J. Sci. Trans. Environ. Technov. 2023, 17 (2) : 25 - 30

Market Basket Analysis: An Overview ofAssociation rule Mining

Priyadharshini M K and Mahadevi M

[1] Student, PG Department of Computer Science Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women, Chrompet, Chennai-44. email : priyadharshini19cs038@gmail.com
[2] Associate Professor, PG Department of Computer Science Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women, Chrompet, Chennai-44. email : drmahadevi2017@gmail.com

Article History
Received: **06.10.2023**
Revised and Accepted: **10.11.2023**
Published: **16.12.2023**

https://doi.org/10.56343/STET.116.017.002.004
www.stetjournals.com

ABSTRACT

Lack of awareness about blood donation, this paper provides a comprehensive overview of association rule mining, focusing on its application in market basket analysis. It discusses foundational approaches like the Apriori and FP-Growth algorithms, as well as alternative methods such as FPP, Eclat, FP Max etc., for efficient and profitable itemset discovery. This review also highlights the challenges and limitations of market basket analysis, emphasising the need for scalable, efficient, and adaptable approaches to handle evolving data trends.

Keywords: association rule learning, Eclat, Item set discovery, FPP, FP Max, market basket analysis

1. INTRODUCTION

Association rules and frequent itemsets are fundamental concepts in data mining and are primarily used for discovering interesting relationships or patterns within large datasets. They are widely employed in various applications, such as market basket analysis, recommendation systems, and healthcare data analysis.

Market basket analysis, a subset of association rule mining, plays a pivotal role in understanding consumer behaviour, enhancing business strategies, and optimising inventory management. This paper offers a comprehensive view of association rules by examining the types, limitations, and emerging challenges for market basket analysis.

Priyadharshini M K
Student, PG Department of Computer Science Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women, Chrompet, Chennai-44.
email : priyadharshini19cs038@gmail.com

2. Association rule Mining

Association rule mining is a data mining technique used to discover interesting relationships and patterns within large datasets. It focuses on identifying associations between items in transactions, revealing which items tend to be purchased together. Association rule mininghas numerous applications, such as market basket analysis, recommendation systems, and understanding customer behaviour in retail and e-commerce. It helps businesses make informed decisions, optimise product placement, and enhance customer experiences by uncovering meaningful associations and patterns in their data. It consists of two main components: antecedent (premise) and consequent (outcome) (Han *et al*., 2020). These rules are written as:

"If {antecedent} then {consequent}" (Han *et al*., 2020)

P-ISSN 0973-9157
E-ISSN 2393-9249

Association rules can take various forms, depending on the number of items in the antecedent(premise) and consequent (outcome), as well as the data types involved. The significance of association rules can be determined by the following metrics

- **Support:** Measures itemset frequency.
- **Confidence:** Gauges rule strength.
- **Lift:** Compares observed and expected support.
- **Conviction:** Assesses rule dependency.
- **Leverage:** Quantifies item co-occurrence difference.

2.1 Foundational approach

2.1.1 Apriori Algorithm

Agrawal and Srikant's Study on Market Basket Analysis (1994) introduced the Apriori algorithm, a pivotal method for efficiently discovering association rules in large transactional datasets. It uses a **"bottom-up" approach**, starting with frequent individual items and gradually expanding to larger itemsets. The key idea is the Apriori principle, which states that if an itemset is infrequent, then its supersets are also infrequent. Their work established the concepts of **support and confidence** to measure rule significance.

A study conducted by Kawle and Dahima (2018), utilising data from the Instacart shopping website comprising **3 million records**, has yielded valuable insights into frequent itemsets, featuring **multiple support and confidence values (Figs. 1 & 2)**. This study also incorporated diverse **visualisations**, enabling retailers to enhance their sales strategies based on the analysis.

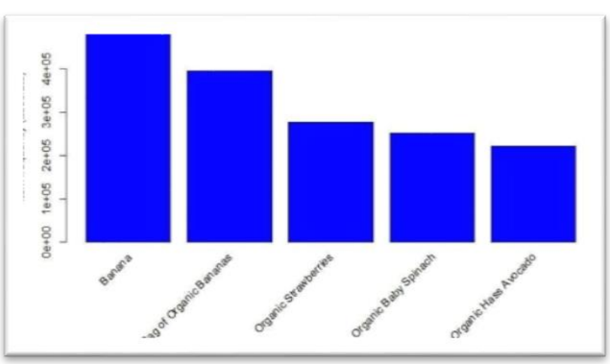

Figure 1. Frequent itemsets histogram

2.1.1 FP-Growth algorithm

The FP-Growth algorithm, introduced by **Han *et.al* in 2000** as an **alternative to the Apriori Algorithm,** offers a distinct approach. **It builds an FP-tree, eliminating the necessity of generating candidate itemsets**. This key difference contributes to its efficiency advantage over Apriori, as it recursively extracts frequent itemsets directly from the FP-tree.

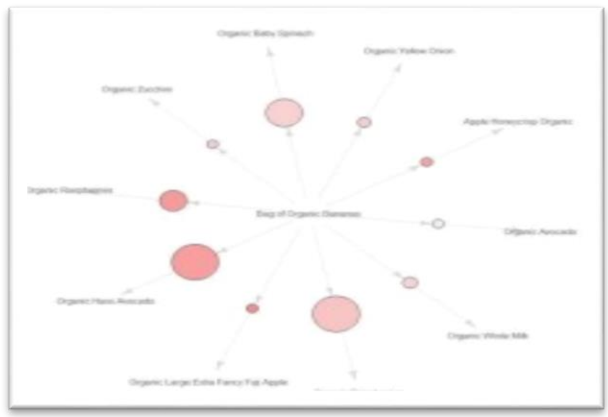

Figure 2. Network graph visualisation

A study that employed supermarket data (Kaur and Kang, 2014, Ignatius *et al*., 2022), consisting of 7501 records, has provided valuable insights regarding the runtime comparison between the Apriori and FP-Growth algorithms. The study highlights that the creation of the FP-tree is significantly swiftercompared to the generation of candidate sets in Apriori, resulting in FP-Growth being **five times faster than the Apriori Algorithm in terms of runtime (Fig.3)**

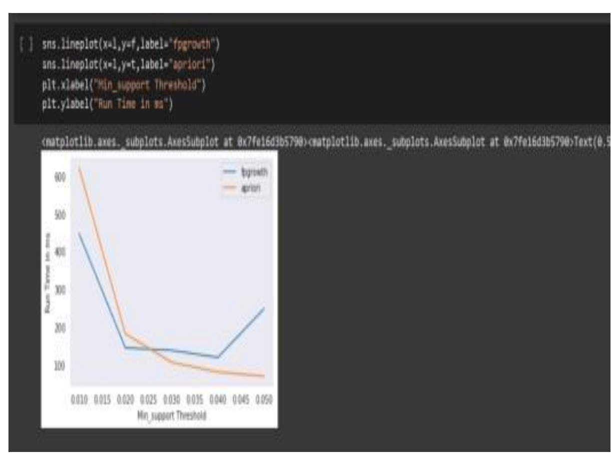

Fig-3. Run time Comparison between Apriori and FP Growth Algorithm [4]

2.2 Other Approaches

2.2.1 Scalable Algorithm for Market Basket Analysis

A study by Kaur and Kang, (2016) addressing the identification of evolving trends in market data provides an overview of several association rules mining algorithms. **It points out that current algorithms are designed for static data** but introduces a novel approach. **This new method concentrates on mining association rules within dynamic data by conducting periodic analyses.** The approach is demonstrated using an extended bakery dataset stored in 4 windows(contains association rules for particular time periods), which includes 2000 transactions and 26 items in each window, and can be scaled to accommodate a variable number of items.

A research study investigating the scalability of Market Basket Analysis (Cavique 2000) has indicated that certain algorithms exhibit **inefficiencies in terms of computational time.** This study primarily concentrates on converting input data into **maximum weighted cliques and constructing frequent itemsets through a meta-heuristic approach.** The findings from thisstudy unveil extensive patterns of large itemsets with favourable scalability characteristics.

2.2.2 Efficient algorithm for Market Basket Analysis

A study conducted by Shrivastava and Sahu (2017) highlights that both the **Apriori and FP-Growth algorithms, commonly used for mining frequent itemsets, lack efficiency when it comes to aiding market analysts in making decisions about supermarket shelf space planning.** The research introduces an alternative method for discovering profitable frequent itemsets by utilising two datasets: transaction and profit datasets (Figs. 4 & 5). This novel approach involves the generation of an FPP (Frequent Profit Pattern)-tree, which successfully identifies the frequent profitable patterns {f, c, a, m, p}. In contrast, the FP-Growth and Apriori Algorithms only identified the frequent patterns {f, c, a} within the same dataset (Fig.6).

TID	item1	item2	item3	item4	item5	item6	item7	item8
1	f	a	c	d	g	i	m	p
2	a	b	c	f	l	m	o	
3	b	f	h	j	o	w		
4	b	c	k	s	p			
5	a	f	c	e	l	p	m	n

Figure 4. Transaction data

itm_name	itm_profit
A	4
B	3
C	5
D	2
E	4
F	3
G	3
H	2
I	5
J	3
K	4
L	6
m	3
N	2
O	5
P	3
S	2

Figure 5. Profit data

Algorithm	Apriori	FP-Tree	FTTP
Results	*f, c, a*	*f, c, a*	*f, c, a, m, p*

Figure 6 Comparison of Apriori, FP-TREE and FPP.

2.2.3 Rapid Association Rule Mining (RARM):

A novel approach in rapid association rule mining (Das *et al*., 2001) seeks to expedite the generation of large 1-itemsets and 2-itemsets without the need for recursive dataset scanning. This approach emphasises that accelerating the candidate generation process for 2-itemsets addresses a critical bottleneck encountered in the Apriori Algorithm.

2.2.4 Eclat Algorithm:

The Eclat (Equivalence Class Transformation) algorithm, introduced by Zaki in 2000, employs a methodology centred on equivalence classes and vertical data representation. It uses a depth-first search strategy to efficiently mine frequent itemsetswithout candidate generation, instead focusing on intersecting transaction IDs of itemsets with shared items. This approach significantly reduces computational complexity, making it particularly effective for sparse datasets where items occur infrequently.

2.2.5. FP Max Algorithm:

The FPMax algorithm, an extension of the FP-Growth algorithm (Li *et al*., 2008), is designed to identify maximal frequent item sets. These maximal itemsets are characterised by their inability to be further extended while still retaining their status as frequent. This research approach is particularly valuable when the aim is to streamline the rule generation process by eliminating itemsets that cannot be expanded to form additional frequent itemsets. In essence, FPMax helps reduce the redundancy in generated rules, making it a useful tool for more focused and efficient association rule mining, especially when dealing with large datasets.

A comparative analysis of association rule mining algorithm is given in Table 1.

SI.No.	Algorithms	Function	Merits	Drawbacks
1	Apriori Algorithm	Generates frequent itemsets using candidate sets	Simplest method to implement. A large group of frequent items can be obtained.	Scans data multipletimes. Computational timeand cost is comparatively high.
2	FP-Growth Algorithm	Generates FP-tree without the use of candidate sets.	Faster than Apriori.	Tree data structure increases complexity.
3	FPP Algorithm	Identifies profitable frequent itemsets	Profitable patterns allow retailers to implement new sales strategies.	Generation of profit tables may be complex.
4	RARM	Faster generation of candidate sets with 2-itemsets	Performance is relatively high compared to Apriori Algorithm	Rapid mining maylead to false associations.
5	Eclat Algorithm	It is used for mining sparse datasets	Uses depth-first search (vertical data mining). Reduces computational complexity.	Fails when candidate sets arelarge.
6	FP Max Algorithm	Finds maximal frequent itemsets.	More efficient when dealing with large datasets. Avoids redundancy in rules generation.	Frequent sets obtain only maximal itemsets and cannotbe extended to other frequencies.

Table-1. Analysis of Association Rule Mining Algorithms

3. Challenges and Limitations:

Market basket analysis, while a powerful tool for uncovering associations and patterns in transactional data, comes with its own set of challenges. These challenges can impact the effectiveness and accuracy of the analysis. Here are some common challenges in market basket analysis:

1. **Scalability:** Handling extensive datasets, particularly with algorithms like Apriori (Cavique *et al.*, 2007), can be computationally demanding. As data trends evolve dynamically, adapting to these changes requires the use of dynamic algorithms to accommodate growing data volumes.

2. **Efficiency:** Merely analysing frequent itemsets may not provide sufficient insights forretailers to make informed decisions about customer purchases. Factors like Profit Pattern Growth (Srivastava and Sahu, 2007)

and climate-related purchase patterns must also be considered to enhance sales strategies effectively.

3. **Computational Complexities:** Complex rules pose challenges in terms of interpretation and can also extend computational time. Notably, when comparingApriori and FP-Growth algorithms, generating candidate itemsets in Apriori demands significantly more computational time than the FP-Growth algorithm (Li et al., 2008).

4. **Threshold Selection:** Establishing appropriate support and confidence thresholds poses a challenge. Setting thresholds too low can yield an overwhelming number of rules, while setting them too high may result in an insufficient quantity of meaningful rules.

5. **Algorithm Selection:** Selecting the most suitable algorithm for a specific dataset and problem can be a daunting task. There isn't a universal solution, and the choice of algorithm hinges on the dataset's attributes and the analysis objectives.

4. Conclusion

Market basket analysis, powered by association rule mining, remains indispensable for understanding consumer behaviour and optimising business strategies. Businesses must carefully select algorithms tailored to their datasets and objectives. Despite challenges, these techniques empower companies to glean valuable insights, fostering informed decision-making in our ever-evolving marketplace. Staying adaptive and data-centric is keyto continued success in the dynamic world of retail and e-commerce.

References

Agrawal, R., & Srikant, R. 1994. Fast algorithms for mining association rules in large databases. *In:* Proceedings of the 20th VLDB Conference Santiago, Chile, 1994. (pp. 487-499).

Cavique, L. 2007. A scalable algorithm for the market basket analysis. *Journal of Retailing and Consumer Services*,14(6),400-407. https://doi.org/10.1016/j.jretconser.2007.02.003

Das, Amitabha,Ng, Wee Keong and Woon, Yew-Kwong. 2001. Rapid Association Rule Mining.. *In*: Proc. Tenth International Conference on Information and Knowledge Mabagement.474-481. https://doi.org/10.1145/502585.502665

Girotra, M., Nagpal, K. C., Minocha, S., and Sharma, N. 2013. Comparative Survey on Association rule mining Algorithms. *International Journal of Computer Applications*, 84(10), 18-22 https://doi.org/10.5120/14612-2862

Han, J., Pei, J., and Yin, Y. 2000. Mining frequent patterns without candidate generation. *In:* ACM SIGMOD Record, 29,(2): 1-12. https://doi.org/10.1145/335191.335372

Ignatius, Cecil, Hermina, B.Gopalakrishnan, and Aishwaryalakshmi.B. 2022. Marlet Basket Analysis FOR a supermarket. *International Journal of Management, Technology and Engineering,* 12(11), 106-113. \https://www.researchgate.net/publication/365489098_MARKET_BASKET_ANALYSIS_FOR_A_SUPERMARKET

Kaur, M., & Kang, S. 2016. Market Basket Analysis: Identify the changing trends of market data using association rule mining. *Procedia Computer Science,* 85, 78-85. https://doi.org/10.1016/j.procs.2016.05.180

Kawale, Nidhi & Dahima, Snehil. (2018). Market Basket Analysis using Apriori Algorithm in R Language. *International Journal of Trend in Scientific Research and Development*. Volume-2. 2628-2633. 10.31142/ijtsrd15677.. https://doi.org/10.31142/ijtsrd15677

Kumbhare, T. A., and Chobe, S. V. 2014. An overview of association rule mining algorithms. *International Journal of*

Computer Science and Information Technologies, 5(1), 927-930.

Li, J., Pei, J., and Han, J. 2008. CMAR: Accurate and efficient classification based on multiple class-association rules. *In:* Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM) (pp. 369-376).

Savi Gupta, and Roopal Mamtora. 2014. A Survey on Association Rule Mining in Market Basket Analysis. *International Journal of Information and Computation Technology*., 4(4), pp. 409-414 .

Shrivastava, A., and Sahu, R. 2007. Efficient association rule mining for market basket analysis. *Global Journal of e-Business & Knowledge Management*, 3(1), 21-25.

Zaki, M. J. 2000. Scalable algorithms for association mining. IEEE *Transactions on Knowledge and Data Engineering*, 12(3), 372-390. https://doi.org/10.1109/69.846291