International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Indexed, Double Blind, Open Access, Peer-Reviewed, Refereed-International Journal.
SJIF Impact Factor =7.938, January-June 2024, Submitted in April 2024, ISSN -2393-8048

Optimisation of Apriori Algorithm to improve performance in E-Commerce Applications

Ashfaq Ahmed Khan, Research Scholar, Department of Computer Science, Mewar University, Gangrar
Dr. Ganesh Gopal Varshney, Supervisor, Department of Computer Science, Mewar University, Gangrar

1. Introduction

E-commerce markets are growing at noticeable rates. The online market is expected to grow by 56% in 2015–2020.E-commerce allows customers to overcome geographical barriers and allows them to purchase products anytime and from anywhere.

Retailing in India is one of the pillars of its economy and accounts for about 10 percent of its GDP. The Indian retail market is estimated to be worth $1.3 trillion as of 2022 and estimated to reach $2 Tn by 2032. India is one of the fastest growing retail markets in the world, with 1.4 billion people.

The major Retailing industries in India are Reliance, More, Bigbazar, Amazon, Flipcart etc.

The retail industry is a major application area for data mining because it collects huge amounts of records on sales, users shopping history, goods transportation, consumption, and service. The quantity of data collected continues to expand promptly because of the increasing ease, accessibility, and popularity of business conducted on the internet, or e-commerce.

Retail data mining can help identify user buying behaviours, find user shopping patterns and trends, enhance the quality of user service, achieve better user retention and satisfaction, increase goods consumption ratios, design more effective goods transportation and distribution policies, and decrease the cost of business.

2. Data Mining Techniques – Apriori Algorithm

There are several data mining techniques used the retail sector. The top five techniques include:

- Classification analysis
- Regression analysis
- Clustering analysis
- Anomaly detection and
- Association analysis.

Each technique solves a specific problem and offers unique insights.

In our Research we have taken up Association Analysis and in particular **Apriori Algorithm** for improving the prediction of Customer Behaviours, Patterns so as to improve the retail sector.

Apriori algorithm was the first algorithm that was proposed for frequent itemset mining. This algorithm uses two steps “join” and “prune” to reduce the search space. It is an iterative approach to discover the most frequent itemsets.

Apriori says:

The probability that item I is not frequent is if:

•P(I) < minimum support threshold, then I is not frequent.

•P (I+A) < minimum support threshold, then I+A is not frequent, where A also belongs to itemset.

•If an itemset set has value less than minimum support then all of its supersets will also fall below min support, and thus can be ignored. This property is called the Antimonotone property.

The steps followed in the Apriori Algorithm of data mining are:

Join Step: This step generates (K+1) itemset from K-itemsets by joining each item with itself.

Prune Step: This step scans the count of each item in the database. If the candidate item does not meet minimum support, then it is regarded as infrequent and thus it is removed. This step is performed to reduce the size of the candidate itemsets.

Apriori algorithm is a sequence of steps to be followed to find the most frequent itemset in the given database. This data mining technique follows the join and the prune steps iteratively until the most frequent itemset is achieved. A minimum support threshold is given in the problem or it is assumed by the user.

#1) In the first iteration of the algorithm, each item is taken as a 1-itemsets candidate. The algorithm will count the occurrences of each item.

#2) Let there be some minimum support, min_sup (eg 2). The set of 1 – itemsets whose occurrence is satisfying the min sup are determined. Only those candidates which count more than or equal to min_sup, are taken ahead for the next iteration and the others are pruned.

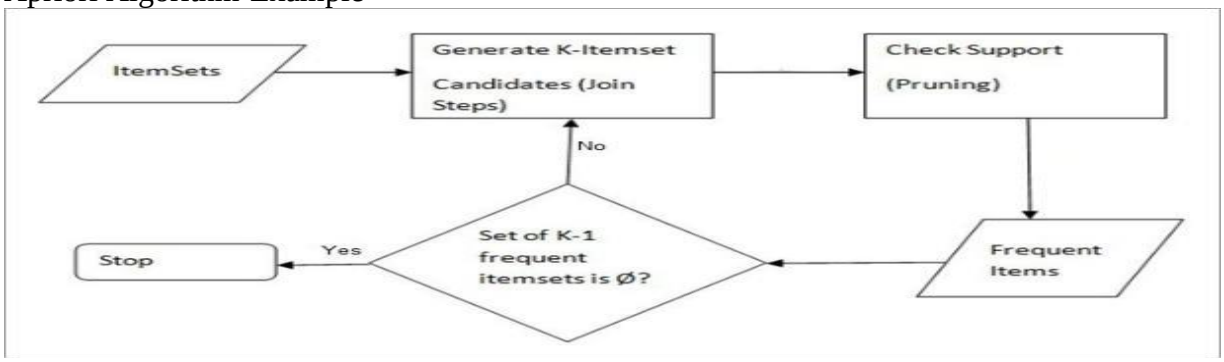
#3) Next, 2-itemset frequent items with min_sup are discovered. For this in the join step, the 2-itemset is generated by forming a group of 2 by combining items with itself.

#4) The 2-itemset candidates are pruned using min-sup threshold value. Now the table will have 2 –itemsets with min-sup only.

#5) The next iteration will form 3 –itemsets using join and prune step. This iteration will follow antimonotone property where the subsets of 3-itemsets, that is the 2 –itemset subsets of each group fall in min_sup. If all 2-itemset subsets are frequent then the superset will be frequent otherwise it is pruned.

#6) Next step will follow making 4-itemset by joining 3-itemset with itself and pruning if its subset does not meet the min_sup criteria. The algorithm is stopped when the most frequent itemset is achieved.

Apriori Algorithm-Example



Example of Apriori: Support threshold=50%, Confidence= 60%

- Applications Of Apriori Algorithm
- In Education Field: Extracting association rules in data mining of admitted students through characteristics and specialties.
- In the Medical field: For example Analysis of the patient's database.
- In Forestry: Analysis of probability and intensity of forest fire with the forest fire data.
- In Retail Sector: Apriori is used by many companies like Amazon in the Recommender System and by Google for the auto-complete feature.

Apriori Algorithm-Limitations

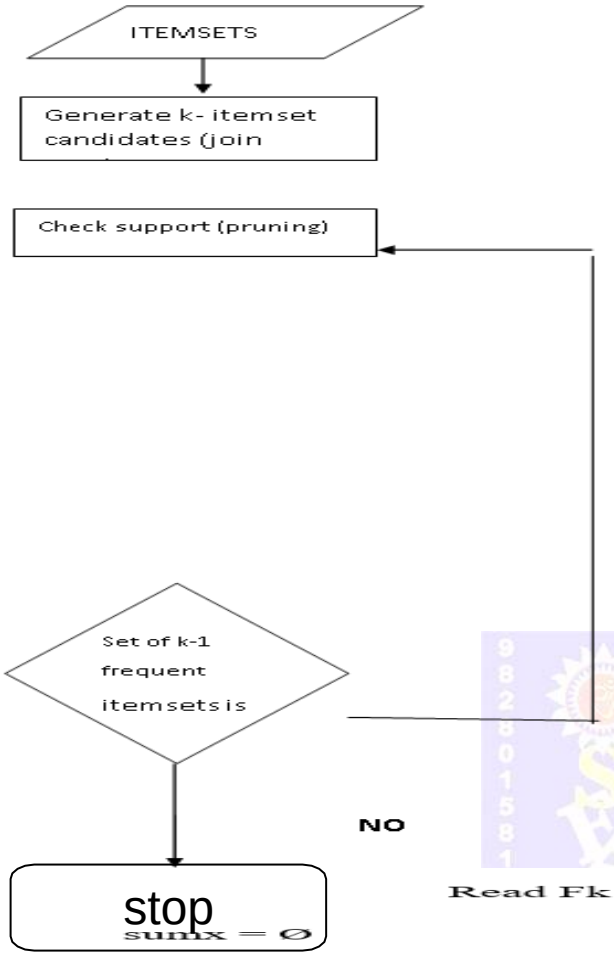
- If the dataset is small, the algorithm can find many false associations that happened simply by chance.
- when working with large datasets the Apriori algorithm is slow, inefficient, and uses a lot of resources as it has to scan the database many times, generate a large number of candidate sets and check each of them.
- There are some factors which influence to the time complexity of an a priori algorithm. These are the minimum support threshold, the number of items, the number of transactions, the average transaction width, and the generation of frequent 1-itemsets, candidate generation and support counting.

3. Modified Apriori Algorithm

In order to address the limitations in the Apriori Algorithm, tproposed amodified Apriori Algorithm where it combines regression methods with an apriori algorithm to filter out

superfluous supermarket patterns and zero in on the most common ones, all while speeding up the search for these patterns' frequent occurrences. This approach was created with the intention of creating the common pattern in less time.

FLOW CHART-MODIFIED APRIORI ALGORITHM



Proposed Algorithm

- **i= 1**

Fk = {i|I Є I ∧ σ ({i}) ≥N x min sup}

- **{Find all frequent 1-itemsets} Repeat**
- **K = k+1**

Ck = apriori-gen (Fk-1)

- **{Generate candidate itemsets} For each transaction t Є T do Ct = subset (Ck, t)K**
- **{identify all candidates that belong to t} for each candidate itemset**
- **C Є Ct do**

σ (c) = σ (c) + 1

- **{increment support count} end for**
- **end for**

Fk = {c|c Є Ck ∧ σ ({c}) ≥N x min sup}

{extract the frequent k – itemsets By Pruning Using Linear Regression

sumx = Ø

sumxsq = Ø

sumy = Ø

sumxy = Ø

For each transaction do

Read Fk, Uk

Sumxq = sumxq +2

Sumy = sumy + U_k * sumxsq

Sumxy = sumxy + F*U end for

Denom = n * sumxsq – sumx *sumx

F_k Ø = (sumy * sumxsq – sumx * sumxy) / denom

U_k = (n * sumxy – sumx * sumy) / denom

write U_k, F_k Ø

until F_k = Ø

Result = $U_k F_k$

4. Implementation & Results

The proposed algorithm is implemented in WEKA Tool.The metrics used is Execution Time. Both the Apriori Algorithm and Proposed modified Apriori Algorithm has been compared and results are observed

We have used Data Set 'Super-Market'

RESULT ANALYSIS USING REGRESSION TECHNIQUES WITH DATA MINING.

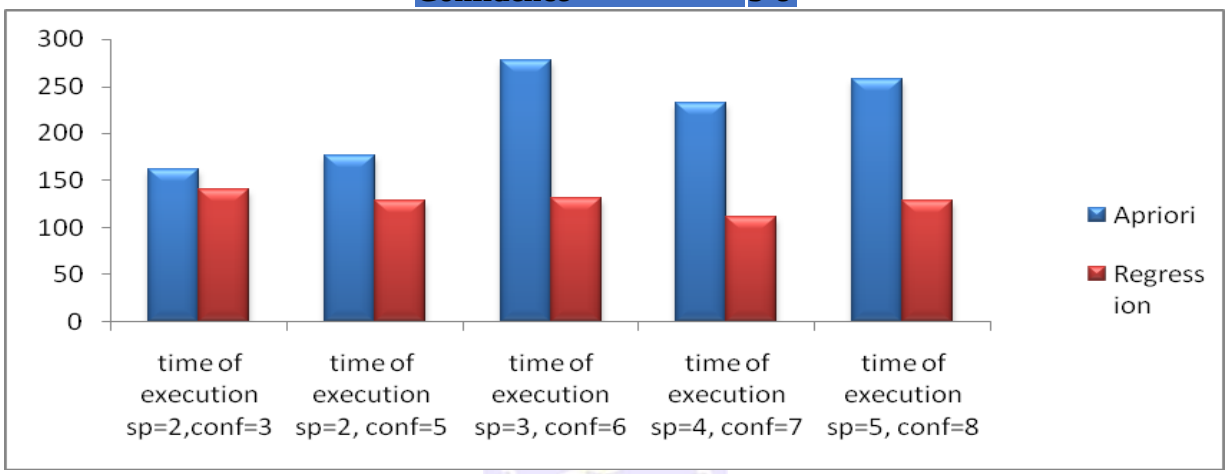
SIMULATION PARAMETERS

In this table, there are different time of execution (in sec) of both algorithm i.e., apriori algorithm and apriori with regression technique on different support count and different confidence.

-- item set are all parameters

Table 1 Comparison of Time Execution in Apriori and Apriori with Regression on different support and different Confidence

Algo	Sec
Apriori Algo	281
Apriori with Regression	
Support Count	2-5
Confidence	3-8



In this result analysis we will create the graph based on different support and different confidence rules. As well as we have used the dot net frame work for implementation of this time execution result. We analysis the linear regression technique reduced the time of execution of those item set values whose confidence to be predict should be zero.

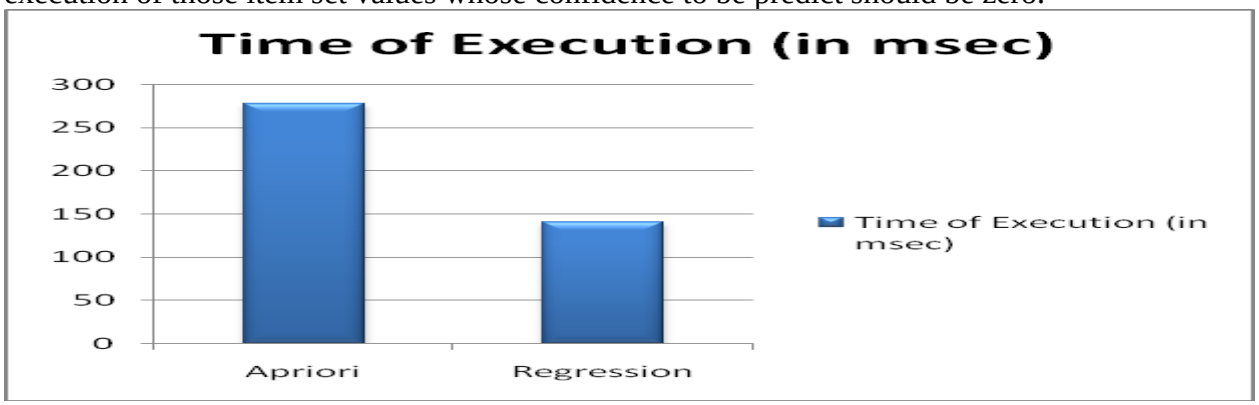


Fig.2 Time optimization of Apriori with Regression on different support and different confidence

The above graph shows the difference of time on different support and different confidence value in between the Apriori algorithm and implemented algorithm i.e. Apriroi with regression algorithm based on Table 1.

5. Conclusion

After implementing the developed approach get the conclusion that the Apriori with Regression algorithm is proposed an effective algorithm to reduce the consumption of time. The work is carried out on partitions of a dataset rather than applying on full dataset which results in reduction of time taken by the Apriori Algorithm. Instead of repeated scan of the

original database, it is scanned only once to form large 1 item-set from which further computations are carried out. This reduces the time involved in scanning the dataset which in turn reduces the overall time to a greater extent. The minimum support value is also calculated at each pass which removes the unnecessary formed sets. Although the algorithm is simple, it carries out more effective pruning.

Future Research Scope

The results of the Apriori algorithm with regression technique are satisfactory on linear regression. The future work may include the implementation of the Apriori algorithm with regression technique on multiple regression and logistic regession. The results are expected to be different in that case. And also wish to see whether the database scans reduces by using this approach. We performed the regression as a linear approach. Anybody will use the multiple regression approach for increase the productivity of pattern rules matching.

References

Dr.Sarika Agarwal, Dr. Mukesh Agarwal,"use of data mining in E-Commerce platforms in india",International journal of innovations & research analysis, vol.02, No.2, April-June 2022

S. Kavitha, Dr. S. Manikandan, "Customer Behavior on shopping using data mining techniques", Journal of emerging technologies and innovative research, vol. 6, Issue.6, June 2019

Swati Mahesh Joshi," Market basket analysis using Apriori Algorithm in data mining", International Journal of engineering and technology, vol.05, Issue.04, April 2018

Prateeksha Tomar, Amit Kumar Manjhvar: Survey Report on Various Decision Tree Classification Algorithm Weka Tool. International Journal of Computer Science and Engineering (IJCSE), March 2019.

Swati Gupta " A Regression Modeling Technique on Data Mining", International Journal of Computer Applications (0975 – 8887) ,Volume.116 No. 9, April 2015.

Himani Bathla, Ms.Kavita Kathuria "Apriori Algorithm and Filtered Associator in Association Rule Mining", International Journal of Computer Science and Mobile Computing, Vol.4 Issue.6, June- 2015