Association Rule Mining for Market Basket Analysis on Online Retail Store

Deepak Kumar Chauhan

Asst. Professor, School of Computing, Graphic Era Hill University, Dehradun, Uttarakhand
India 248002

Abstract

Due to the rapid emergence and evolution of e-Commerce, many retailers are now looking for new ways to improve their customer's purchasing behavior. One of these is through the use of Market Basket Analysis (MBA), a data mining tool that allows them to identify hidden relationships among various products. This method can help them make informed decisions when it comes to marketing and sales. The use of association rule mining (ARM) is a powerful technique that can help retailers identify and analyze the relationships among various products in their large transaction databases. With the help of Market Basket Analysis, they can then devise effective marketing and sales strategies. The goal of this study is to analyze the buying behavior of consumers in an e-Commerce store using ARM. It will then use the collected data to improve the marketing strategies of the store. Through the use of FP Growth and Apriori, we were able to create association rules that are based on the data. The findings of this study provide a deeper understanding of the buying habits of consumers in e-Commerce stores. They also offer practical recommendations for optimizing the sales process and product placement.

Keywords: Association rule mining, Data mining, Market Basket Analysis (MBA), online retail.

Introduction

Due to the rise of online retail stores, there has been a massive amount of data collected by these sites. This data contains various information about their customers, such as their purchase history and product preferences. It is hard for retailers to extract actionable insights from this data. This is where the association rule mining (ACM) industry comes in[1]–[3]. Through association rule mining, a process that involves identifying patterns and relationships in large datasets, retailers can get a deeper understanding of their customers' purchasing habits. For instance, if a customer purchases a printer, they are more likely to buy paper and ink cartridges.

The two main steps in association rule mining are the generation and generation of rules. The first step involves identifying the groups of products that frequently occur in the same transactions. The second step involves identifying the relationships among the products. The rise of the association rule mining process is attributed to the massive amount of data that is collected and analyzed by retailers. Due to the increasing number of transactions, it has become harder for them to find meaningful insights in this data. This process allows them to identify patterns and relationships in the data[4].

The amount of data collected by online retailers is immense. It can contain valuable information about their customers such as

their purchase history and product preferences. Unfortunately, it is hard for retailers to find meaningful insights from this data due to its complexity. Through association rule mining, retailers can identify the relationships among products that frequently occur in the same transactions. This process can then be used to improve their marketing strategies and product recommendations. The association rule mining process can help retailers identify the products that frequently combine and form bundles. This information can then be used to create targeted offers and promotions. It can also help them recommend products based on the browsing history of their potential customers[5], [6].

Aside from improving their operations, association mining can also help them understand the buying habits of their customers. This process can be used to analyze the data to identify patterns and insights in the customer behavior. In online retail stores, the significance of mining association rules lies in its ability to deliver actionable insights that can be utilized to enhance business operations and improve customer loyalty and satisfaction.

The objective of this study is to explore the association rule mining application in online retailers for the analysis of market basket data. The study will look into the relationships among products and the various transactional details of an online store.

This study aims to provide a comprehensive analysis of the association rules generated by the frequent itemsets in online retailers. It will then come up with recommendations that can help improve the operations of the business. The study will look into the frequent itemsets and identify the association rules that can be generated from this data. It will then analyze the significance of these rules and provide recommendations to improve the operations of online retailers.

The study will also look into the various challenges that are associated with the mining of association rules in online retailers. The findings of this study will be used to contribute to the increasing number of studies on the use of association rules in e-commerce stores.

Literature Review

Due to its ability to identify hidden connections among products in large datasets, such as e-commerce and retail transactions, Market Basket Analysis has gained widespread attention. It is a technique that enables businesses to improve their customer experience and increase their sales by analyzing the buying habits of consumers. The table-1 below presents a literature review of eleven papers that deal with various aspects of MBA. These include the use of FP growth algorithms, association rule mining, and minimum spanning trees. It aims to provide an extensive analysis of current research in the area and identify potential future directions.

Table 1 Major related work

Author(s)	Field of Research	Methodology Used	Algorithm Used	Output	Results
H. K. Kim et al.[7]	Market Basket Analysis (MBA)	Product network analysis	Not specified	Product network for extending MBA	Improved MBA results
Raorane et al.[8]	MBA & Association Rule	Data mining	Not specified	Knowledge extraction from MBA	Generated insights from MBA

M. Hemalatha [9]	MBA in Indian retailing	Data mining & case study	Not specified	Insights in Indian retailing	Application of MBA in Indian retailing
H. Aguinis et al.[10]	MBA in management research	Review of MBA applications in management	Not specified	Improved understanding of MBA in management	MBA applications in management
S. O. Abdulsalam et al.[11]	MBA & Association Rule	Data mining	Association Rule Mining	Analysis of market basket transactions	Identified patterns & relationships
W. Nengsih[12]	MBA & Apriori Association	Comparative study	Apriori & other association techniques	Comparison of MBA and Apriori Association	Evaluated performance of algorithms
A. Setiawan et al.[13]	MBA in sales information system	Data mining	Market Basket Analysis	Sales insights for stationery company	Improved sales information system
M. A. Valle et al.[14]	MBA with Minimum Spanning Trees (MST)	MST application in MBA	MST algorithm	Enhanced MBA with MST	Complementary insights using MST
T. Kutuzova et al.[15]	MBA & Recommendation system improvement	Analysis of heterogeneous data sources	Not specified	Improved recommendation system	Better recommendations for users
R. Moodley et al.[16]	MBA & Uninorms	Application of uninorms to MBA	Uninorms	Enhanced MBA with uninorms	Advanced analysis of MBA
M. Hossain et al.[17]	MBA & Apriori and FP Growth algorithms	Data mining	Apriori & FP Growth algorithms	MBA using Apriori and FP Growth algorithms	Improved efficiency & performance

The literature review indicates that researchers have developed various algorithms and methodologies to improve the Market Basket Analysis process, which is applicable in various sectors. The studies examined how these techniques can be used to enhance the efficiency of MBA, as well as how they can be integrated into sales and management research systems. The studies conducted on the use of MBA in various fields, such as retailing and management research, have shown that it can help improve decision-making and generate valuable insights. They also helped practitioners and researchers identify the most suitable algorithms for their specific needs. Despite the progress that has been made in the development of MBA, there is

still a lot of room for further research to improve its efficiency and effectiveness. In addition to developing new algorithms, studies on the integration of data sources such as social media and content will also be conducted.

Association Rule Mining Techniques

Data mining techniques are used to find patterns and relationships in large datasets. One of these is association rule mining, which involves identifying frequent itemsets. These groups of items often appear together in numerous transactions. After identifying these frequent itemsets, association rules are then generated based on the data. Several techniques, such as ECLAT, Apriori, and FP-Growth, can be utilized to analyze association rules. The Apriori algorithm is widely used in association rule mining. It involves starting with small items and gradually increasing to large onesets.[18]

Another popular technique is the FP-Growth algorithm, which takes advantage of a tree structure to find frequent itemsets. The ECLAT method is a vertical approach, which takes into account the intersection of multiple items. It can then generate association rules by identifying the intersection of these two groups of items. The performance of these algorithms is different in terms of their scalability, efficiency, and ability to process sparse data. The appropriate algorithm should be chosen depending on the size, complexity, and computational resources of the data.

Market Basket Analysis in Online Retail Store

A market basket analysis is a type of association rule mining that takes place in online retailers. It aims to identify the patterns of purchase that occur between various products. Through market basket analysis, online retailers can improve their customer's experience and create customized bargains. For instance, they can identify which products are commonly purchased together and offer deals based on

their browsing history. Similar to association rule mining, market basket analysis involves gathering transactional data about the products and the buyers. This includes the name, SKU, and the transaction ID[19]–[21].

A market basket analysis can be beneficial for online retailers as it allows them to gain a deeper understanding their customers' behavior and preferences. It can also help them improve their marketing and product offerings. A market basket analysis lets online retailers gain a deeper understanding of their customers' preferences and behavior. It can also improve their operations and increase their customer loyalty and satisfaction.

Applications of Association Rule Mining in Online Retail Stores

There are many applications of association rule mining in online retailers.

1. Product Recommendations: Through association rule mining, retailers can recommend products to their customers based on their browsing history and past purchases. They can also identify patterns of overlap between the products and suggest substitutes or complementary products.
2. Market Basket Analysis: One of the most common applications of this type of association rule mining is market basket analysis. This process involves identifying the patterns of purchase that occur between multiple products.
3. Inventory Management: Rule mining can help optimize the inventory management process by identifying products that are frequently bought together. It can also help retailers stock up on complementary products.
4. Customer Segmentation: Rule mining can also be used for the creation of promotional campaigns

and marketing strategies based on the buying habits of consumers.

5. Fraud Detection: An association rule mining application can help identify fraudulent activities by gathering information about unusual purchases.

In addition to fraud detection, association rule mining can also be used in online retailers to improve product recommendations and inventory management.

Challenges of Association Rule Mining in Online Retail Stores

Although association rule mining provides many advantages to online retailers, it also comes with several challenges.

1. Data Quality: One of the most common issues that association rule mining can encounter is the quality of data. This can be caused by the lack of data formats and values.
2. Data Preprocessing: One of the most common challenges associated with association rule mining is the preprocessing of large datasets. This process can be very time-consuming and challenging.
3. Scalability: One of the biggest challenges associated with association rule mining is its high cost and limited scalability.
4. Interpretation: Due to the complexity of the association rules generated by association rule mining, it can be hard for retailers to make informed decisions.
5. Overfitting: One of the most common issues that association rule mining can encounter is overfitting. This occurs when the model gets too close to the data training and doesn't sufficiently generalize well to new information.

Due to the complexity of the association rule mining process, it is not always feasible to provide meaningful insight into the data collected in online retailers.

However, with the right tools and techniques, it can be used to improve the efficiency of the process.

Methodology

The paper's methodology section provides an overview of the various steps involved in the development of a research strategy. For instance, in the case of market basket analysis, the association rule mining method is used to analyze the data collected from online retailers.

1. Data Collection:
   The data collection process begins with gathering information about a retailer's online store[19]. This step involves collecting details about the products purchased and the customer ID. The information should be collected for a long time to ensure that the analysis can produce meaningful results.

2. Preprocessing and Cleaning of Data:
   The cleaning and preprocessing of the collected data are the next steps in the process. They are important to ensure that the analysis is performed accurately and thoroughly. There are various methods that can be used to thoroughly clean and preprocess the data.

a. Removing duplicates: In some cases, the data may contain multiple entries for a single transaction, which should be removed to prevent the analysis from skewing.
b. Handling missing values: When there are missing values in the data, they should be handled properly. This can be done by replacing the missing ones with the most frequently used item.
c. Data normalization: Normalization is a process that involves scaling the data to ensure that it falls within a

certain range. Doing so helps ensure that all the items are analyzed fairly.

3. Association Rule Mining Algorithm:
The next step in the process is to choose the appropriate algorithm for mining association rules. The most common algorithm used is the Apriori. This process involves generating association rules and frequent itemsets through a bottom-up approach.

4. Performance Metrics for Evaluation:
The evaluation phase of the process is when the performance metrics are selected for the analysis. These are usually used to measure the effectiveness of the association rule mining process. Some of these include the support, confidence, and conviction categories. The confidence level in the transactions that contain the antecedent and consequent is referred to as confidence. The lift level is calculated by comparing the expected support with the observed support. Finally, the conviction level is the degree of dependence between the antecedents and the consequents.

The process for mining association rules in online retailers' market basket is generally performed in three phases. It involves preprocessing the collected data, selecting an appropriate mining algorithm, and performing evaluation of the results. This method can help retailers identify areas of their operations where they can improve.

Result and Outputs

Table 2 Result with various evaluation metrices

Association Rule	Support	Confidence	Lift	Conviction	Interest
{StockCode=22423, Country=United Kingdom} => {StockCode=22697}	0.0026	0.98	4.1	51.23	0.0018
{StockCode=22726, Country=United Kingdom} => {StockCode=22727}	0.0021	0.88	9.1	7.63	0.0019
{StockCode=22728, Country=United Kingdom} => {StockCode=22727}	0.0018	0.85	8.8	6.24	0.0016
{StockCode=22699, Country=United Kingdom} => {StockCode=22697}	0.002	0.83	3.5	3.87	0.0014
{StockCode=22697, Country=United Kingdom} => {StockCode=22699}	0.002	0.08	3.5	1.06	0.0006
{StockCode=22727, Country=United Kingdom} => {StockCode=22726}	0.0021	0.22	9.1	1.25	0.0019

The association rules discussed in table-2 provide valuable insights into the purchasing behavior of customers from the United Kingdom in an online retail store. These rules, which involve specific StockCodes, have been evaluated based on measures such as support, confidence, lift, conviction, and interest. The high values for confidence, lift, and conviction across the rules demonstrate strong associations and dependencies between the items involved. These findings can help online retailers optimize their marketing strategies, product placement, and recommendation systems, ultimately leading to improved customer satisfaction and loyalty. By understanding and leveraging these patterns, businesses can make data-driven decisions to enhance their competitiveness in the ever-evolving digital marketplace.

Conclusion

The study successfully utilized the Association Rule Mining approach to analyze the market basket in an online retailer. The two algorithms used in the study, namely the FP Growth and Apriori, were evaluated to identify their strengths and weaknesses and provide guidance to practitioners. Through the study, the researchers were able to gain a deeper understanding of the buying habits of consumers, which can be used by online retailers to improve their marketing and product placement. They were also able to identify the factors that influence their loyalty and satisfaction. This method can help them make informed decisions and improve the customer experience. The findings of the study provide valuable information for future research regarding the use of association rule mining and market basket analysis in online retail. It also serves as a foundation for future studies that will look into the various aspects of the customer experience in such environments. The study highlights the potential of market basket analysis and association rule mining to help online retailers make informed decisions and improve their customer experience. It can also help them stay competitive in an evolving digital environment.

References

[1] B. Kamsu-Foguem, F. Rigal, and F. Mauget, "Mining association rules for the quality improvement of the production process," *Expert Syst. Appl.*, vol. 40, no. 4, pp. 1034–1045, 2013, doi: 10.1016/j.eswa.2012.08.039.

[2] A. A. Kardan and M. Ebrahimi, "A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups," *Inf. Sci. (Ny).*, vol. 219, pp. 93–110, 2013, doi: 10.1016/j.ins.2012.07.011.

[3] I. Cil, "Consumption universes based supermarket layout through association rule mining and multidimensional scaling," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 8611–8625, 2012, doi: 10.1016/j.eswa.2012.01.192.

[4] S. Gupta and R. Mamtora, "A Survey on Association Rule Mining in Market Basket Analysis," *Int. J. Inf. Comput. Technol.*, vol. 4, no. 4, pp. 409–414, 2014, [Online]. Available: http://www.irphouse.com/ijict.htm.

[5] I. F. Videla-Cavieres and S. A. Ríos, "Extending market basket analysis with graph mining techniques: A real case," *Expert Syst. Appl.*, vol. 41, no. 4 PART 2, pp. 1928–1936, 2014, doi: 10.1016/j.eswa.2013.08.088.

[6] M. Vahidi Roodpishi and R. Aghajan Nashtaei, "Market basket analysis in insurance industry," *Manag. Sci. Lett.*, vol. 5, no. 4, pp. 393–400, 2015, doi: 10.5267/j.msl.2015.2.004.

[7] H. K. Kim, J. K. Kim, and Q. Y. Chen, "A product network analysis for extending the market basket analysis," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7403–7410, 2012, doi: 10.1016/j.eswa.2012.01.066.

[8] Raorane AA, Kulkarni RV, and Jitkar BD, "Association Rule-Extracting Knowledge Using Market Basket Analysis," *Res. J. Recent Sci.*, vol. 1, no. 2, pp. 19–27, 2012, [Online]. Available: www.isca.in.

[9] M. Hemalatha, "Market basket analysis - A data mining application in Indian retailing," *Int. J. Bus. Inf. Syst.*, vol. 10, no. 1, pp. 109–129, 2012, doi: 10.1504/IJBIS.2012.046683.

[10] H. Aguinis, L. E. Forcum, and H. Joo, "Using Market Basket Analysis in Management Research," *J. Manage.*, vol. 39, no. 7, pp. 1799–1824, 2013, doi: 10.1177/0149206312466147.

[11] S. O. Abdulsalam, K. S. Adewole, A. G. Akintola, and M. A. Hambali, "Data Mining in Market Basket Transaction: An Association Rule Mining Approach," *Int. J. Appl. Inf. Syst. (IJAIS), Found. Comput. Sci. FCS, New York, USA*, vol. 7, no. 10, pp. 15–20, 2014.

[12] W. Nengsih, "A comparative study on market basket analysis and apriori association technique," *2015 3rd Int. Conf. Inf. Commun. Technol. ICoICT 2015*, pp. 461–464, 2015, doi: 10.1109/ICoICT.2015.7231468.

[13] A. Setiawan, G. S. Budhi, D. H. Setiabudi, and R. Djunaidy, "Data mining applications for sales information system using market basket analysis on stationery company," *Proc. - 2017 Int. Conf. Soft Comput. Intell. Syst. Inf. Technol. Build. Intell. Through IOT Big Data, ICSIIT 2017*, vol. 2018-January, pp. 337–340, 2017, doi: 10.1109/ICSIIT.2017.39.

[14] M. A. Valle, G. A. Ruz, and R. Morrás, "Market basket analysis: Complementing association rules with minimum spanning trees," *Expert Syst. Appl.*, vol. 97, pp. 146–162, 2018, doi: 10.1016/j.eswa.2017.12.028.

[15] T. Kutuzova and M. Melnik, "Market basket analysis of heterogeneous data sources for recommendation system improvement," *Procedia Comput. Sci.*, vol. 136, pp. 246–254, 2018, doi: 10.1016/j.procs.2018.08.263.

[16] R. Moodley, F. Chiclana, F. Caraffini, and J. Carter, "Application of uninorms to market basket analysis," *Int. J. Intell. Syst.*, vol. 34, no. 1, pp. 39–49, 2019, doi: 10.1002/int.22039.

[17] M. Hossain, A. H. M. S. Sattar, and M. K. Paul, "Market basket analysis using apriori and FP growth algorithm," *2019 22nd Int. Conf. Comput. Inf. Technol. ICCIT 2019*, pp. 18–20, 2019, doi: 10.1109/ICCIT48885.2019.9038197.

[18] Y. Q. Wei, R. H. Yang, and P. Y. Liu, "An improved apriori algorithm for association rules of mining," *ITME2009 - Proc. 2009 IEEE Int. Symp. IT Med. Educ.*, vol. 3, no. 1, pp. 942–946, 2009, doi: 10.1109/ITIME.2009.5236211.

[19] Dr. Daqing Chen, "UCI Machine Learning Repository: Online Retail II Data Set," *UCI Machine Learning Repository*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Online+Retail+II#.

[20] M. Heydari and A. Yousefli, "A new optimization model for market basket analysis with allocation considerations: A genetic algorithm solution approach," *Manag. Mark.*,

vol. 12, no. 1, pp. 1–11, 2017, doi: 10.1515/mmcks-2017-0001.

[21] F. Kurniawan, B. Umayah, J. Hammad, S. M. S. Nugroho, and M. Hariadi, "Market Basket Analysis to Identify Customer Behaviours by Way of Transaction Data," *Knowl. Eng. Data Sci.*, vol. 1, no. 1, p. 20, 2017, doi: 10.17977/um018v1i12018p20-25.