Comparative study of Apriori and FP-Growth Algorithm using WEKA tool

[1]Nitisha Yadav, [2]Palak Baraiya, [3]Nitika Goswami

Students
Computer Science
Acropolis Institute of technology and Research, Indore, India

Abstract— **Manually analyzing pattern for frequently bought item set is a cumbersome task. To solve this problem, many analytical data mining tools are available. Association rule mining is one of the data mining rules, which discovers interesting relation between variable from large dataset. Apriori & FP-Growth algorithm is the most common algorithm of association rule mining.**

This paper present the comparison of Apriori and FP-Growth algorithm on the basis of their execution time and memory space on supermarket dataset using WEKA tool. And the comparative result showed that the execution time of FP-Growth algorithm is much faster than Apriori algorithm & comes to the conclusion that FP-Growth is better for analyzing data quickly.

Index Terms—Data mining, Association rule, Apriori, FP-Growth, WEKA

I. INTRODUCTION

In today's world large amount of data has been found. And manually extracting useful information from that data becomes a complex task. Thus this mining of data can be done with data mining engine which comprises of modules or algorithms for performing mining tasks like Classification, Association Rule, and Clustering. Here we work on supermarket dataset for finding association rule among items with the help of Apriori and FP-Growth algorithm, and also practically differentiated both the algorithm on the basis of execution time by using WEKA tool.

II. ASSOCIATION RULE

They are used to show if-then relationship among various data items. Support (S) and Confidence (C) are normal methods used to generate association rules. Support means "Both together". Confidence means "If implies then". An example of an association rule is: "Confidence is 20% of ladies that purchase saries also purchase matching bangles and support is 5% of all ladies purchase both of these items".

III. APRIORI ALGORITHM

This algorithm is one of finest approaches to find frequent item sets from transaction database and derive association rules. It was proposed by R.Agrawal and R.Srikant in 1994 for mining the frequent item sets. If item sets are obtained, then they are used to generate association rule. The algorithm terminates when frequent item sets cannot be extended any more. But it has to generate a large amount of candidate item sets and scans the data set as many times as the length of the longest frequent item sets.

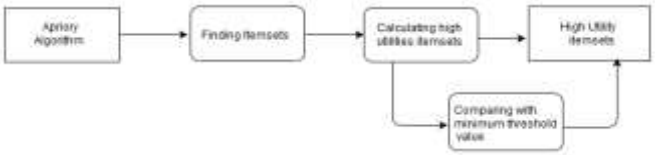

Fig. 1. Apriori Algorithm Block Diagram

IV. FP-GROWTH ALGORITHM

The most outstanding improvement over Apriori would be a method called FP-growth(frequent pattern growth). It adopts a divide and conquer strategy by compressing the database representing frequent items into a structure called FP-tree (frequent pattern tree). It scans the database only twice. In the first scan, all the frequent items and their support counts (frequencies) are derived and they are sorted in the order of descending support count in each transaction. In the second scan, items in each transaction are merged into a prefix tree and items (nodes) that appear in common in different transactions are counted.

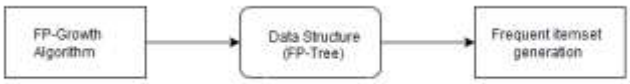

Fig. 2. FP-Growth Algorithm Block Diagram

V. SUPERMARKET DATASET

This dataset contains attributes i.e. number of different products in the shops and instances i.e. products which are bought together by the customer. Here we have taken five different datasets having same number of attributes i.e. 217 and different instances like 100, 1000, 5000, 10000, and 15000.

VI. WEKA TOOL

(Waikato Environment for Knowledge Analysis) is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. WEKA is free software available under the GNU General Public License. The WEKA workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality.

Advantages of WEKA include:

- It provides many different algorithms for data mining and machine learning
- It is open source and freely available
- It is platform-independent
- It is easily useable by people who are not data mining specialists
- It provides flexible facilities for scripting experiments
- It has kept up-to-date, with new algorithms being added as they appear in the research literature.

VII. IMPLEMENTATION IN WEKA

Here we have taken a small supermarket dataset having 5 instances. With the help of this example we saw the implementation step by step. Initially we have to save the notepad file with "arff" extension for opening in the WEKA tool.

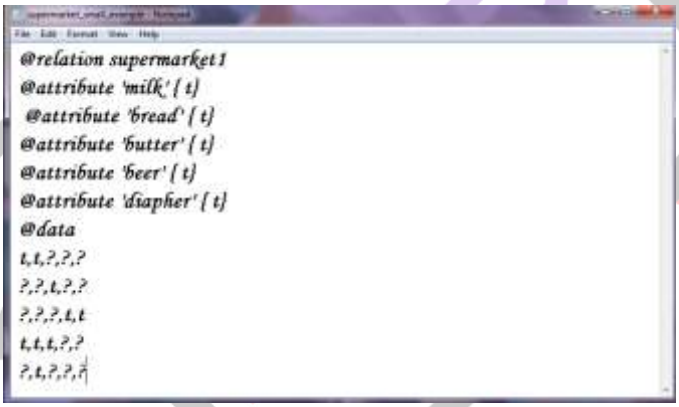

Fig. 3. Dataset in notpad with "supermarket.arff" extension

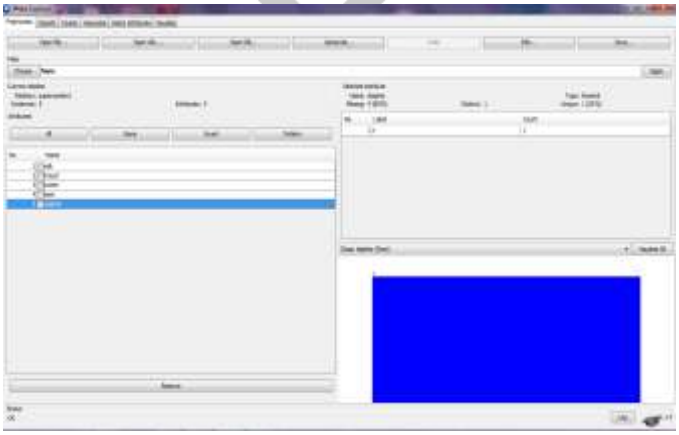

Fig. 4. Opening of .supermarket.arff file

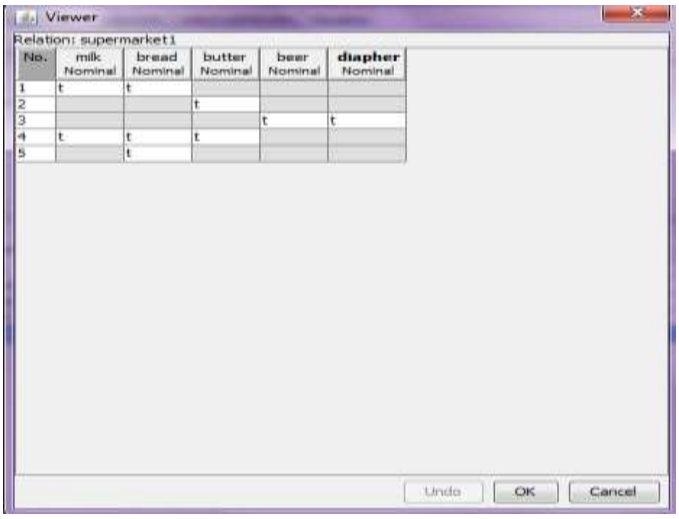

Fig. 5. Dataset tabular view in WEKA tool

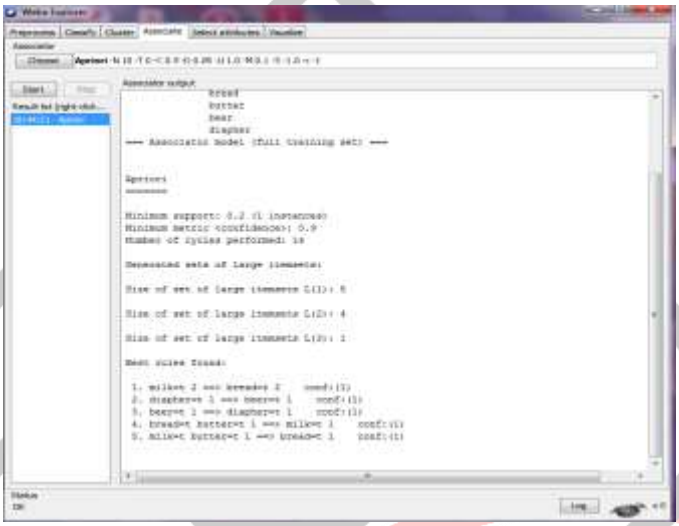

Fig. 5. Implementation of Apriori algorithm on supermarket dataset and its result

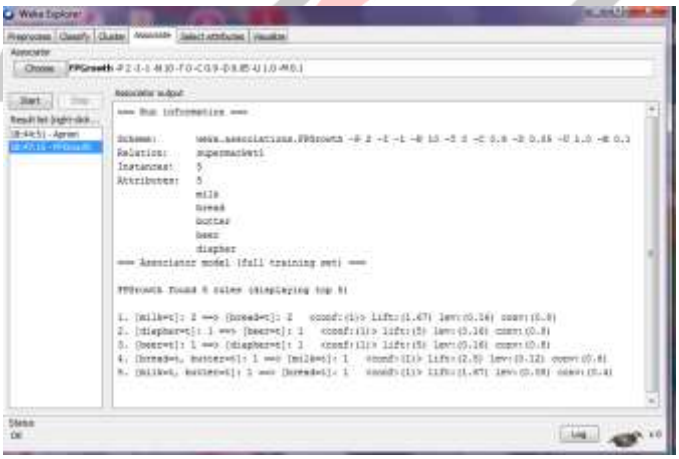

Fig. 6. Implementation of FP-Growth algorithm on supermarket dataset and its result

VIII. RESULT

Here we saw the performance of Apriori and FP-Growth algorithm varies from one dataset to another.

No. of Instance	Execution times(Secs.)	
	Apriori algorithm	FP-Growth algorithm
100	1	1
1000	11	2
5000	92	3
10000	95	3
15000	398	4

Fig. 7. Execution time in seconds for different datasets

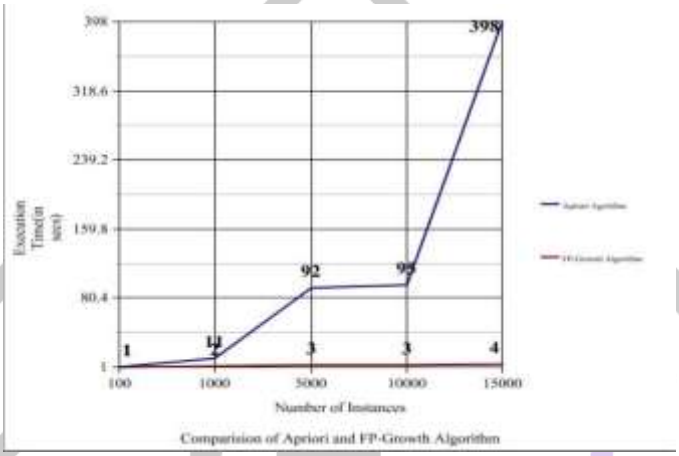

Fig. 8. Comparision graph of Apriori and FP-Growth algorithm

IX. CONCLUSION

In this research paper we have taken 5 different datasets of supermarket with instances 100, 1000, 5000, 10000, & 15000. And with the help of WEKA tool, we practically differentiated the performance of Apriori and FP-Growth algorithm on the basis of execution time. And the result comes in the favor of FP-Growth algorithm, which is faster than Apriori

X. ACKNOWLEDGMENT

It is our pleasant duty to acknowledge all those people who have helped us in the successful completion of the research paper. We extend our heartfelt gratitude to our guide, **Ms. Vandana Kate, Asst. Professor,** whose profound knowledge and perpetual encouragement provided impetus to the progress of the paper.

We also thank our coordinator, **Ms. Kavita Namdeo, Acropolis Institute of Technology and Research (Indore),** for the guidance they gave us during the complete research work duration.

We are also thankful to the all faculty members of computer science and Engineering department, Acropolis Institute of Technology and Research (Indore), friends and family for providing all the necessary facilities required to complete the paper.

References

[1] Gagandeep Kaur, Shruti Aggarwal "Performance Analysis of Association Rule Mining Algorithms" Volume 3, Issue 8, August 2013 "International Journal of Computer Science and Engineering (IJCSE)".

[2] http://www.iasri.res.in/ebook/win_school_aa/notes/WEKA.pdf

[3] Applications (IJCA), ETCSIT- Number 3", 2012.Vishal Jain, Gagandeep Singh Narula & Mayank Singh, ―Implementation Of Data Mining In Online Shopping System Using Tanagra Tool‖, International Journal of Computer Science and Engineering (IJCSE) , ISSN 2278-9960 ,Vol. 2, Issue 1, Feb 2013, 47-58.

[4] Ferenc Bodon, "A fast Apriori implementation", "In Proceedings of the IEEE ICDM Workshop on Frequent Item

[5] set Mining Implementations", 2003.Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[6] Dhanashree S. Deshpande, "A Survey on Web Data Mining Applications", "International Journal of Computer Science and Engineering (IJCSE)".
[7] Han J. and M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann publishers, 2nd Edition.
[8] http://www.cs.ccsu.edu/~markov/weka-tutorial.pdf
[9] http://en.wikipedia.org/wiki/Data_mining
[10] http://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Frequent_Pattern_Mining/The_FP-Growth_Algorithm#FP-Tree_structure
[11] http://www.information-drivers.com/market_basket_analysis.php
[12] Venkatadri.M and Dr. Lokanatha C. Reddy, "A Review on Data mining from Past to the Future", International Journal of Computer Applications, Volume 15– No.7, pp. 19-22, February 2011.