented. The prominent conclusions of the reviews are depicted in following figures.

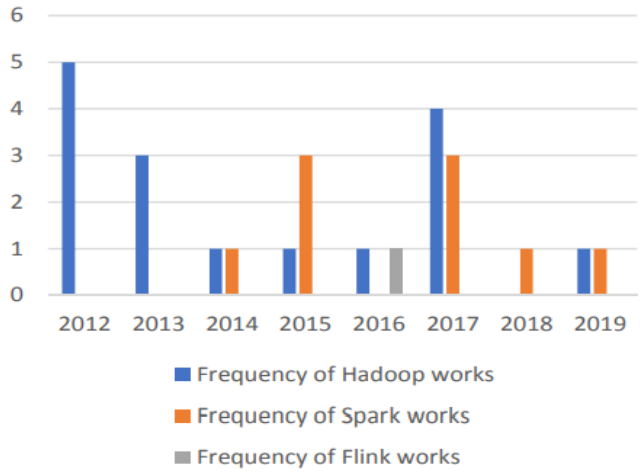

Fig.6: Frequency of new algorithms based on Apriori, introduced in each year [6]

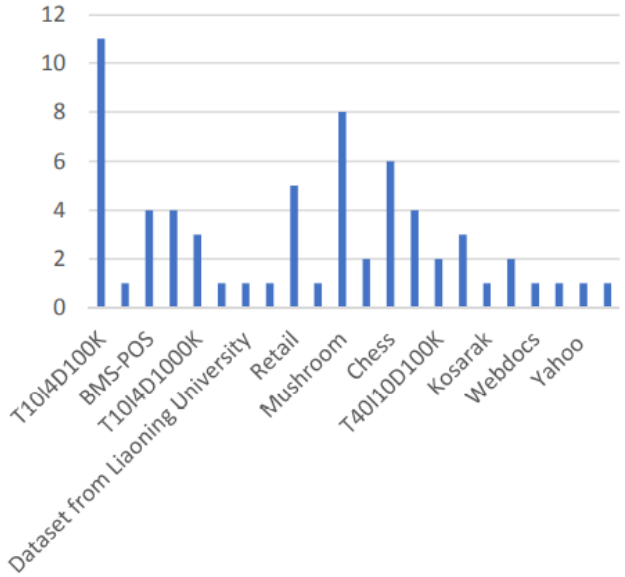

Fig.7: Frequency of various datasets used in reviews [6]

In [7] Fulan Ye et. al. see that the common point of the improved algorithm and Apriori Algorithm is to get the frequent item sets Lk whose support degree is not less than the minimum support degree given by users by scanning data The improved algorithm first transforms the database into a HASH table, so only the items in the k-item set need to be scanned when calculating the support, without scanning the whole HASH table. Secondly, the improved algorithm counts the elements that will participate in the combination before considering the candidate item set of combination, and decides to exclude some elements that do not meet the combination condition according to the result of counting, which reduces the possibility of combination Directly reduces the number of loop judgments. The improved algorithm improves the efficiency largely. The results are shown in fig. below.

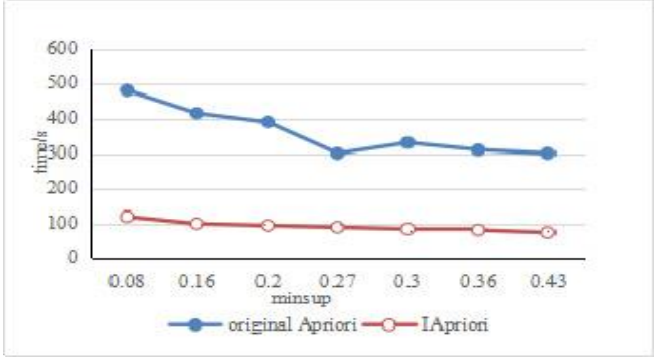

Fig.8: Algorithm efficiency comparison [7]

In [8] Hongqin Wang, Hongxia Wang, Huiyong Jiang, Lina Yuan et.al., authors proposed the approach in which Apriori algorithm is parallelized and improved by using MapReduce in the Hadoop platform. The improved algorithm is called IApriori. The IApriori uses the Map function to block the original data set, which is distributed by the main process to the computers with Hadoop cluster, all of the partitioned k-item candidate set support numbers obtained by the Map process are combined to obtain the global k-item candidate set support number, which is done by the Reduce function in MapReduce. The complete set of K frequent itemsets is calculated from the support for the complete set of candidate items. Each computer starts the next Map process to handle the second block of data, after the global set of frequent k terms for a block of data has been calculated, and so on, until all the data

blocks are processed. The results obtained by comparing the Apriori and the IApriori are shown in the figure below with minimum confidence and support .

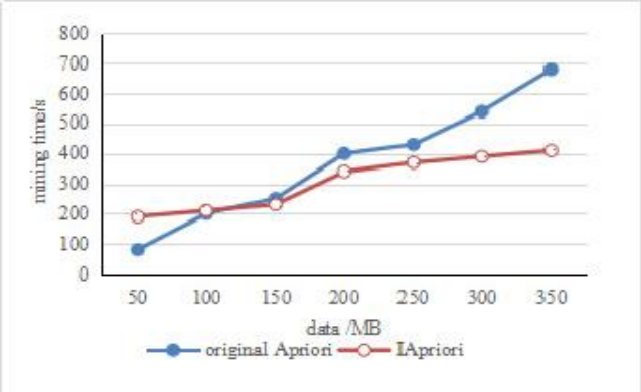

Fig.9: performance comparison of the two algorithms [8]

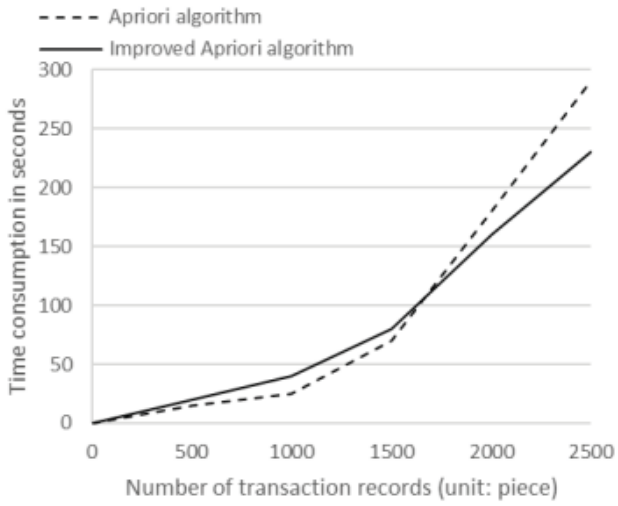

Fig.10: Comparison of response time with minimum support [8]

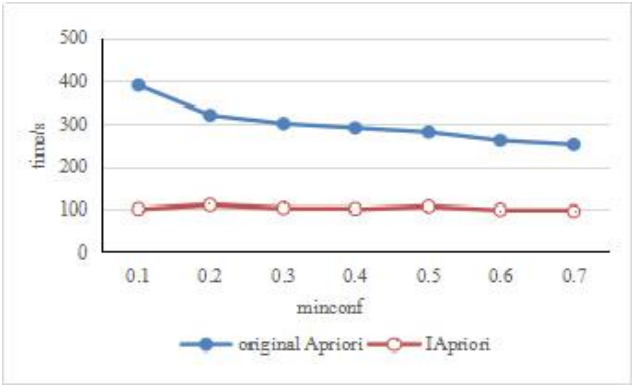

Fig.11: Comparison of response time with change in minimum confidence[8]

III. CONCLUSION

In this review, we reviewed many approaches that are related with improving the original classic Apriori Algorithm some of which are resource efficient and some are output accuracy specific. The best improvement in an Apriori algorithm is in terms of the saving of memory and time complexity and the same time producing the most relevant association rules that can yield the better profit by helping in taking the optimized inventory and business decisions.

IV. REFERENCES

[1]. https://www.javatpoint.com/apriori-algorithm

[2]. Ashish Shah," Association Rule Mining with Modified Apriori Algorithm using Top down Approach", 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT).

[3]. Yi Cong, "Research on Data Association Rules Mining Method based on Improved Apriori Algorithm", International conference on Big Data and Artificial Intelligence and Software Engineering (ICBASE) 2020.

[4]. Hanxiao Zhang, Wei Song, Lizhen Liu, Hanshi Wang,"The Application of Matrix Apriori Algorithm in Web Log Mining", 2017 IEEE 2nd International Conference on Big Data Analysis.

[5]. XU Hongfei , LIANG Xuesong , CUI Wei , LIU Wei," Research on an Improved Association Rule Mining Algorithm", International conference on Power Data Science (IPCDS) 2019, IEEE.

[6]. Mohammad Javad ShayeganFard ,ParsaAsgariNamin," Review of Apriori based Frequent Itemset Mining Solutions on Big Data", 2020 6th International Conference on Web Research (ICWR) .

[7]. Fulan Ye," Research and Application of Improved APRIORI Algorithm Based on Hash Technology", Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC) 2020

[8]. Hongqin Wang, HongxiaWang, Huiyong Jiang, Lina Yuan,"Research on an improved algorithm of Apriori based on Hadoop", 2020 International Conference on Information Science, Parallel and Distributed Systems (ISPDS).

[9]. Zhou Guojun. Study on multi-keyword sorting method based on HadoopJ]. Computer Engineering and Applications.2016, 52(17), pp.79-83. (in Chinese).

[10]. CHEN Z F, FENG J. An optimized data mining algorithm based on Apriori algorithm J]. Computer and Modernization, 2016 (9). pp. 1-5. (in Chinese).

[11]. Lin, X.: Mr-apriori: Association rules algorithm based on mapreduce. In Software Engineering and Service Science (ICSESS), 2014 5th IEEE international Conference on.pp.141-144.

[12]. SUN Xuebo, SHI Feida. Research and optimization of Apriori algorithm based on Hadoop J]. Computer engineering and design.2018, 39(1), pp.126-133. (in Chinese).

[13]. Agaoglu, Mustafa.Predicting Instructor Performance Using Data Mining Techniques in Higher EducationJ]. IEEE Access, 2017, pp.1-1

[14]. Bin Chen, Peter Hass, Peter Scheuermann, "A New Two-Phase Sampling Based Algorithm for Discovering Association Rules", SIGKDD '02 Edmonton, Albetra, Canada © 2002 ACM 1 58113 567 X/02/2007.

[15]. Judith Pavón, Judith N, Viana S, and S mez," Matrix apriori: speeding up the search for frequent patterns C]. Iasted" International Conference on Database and Applications. ACTA Press, pp 75—82, 2016.

[16]. Hu J, Li B, "An improved algorithm of association rules in the application of web logs J]". Energy Procedia, 13, pp 1282-1286, 2011.

[17]. Wang Shuyan, Wang Xiaomei. Department of Computer, An X,"Application of Association Rules for Medical Data Classification J]", Computer & Digital Engineering, 2009. 18J.Hu, A. Mojsilovic

[18]. Duan Q, Li J, Wang Y,"The Application of Fuzzy Association Rule Mining in E-Commerce Information System Mining J]", pp 631-635, 2012.

[19]. Agrawal, Imielinski, and Swami,"Mining Association Rules between Sets of Items of the ACM SIGMOG", International Conference on Management of data, pp 207—216, 1993

[20]. Goethals B,"Survey on Frequent Pattern MiningJ].", Univ of Helsinki, 2003, 63(14):47-Lu Z, Yao Y, Zhong N. Web Log MiningJ]. Web Intelligence, 2003:173-194

[21]. Guo Tao, Zhang Daiyuan,"Research and application of data mining association rules Apriori algorithm", computer technology and development, 2011,06:101-103

[22]. Wu Yugang, Qin Yong, Song Jiguang, Yang Zhongming," Review of the research on Intrusion Detection Algorithm Based on association rules", computer engineering and design, 2011,03:834-838

[23]. Li Yuanbo, "Tourism recommendation based on association rules algorithm" Shaanxi Normal University, 2015

[24]. Yao Wendy,"Algorithm of association rules data mining research and application in Colleges and Universities Based on the achievements" ,Southwest Jiao Tong University, 2015

[25]. Aman Chang, Li Lingjuan, "Research on Association Rule Mining Algorithm in cloud computing environment", computer technology and development, 2011,02:43-46

Cite this article as :

Prof. Pradeep N. Fale, Narendra Moundekar, RiteshSaudagar, Prajwal Kamdi, Mrunali Rode, Janvi Borkar, "Review on Optimization of Apriori Algorithm for Finding the Association Rules in Different Business and Other Datasets for Retrieval of Relations Between Different Entities", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 2, pp. 271-276, March-April 2022.
Journal URL : https://ijsrset.com/IJSRSET229246