A Comparative Study of Association Rule Mining Algorithms in Data Mining

Pushpa Devi[1], Vikrant Singh Bhardwaj[2], Dr.Kishori Lal Bansal[3]
[1, 2] Dept of Computer Science
[3]Professor, Dept of Computer Science
[1, 2, 3] Himachal Pradesh University, Summerhill, Shimla-5

Abstract- *Tons of data is being generated and processed every data. This data is of no use if we cannot utilize it for the betterment of our organizations or businesses. This gives a rise to Data Mining applications, using data mining techniques we can find interesting insights from huge amount of Data. Among various Data Mining techniques Association Rule Mining is very popular. ARM finds correlation among various Itemsets of Database which helps in understanding market trends, customer behavior and make business decisions accordingly. The most popular association rule mining algorithms are: Apriori, Frequent Pattern Growth, Eclat, Association Outlier etc. This Paper aims to provide a comparative study of Association Rule Mining(ARM) Algorithms in Data Mining. This paper will be helpful for future researcher to understand which algorithm will perform better in what situation and which algorithm to choose over other algorithms based on their requirement of the study.*

Keywords- Data Mining, Association Rule Mining, Apriori, FP Growth, Eclat.

I. INTRODUCTION

Data Mining or Knowledge Discovery in Database is a technique or process which helps to extract useful insights or knowledge from large set of Database[1]. This process of extracting knowledge consists of several steps like Data cleaning, Data Integration, Data Reduction or Selection, Data transformation, Data mining, Pattern evaluation, and Knowledge Representation. Data Mining has many application domains like Market Basket Analysis, Fraud Detection, Customer Relationship Management, Healthcare Sector and Many more. There are various Data Mining Techniques and among these techniques Association Rule Mining is the most popular as it helps in better business Management and Decision making.

Association Rule Mining (ARM)

ARM is the process or technique of finding the interdependence between the Itemsets of the Database. Association rule mining describes the relationship between itemsets of the database through association rules. Association Rules are the simple if then statements. The left part of rule is called antecedent and right part of the rule is called consequent. For e.g. *"If a customer buys bread, he's 70% likely of buying butter."*

Here, bread is an antecedent and butter is the consequent.

Rule Evaluation Metrics:

Support(s): Support gives the manifestation of how frequently the itemset appear in a dataset.
Confidence(c): Confidence value indicates how valuable the rule is or the number of times the association rules are found true [16].
Lift(l): The lift value of an association rule is **the ratio of the confidence of the rule and the expected confidence of the rule**[24].

Applications of Association Rule Mining

Some of the most popular application areas of Association Rule Mining are:

1. Market Basket Analysis: Market Basket Analysis(MBA) is the most popular applications of ARM. Association rules are used to predict the likelihood of buying a product with the purchase of another product. For example, Likelihood of buying Butter when a person buys Bread.
2. Medical Diagnosis: Association rules can be useful for assisting physicians for curing patients, we can identify the probability of the occurrence of illness on the basis of various factors and symptoms.
3. Census Data: This data can be used to plan efficient public services like education, health, transport as well as help public businesses for setting up new factories, shopping malls, and even marketing particular products.

Association Rule Mining Algorithms

Number of algorithms are available for association rule mining namely: A-priori, FP – Growth, Eclat, A- Priori

Prefix Tree, Association Outlier, Frequent Item sets, Supervised Association Rule, Supervised Association Tree, AIS, SETM, Apriori tid, Apriori hybrid etc.A brief description about the frequently used ARM algorithms is given below:

a) Apriori Algorithm:

Apriori was first presented in 1994 by Ramakrishnan Srikant and Rakesh Agarwal. A-priori usethe information of frequent itemset properties that has been previously discovered and "bottom-up" approach to bring out frequent item sets. The Frequent itemsets are extended level by level this process is known as ***candidate generation***. When there are no further successful extensions found Apriori algorithm will stop.

b) **FP-growth algorithm**

Frequent Pattern – Growth algorithm consists of two passes. It counts the occurrences of items in first pass and stores this count of occurrences in a Header Table. In next Pass, FP- Growth algorithm create a FP-Tree structure.

c) Eclat algorithm

Eclat stands for Equivalence Class Transformation. It is a depth-first search algorithm. Eclat algorithm is suitable for both sequential as well as parallel execution.

d) AIS

AIS algorithm is the first algorithm developed in 1993 for association rules mining. This algorithm was mainly designed to discover qualitative rules. In AIS algorithm there can be only one item in the consequent. The main limitation of AIS algorithm is that it unnecessarily generates and count huge number ofcandidatesItemsets which later turn out to be small [25].

e) SETM:

SETM algorithm is proposed in 1994. It uses Relational Database Management System to store and compute association rules. It is a set-oriented algorithm for mining Frequent Itemsets. It uses sort and merge join. The benefits of using SETM is that memory limitation is not theirBuffer management and Query optimization is handled by DBMS.The main Limitation of SETM is that it generates too many candidate Itemsets[14]

II. LITERATURE REVIEW

F.H. AL-Zawaidah, and Y.H. Jbara [4]developed a novel improved algorithm which is based on the traditional Apriori algorithm approach with some additional features. Authors has compared performance of both algorithms and the novel algorithm outperformed better than the traditional algorithms. **A.B.M. Rezbaul Islam, and T.S. Chung [5]**proposed a novel and improved Frequent Pattern Tree with **a table** & **a new algorithm**. This new algorithm mines all the possible frequent Itemsets without generating the conditional Frequent Pattern tree (FP – Tree).**J. Nahar et al., [9]** studied association rule mining to detect factors that are responsible for heart disease in both males and females. They studied both sick and healthy factors which contribute to heart disease. They used UCI Cleveland dataset for this study. Apriori, Predictive Apriori and Tertius Association rule mining algorithms used for rule generation. They found that Females have less chance of coronary heart disease then males. **T.A. Kumbhare et al., [10]**has compared six basic Association Rule Mining Algorithms named as AIS (1993), SETM(Set Oriented-Mining), Frequent Pattern-Growth, Apriori, A-priori TID, A-priori-Hybrid algorithms. Parameters used for comparison are accuracy, data support and execution speed. Results shows that Frequent-Pattern Growth algorithm has performed better than the other five algorithms.**G. Kaur [11]** reviewed the basic concepts of Association Rule Mining technique and recent work done in this field. Author addressed the issues and challenges in the field of association rule mining. She also compared the performance of various association rule mining based on their merits and demerits. **H. Bathla and K. Kathuria [12]** compared Apriori and Filter Associator algorithms on weka tool. Authors found that Apriori algorithm generates a large number of cycles and itemsets which degrades the performance. Lastly, they Concluded that Filter Associator is efficient algorithm than Apriori algorithm. **S. Vijayarani and R. Prasannalakshmi [13]**compared FP Tree and Apriori M/R (Map/Reduce)algorithms on Tanagr a tool. They used two performance metrics number of rules generated and execution time. Based on the experimental results authors found that the performance of FP-Tree Growth algorithm efficient than that of Apriori Map/Reduce (Apriori M/R)ARM algorithm. **S.O. Fageeri et al., [15]**compared the performance of ELACT and Binary-Based algorithm. Performance metric used are: Memory Usage and Execution Time. They found that Binary-Based algorithm has performed better than ELACT algorithm. **K. Vijayalakshmi et al., [18]**developed a prototype intelligent thyroid Prediction System by using big data and Information Mining strategies. This framework is capable of finding out hidden information from a thyroid database. They concluded that naive base classifier performed better in terms of execution time and accuracy. **M NAIR and F. KAYAALP [19]** used dataset of a company selling car maintenance and repair products in Turkey. **Memory usage**

and **execution times** against varying support values and varying record numbers are parameters used for comparison. **dEclat_bitset** algorithm is efficient for 6-months and 12-months dataset. **Eclat** algorithm did well for 0.7 and 0.3 support values. **dEclat_bitset** algorithm is the most efficient algorithm for 0.3 and 0.1 support values on 22-months dataset. **P. Devi and K.L. Bansal [21]**compared A-priori, A-priori Prefix Tree, Assoc Outlier, Frequent Itemsets, Supervised Association Rule and Supervised Association Tree using Tanagra Tool. Parameters used for performance comparison are Execution Time, Memory used and No. of Association Rules generated. Datasets used are Hypothyroid, chess game and Mushroom. Authors concluded that Apriori algorithm and Association Outlier algorithm take same execution time, Apriori take more memory, Apriori, Frequent Itemsets and Apriori Prefix Tree (Apriori PT) generates more rules than other algorithms. The Supervised Association Tree and Supervised Association Rule filter rules based on the discrete class value. Therefore, these two algorithms require less execution time, less memory space and generate few rules.**M. Kaushik et al. [22]**mentioned that there are variety of data attributes like numerical, categorical, Boolean but the traditional Association rule mining algorithms deal with binary attributes. To deal with varied data attributes we need NARM algorithms. Authors concluded that no single Numerical Association Rule Mining method is perfect for discovering patterns from real world datasets.

III. COMPARATIVE STUDY AND DISCUSSION

This study presents the complete analysis of ARM algorithms that are being used for conducting experiments.

Table1: Comparison of Different ARM Algorithms.

Author (s)	Problem Identified	Parameters Used	Technology Applied	Interpretation
K. Vanitha[2]	compared A-priori and FP-growth	Number of Instances, Confidence	Weka	FP- Growth is more Efficient and Scalable
L. Li, and M. Zhang [3]	Implemented the improved association rule mining algorithm	data set partition, data set distribution	MapReduce, Hadoop	improved algorithm is efficient in cloudcomputing environment.
S.B. Aher, and Lobo L.M.R.J. [6]	compared the result of Apriori, Predictive Apriori, Tertius and Filtered Associator	"yes" or "no" for course selection by the student	Weka	Apriori is performing better
M. Sharma et al., [7]	compared the performance of two algorithms namely: Apriori and Predictive Apriori	Mean, Median, Mode, Range, Std_Dev, Ave_Dev, Variance	Weka3.7.5	Predictive Apriori has performed better
K. Khurana, and S. Sharma [8]	comparison of Apriori, Apriori Tid, Apriori Hybrid and AIS, SETM	Data support, Speed,	———	speed and accuracy are improved in Apriori Hybrid
D. Kerana Hanirexet al., [14]	Implemented SETM algorithm and compared with Apriori Algorithm		Weka3.6.4	SETMprovides similar performance with the traditional Apriori
B. Nigam and P. Dalal [17]	compared the performance of Apriori, FP-Growth, Eclat and dEclat algorithms	memory usage, processing time, dataset and support	———	dEclat algorithm is faster

Table1 shows the theoretical Performance of various Association Rule Mining Algorithms achieved by the different researchers. It has been observed that most of the existing studies have used Apriori Algorithm and the experiments are being done on the Weka Tool.

It has been found that the performance of Apriori algorithm is not fully satisfiable as lot of limitations are there in Apriori Algorithms working. Some of the researchers have designed their Algorithms using the traditional Apriori Algorithms. But there is no such universally acceptable algorithm developed which overcomes all the limitations of Apriori algorithm. Also, we found out most of the researchers have used Weka tool for experimentations and less work is done in other data mining tool.

IV. CONCLUSION

This study has given the analysis and comparison of various association rule mining algorithms that are being used in various fields and their performance comparison with other algorithms. In this study we found out that Apriori algorithm and Weka tool are frequently being used for experiments. In many cases, FP growth algorithm perform better than the Apriori algorithm. Other than Weka Tool there are many other tools like Tanagra Tool which are providing many Inbuilt Association Rule Mining algorithms for different functionalities. Mostly the Parameters used for performance comparison are Execution Time and Memory used.

V. FUTURE SCOPE

In the future, we can evaluate the performance of different Association Rule Mining algorithms based on different parameter by using different data mining tools and also develop one algorithm which will overcome the limitations of all the existing algorithms Association Rule Mining.

REFERENCES

[1] B. Thuraisingham, Data mining: technologies, techniques, tools, and trends. CRC press, 2014.

[2] K. Vanitha, "EVALUATING THE PERFORMANCE OF ASSOCIATION RULE MINING ALGORITHMS", Journal of Global Research in Computer Science, vol. 2, pp.101-103, June 2011.

[3] L. Li, and M. Zhang, "The Strategy of Mining Association Rule Based on Cloud Computing", International Conference on Business Computing and Global Informatization, 2011, pp. 475-478 IEEE.

[4] F.H. AL-Zawaidah, and Y.H. Jbara, "An Improved Algorithm for Mining Association Rules in Large Databases", World of Computer Science and Information Technology Journal (WCSIT), Vol. 1, Issue 7, pp.311-316,2011.

[5] A.B.M. Rezbaul Islam, and T.S. Chung, "An Improved Frequent Pattern Tree Based Association Rule Mining Technique", International Conference on Information Science and Applications, 2011, pp. 1-8. IEEE.

[6] S.B. Aher, and Lobo L.M.R.J., "A Comparative Study of Association Rule Algorithms for Course Recommender System in E-learning", International Journal of Computer Applications, Vol. 39, Issue 1, February 2012.

[7] M. Sharma et al., 2012, "Evaluating the Performance of Apriori and Predictive Apriori Algorithm to Find New Association Rules Based on the Statistical Measures of Datasets", INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT), Vol. 01, Issue 06, August 2012.

[8] K. Khurana, and S. Sharma, "A Comparative Analysis of Association Rules Mining Algorithms", International Journal of Scientific and Research Publications, Vol. 3, Issue 5, May 2013.

[9] J. Nahar et al., "Association rule mining to detect factors which contribute to heart disease in males and females", Expert Systems with Applications, Vol. 40, Issue 4, pp.1086-1093,2013.

[10] T.A. Kumbhare, and S.V. Chobe, "An overview of association rule mining algorithms", International Journal of Computer Science and Information Technologies", Vol. 5, Issue 1, pp.927-930, 2014.

[11] G. Kaur, "Association Rule Mining: A Survey", International Journal of Computer Science and Information Technologies", Vol. 5, Issue 2, 2014.

[12] H. Bathla and K. Kathuria, "Association Rule Mining: Algorithms Used", International Journal of Computer Science and Mobile Computing, Vol. 4, Issue 6, pp. 271 – 277, June 2015.

[13] S. Vijayarani and R. Prasannalakshmi, "Association Rule Generation in Data Streams using FP-Growth and APRIORI MR Algorithms", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 9, pp.15-25, September 2015.

[14] D. Kerana Hanirex and K. P. Kaliyamurthie, "Mining Frequent Item Sets for Association Rule Mining in Relational Databases: An Implementation of SETM Algorithm Using Super Market Dataset", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 8, Issue 31, November 2015.

[15] S.O. Fageeri et al., "A performance analysis of association rule mining algorithms",3rd International Conference on Computer and Information Sciences (ICCOINS),2016, pp. 328-333.

[16] M. Shridhar and M. Parmar, "Survey on Association Rule Mining and Its Approaches", World Applied Sciences Journal, Vol. 5, Issue 3, Mar. 2017.

[17] B. Nigam and P. Dalal, "PERFORMANCE EVALUATION OF ASSOCIATION RULE MINING ALGORITHMS", INTERNATIONAL JOURNAL OF CURRENT ENGINEERING AND SCIENTIFIC RESEARCH (IJCESR), Vol. 4, Issue-7, 2017.

[18] K. Vijayalakshmi et al., "Intelligent Thyroid prediction system using Big data", International Journal of Computer Sciences and Engineering, Vol. 6, Issue 1, 31 Jan 2018.

[19] M. NAIR and F. KAYAALP, "Performance Comparison of Association Rule Algorithms with SPMF on Automotive Industry Data", Düzce University Journal of Science & Technology, 7 (2019) 1985-2000.

[20] P. Devi et al., "Association Rule Mining on Spambase Dataset using Tanagra", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-9, Issue-6, April 2020.

[21] P. Devi and K.L. Bansal, "Performance Comparison of Association Rule Mining algorithms on Different Datasets using Tanagra Tool", MuktShabd Journal, Volume-1X, Issue-VI, June 2020.

[22] M. Kaushik et al., "A Systematic Assessment of Numerical Association Rule Mining Methods", SN Computer Science, volume-2, Issue-5, pp. 1-13, 22 June 2021.

[23] T. Wang et al., "Student Behavior Data Analysis Based on Association Rule Mining", International Journal of Computational Intelligence Systems, volume-15, Issue-1, pp. 1-9, 19 May 2022.

[24] https://www.ibm.com/docs/en/db2/10.5?topic=SSEPGG_10.5.0/com.ibm.im.model.doc/c_lift_in_an_association_rule.html [accessed on 27-06-2022 1:00 PM]

[25] https://www.saedsayad.com/association_rules.htm [accessed on 27-06-2022 5:00 PM]