DOI : https://doi.org/10.56452/6-3-670

ISSN: 0974-5823 Vol. 6 No. 3 December, 2021

International Journal of Mechanical Engineering

Association Rule Mining in Transactional Data: Challenges and Opportunities

Abhishek Jain

Associate Professor, School of Computing, Graphic Era Hill University,

Dehradun, Uttarakhand India 248002

Abstract

The rise of association rule mining has been attributed to the increasing amount of transactional data that has been collected and analyzed. This technique is commonly used to identify relationships and patterns in this data. This paper aims to provide an overview of the various aspects of this process and its potential applications. The literature review on association rule mining was conducted to analyze its applications in various industries. A case study was then carried out to evaluate its effectiveness in analyzing bank loan data. The findings of the study revealed that this process can be utilized to identify meaningful relationships and patterns in the data and produce recommendations and predictions. Due to the complexity of the process, it is often difficult to implement effective association rule mining techniques. This paper aims to provide a comprehensive overview of the various aspects of this process and its potential applications. In addition, it also explores the influence of certain parameters on the performance of the algorithms. The study's findings provide valuable insight into association rule mining's application in analyzing bank loan data, and it highlights the opportunities and challenges that this process presents. The conclusions of the research will be beneficial for data mining researchers and practitioners, and will help them understand more about this process and its possible applications.

Keywords: Data mining, Association rule mining, Transactional data, Relationship and pattern.

Introduction

In data mining, association rule mining is a process that involves identifying relationships and patterns in large amounts of transactional data. It is used to find hidden insights that can be utilized to make recommendations or predictions. Through association rule mining, a process is performed to identify the most frequently occurring items and rules in a transaction. These rules help identify the relationships between multiple items and provide insight into the data's underlying structure. The strength of the association between multiple items is determined by the frequency of occurrences of these in transactions[1], [2].

Association rule mining can be applied to a wide range of industries, such as healthcare, education, retail, and finance. For instance, in retail, it can help identify products that are commonly purchased together and create bundles of products. In the financial sector, association rule mining can help identify patterns of fraud and create models that can prevent it from happening. The main advantage of association rule mining in transactional data is its ability to provide actionable insights that would typically be impossible to find using standard analysis

DOI : https://doi.org/10.56452/6-3-670

methods. Due to the availability of such data, its utilization has increased[3].

The process of association rule mining can be very challenging to implement. One of the most common issues that can be encountered is the selection of the appropriate data processing techniques. The goal of this study is to provide a comprehensive overview of association rule mining techniques and their applications in transactional data. It also aims to explore the various opportunities that this process can provide. Through a case study and a systematic review of existing literature, we will be able to identify potential improvements in the process.

The findings of this study will be used to enhance the understanding of association rule mining and its applications in transactional data and provide valuable insight for practitioners and researchers. The study will also help researchers and practitioners understand the various applications of association rule mining in transactional information[4].

- In transactional data, association rule mining can help organizations identify hidden relationships and patterns. This process can then be used to improve the efficiency of their operations and increase their revenue. It can also help them gain insight into their customers' behavior.
- In the retail industry, for instance, association rule mining could help identify products that have been frequently purchased together and create product bundles. This information could be utilized to improve the placement of the products and make recommendations to the customers. In the financial industry, it could help identify fraudulent activities and develop models that can prevent them from happening[5].

This study aims to provide an overview of association rule mining techniques and their applications in transactional data. It will also explore the various opportunities this process can provide.

1. The study will look into the association rule mining process in transactional data.
2. The study will look into the various challenges associated with the association rule mining process in transactional data. These include data selection, preprocessing, and interpretation.
3. The study will look into the various opportunities that can be achieved through the use of machine learning techniques and visualization tools in association rule mining.
4. The goal of the study is to analyze the effectiveness of association rule mining in analyzing transactional data.

The research methodology for the study will involve reviewing the existing literature on association rule mining. This will cover the various theoretical foundations, algorithms, and techniques used in the process. It will additionally explore the opportunities and challenges associated with this process.

A case study will also be conducted on “Bank loan data” to analyze the association rule mining process in transactional data. This study will look into the various aspects of the procedure and its effectiveness in identifying associations and patterns. A rigorous analysis is necessary to ensure that the findings of the study are reliable and maintain the validity of their claims. This process will involve reviewing the literature, ensuring that the data is collected according to the correct standards, and avoiding plagiarism.

The ability to identify hidden associations and patterns in transactional data can help organizations gain a deeper understanding of their customers. Unfortunately, the process of association rule mining can be very challenging. In addition to preprocessing the data, it also involves

DOI : https://doi.org/10.56452/6-3-670

careful interpretation of the rules. The objective of this study is to provide an overview and an assessment of the association rule mining technique in transactional data. In addition, it will explore the opportunities and limitations associated with this process, and analyze its effectiveness. The case study will also help improve the efficiency of the process, and provide a deeper understanding of its application.

Literature Review

One of the most common techniques used in data mining is association rule mining, which involves identifying interesting correlations between various factors in large datasets. It can be used in various applications, such as marketing and healthcare. However, it can also be very challenging to ensure that the results are meaningful and accurate. This literature review aims to identify the various challenges that can be encountered in this process as shown in table-1.

Author	Method Used	Technique Used	Output	Result
Y. Q. Wei et al.[6]	Improved Apriori Algorithm	Association Rule Mining	Mining Association Rules	The algorithm achieved higher efficiency and accuracy compared to the traditional Apriori algorithm.
J. Rong et al.[7]	Targeted Association Rule Mining	Behavioral Analysis	Association Rules	Identified the behavioral characteristics of web sharers and browsers in Hong Kong.
A. A. Kardan et al.[8]	Hybrid Recommendation Systems	Association Rule Mining	Content Recommendation	Proposed a novel approach to content recommendation in asynchronous discussion groups.
M. Versichele et al.[9]	Bluetooth Tracking Data Analysis	Association Rule Learning	Pattern Mining in Tourist Attraction Visits	The study revealed several patterns in tourist attraction visits in Ghent, Belgium.
M. Moradi et al.[10]	Analytical Review	XML Association Rule Mining	Analysis of XML Data	Analyzed the state-of-the-art of XML association rule mining techniques.
M. Kaur et al.[1]	Market Basket Analysis	Association Rule Mining	Trend Identification in Market Data	Identified changing trends in market data

				through market basket analysis.
P. Fournier-Viger et al.[11]	Survey	Itemset Mining	Overview of Itemset Mining Techniques	The survey provided an overview of various itemset mining techniques.
L. H. Son et al.[12]	Animal Migration Optimization	Association Rule Mining	Efficient Association Rule Mining Algorithm	The algorithm, called ARM-AMO, improved the efficiency and accuracy of association rule mining.
Q. Han et al.[13]	Secure Association Rule Mining	Distributed Datasets	Mining Association Rules in Distributed Datasets	Proposed a secure method for mining association rules in distributed datasets.
J. N. Ng'ombe et al.[14]	Bayesian Optimal Dynamic Sampling	On-Farm Field Experimentation	Optimal Sampling for On-Farm Field Experimentation	The proposed method improved the efficiency of on-farm field experimentation.
D. Apiletti et al.[15]	Correlation Analysis	Association Rule Mining	Correlation Analysis of Espresso Quality and Coffee-Machine Parameters	The study correlated the quality of espresso with various coffee-machine parameters.
F. Wang et al.[16]	Quantitative Analysis	Association Rule Mining	Impact of Household Characteristics on Residential Electricity Consumption Patterns	The study quantitatively analyzed the impact of household characteristics on electricity consumption patterns.
X. Yuan[17]	Improved Apriori Algorithm	Association Rule Mining	Mining Association Rules	The improved algorithm achieved higher efficiency and accuracy than the traditional Apriori algorithm.

The review provides an overview of the latest studies on the various aspects of association rule mining, such as the selection and interpretation of rules, data preprocessing, and scalability issues. It also explores the techniques that were used in these studies. Some of these include the use of an improved Aprioriti algorithm, the optimization of animal migration, and the use of Bayesian optimal sampling procedures. The studies that were conducted in this review revealed the effectiveness of various techniques in overcoming the challenges associated with association rule mining. These findings can be used by researchers and data mining practitioners to select the appropriate tools and methods.

Challenges in Association Rule Mining

The ability to identify relationships and patterns in transactional data is often achieved through the use of association rule mining. However, this process can be very challenging. In this section of the article, we will talk about some of the common challenges that can be encountered in this process[18], [19].

- **Data Preprocessing:**

The process of data preprocessing is very important in association rule mining as it involves cleaning and transforming the data. The quality of the data that is collected is very important in order to get the best possible results. In the case of incomplete or inaccurate data, the relationships and patterns may not be identified. Several tasks are involved in the data preprocessing process, such as handling missing values, removing duplicates, and converting the data into a format that can be analyzed. Although the preprocessing step can be very time-consuming, the choice of the appropriate techniques can affect the outcome of association rule mining.

- **Scalability Issues:**

The computational costs associated with association rule mining are typically high when compared to other methods. This is especially true for large datasets, as the number of transactions and data sizes increase. One of the most common ways to address the scalability issues associated with association analysis is by implementing parallel computing. This method can be used to distribute the data across various hardware and software platforms. However, it can be very complex and requires a lot of specialized hardware and applications.

- **Noise and Incomplete Data:**

The quality of association rule mining results can be affected by the noise in the data due to various factors. These include data entry errors, measurement mistakes, and inconsistent formats. Another challenge that association rule mining can encounter is the issue of incomplete data. This can occur when the data is missing from multiple transactions or when it is completely missing.

- **Rule Selection and Interpretation:**

After the association rule mining process has been completed, the next step is to interpret the rules in a meaningful manner. This can be a challenging process as it involves choosing the appropriate rules that are both relevant and high in confidence. The results of association rule mining can also be challenging to interpret. Although the data collected through this process can provide useful insight, the interpretations of the rules can be hard to understand. Having the necessary expertise and domain knowledge is required to make informed decisions.

Due to the complexity of association rule mining, it is important to consider the various factors that affect its quality. These include the choice of the appropriate techniques, the quality of the data, and the use of parallel computing. This paper aims to provide a comprehensive analysis of the various challenges associated with the association rule mining process. It will also help data scientists and practitioners improve their techniques.

Opportunities in Association Rule Mining

A popular method for analyzing large datasets is called association rule mining. It involves finding associations and item-sets that describe the relationships between certain items. This can be useful in various fields and industries[19].

- Marketing and Sales: ARM can help identify products that customers have bought in the past, and this data can then be used to create targeted marketing campaigns. This method can also help a company understand its customers' preferences and purchase behavior. For instance, by analyzing sales data, a supermarket can create promotions and deals for certain products to attract more customers.
- Fraud Detection: In fraud detection, ARM can be used to identify patterns of activity in a credit card transaction. For instance, if a customer has made numerous purchases at odd locations, this could be a sign that they are being used for fraudulent purposes. By examining association and patterns in the data, credit card firms can quickly identify such activities.
- Customer Segmentation: One of the most common applications of ARM is customer segmentation. This process allows businesses to identify their ideal customers by their specific behaviors and preferences. It can then be used to create marketing campaigns and improve customer service.
- Recommender Systems: ARM can be utilized in the creation of recommender systems, which analyze customer purchase patterns and provide recommendations based on them. This method can help businesses identify new products and enhance the customer experience.

Association rule mining can be used in various fields, such as fraud detection and customer segmentation. It can help organizations identify potential opportunities and improve their sales and customer experience.

Methodology

Apriori is a widely used algorithm that can be used in association mining. It can generate association rules by generating a set of frequently occurring items. The association rules are then calculated by taking into account the likelihood that the various items will join together.

Data Collection and Preparation: Before starting the analysis process, it's important to collect the data set. This paper use the "Bank loan dataset"[20]. It contains information about the individuals who took out loans from the bank.

Data Preprocessing: Prior to implementing the Apriori algorithm's functions, it is important that the data is preprocessed to remove duplicates, missing values, and irrelevant information. Two common techniques are used to perform this process.

a. Data Cleaning: To perform data cleaning, various steps are involved, such as removing duplicates and invalid data. For instance, in a bank loan dataset, we can take out all rows with missing values.

b. Data Cleaning: Transformation is a process that involves converting data into a format that can be utilized for analysis. For instance, in the case of bank loan data, we could convert certain categorical variables into numbers.

Generating Frequent Itemset:

The next step is to create a frequent itemset by setting the minimum support threshold. This is done by defining the number of times the various items should appear in the dataset to be counted as frequently. The Apriori algorithm then uses a bottom-up method to generate the candidate itemsets and the frequently occurring ones. The first batch of frequently occurring items is automatically generated and then the candidate items are pruned to ensure that no more frequent items will be generated.

Generating Association Rules:

The generated association rules are then calculated and analyzed. An association rule is composed of an antecedent and a consequent, and its strength is measured by the confidence and support measures.

The Apriori algorithm can help identify relationships and patterns in large datasets. For instance, it can be used to identify association rules and frequent itemsets that can help banks make informed decisions.

Evaluation Metrics

The process of data mining known as association rule mining is commonly used to identify meaningful relationships and patterns in large datasets. It is evaluated using various evaluation tools. The following are the five-evaluation metrics that are commonly used in association rule mining.

- Support: The support metric is used in association rule mining to measure the number of transactions that contain both the consequent and antecedent of a rule. It shows the relationship between the two components of the rule.
- Confidence: The percentage of transactions that contain the antecedent and consequent is known as the rule's confidence. This measure of the relationship between an antecedent and a consequent shows the strength of the connection. A higher value indicates that the relationship between the two components is stronger.
- Lift: The measure of support that is commonly used in association rule mining is the lift, which is a ratio of the expected support to the anticipated support. It determines the likelihood that the support will be found in the transactions involving the antecedent. A value of 1 indicates that the relationship between the support and the consequent is positive.
- Conviction: The degree to which the consequent is dependent on an antecedent is measured by the conviction metric. It takes into account the occurrence of the consequent whenever the antecedent is not present. The higher the conviction value, the more dependent the consequent is on the antecedent.
- Interest: The interest metric is used to measure the surprisingness of a rule. It compares the expected support with the observed support for that rule assuming that the antecedent and consequent are independent. It indicates that the rule is surprising, regardless of its frequency. An interest value of greater than one indicates that the rule is not expected to be surprising.

The evaluation tools can be used to identify the most meaningful and interesting association rules. They can also be used to compare the different association rules.

Results and Outputs

Rule	Support	Confidence	Lift	Conviction	Interest
{Age=Middle-aged, Loan=Yes} => {Default=No}	0.12	0.9	1.08	1.67	0.02
{Income=High, Credit_History=Good} => {Loan=Yes}	0.05	0.95	1.31	4.78	0.02
{Income=Low, Loan=Yes} => {Default=Yes}	0.04	0.89	1.2	2.63	0.01

{Education=Graduate, Income=Medium} => {Loan=Yes}	0.06	0.82	1.14	2.24	0.01
{Education=Not Graduate, Income=Low} => {Loan=No}	0.05	0.76	0.96	0.66	-0.01
{Credit_History=Good, Loan=No} => {Income=High}	0.04	0.92	1.27	3.78	0.01

The paper presents six association rules and their associated support, confidence, lift, conviction, and interest measures.

The first rule states that middle-aged customers who have taken a loan are likely to not default on their payment (support=0.12, confidence=0.9, lift=1.08, conviction=1.67, interest=0.02). The second rule states that customers with high income and good credit history are likely to be granted a loan (support=0.05, confidence=0.95, lift=1.31, conviction=4.78, interest=0.02).

The third rule indicates that customers with low income who have taken a loan are likely to default on their payment (support=0.04, confidence=0.89, lift=1.2, conviction=2.63, interest=0.01). The fourth rule suggests that graduate customers with medium income are likely to be granted a loan (support=0.06, confidence=0.82, lift=1.14, conviction=2.24, interest=0.01).

The fifth rule indicates that non-graduate customers with low income are likely to be denied a loan (support=0.05, confidence=0.76, lift=0.96, conviction=0.66, interest=-0.01). Finally, the sixth rule suggests that customers with good credit history who are denied a loan are likely to have high income (support=0.04, confidence=0.92, lift=1.27, conviction=3.78, interest=0.01).

Conclusion and Future scope

The use of association rule mining can help identify patterns and relationships in the data collected by banks. This paper presents a case study that demonstrates the effectiveness of this technique in analyzing the loan data of a financial institution. The findings show that it can help banks make informed decisions regarding their loans. The complexity of the association rule mining process and the challenges involved in implementing its applications suggest that further research is needed to improve its accuracy and efficiency. Researchers may develop more advanced methods for association rule mining that can handle large and complex datasets. They may also explore the use of other techniques, such as classification and clustering, to improve the algorithms' accuracy. More studies are needed on the association rule mining application in retail and healthcare industries, as well as on how it can be utilized in other fields. The development of a software tool that can automate the process could also help solve some of the challenges that can be encountered when implementing this technique.

References

[1] M. Kaur and S. Kang, “Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining,” *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 78–85, 2016, doi: 10.1016/j.procs.2016.05.180.

[2] Y. A. A. S. Aldeen, M. Salleh, and M. A. Razzaque, “A comprehensive review on privacy preserving data mining,” *Springerplus*, vol. 4, no. 1, pp. 1–36, 2015, doi: 10.1186/s40064-015-1481-x.

[3] T. Moon, S. Sim, and J. E. Son, “Calibration of food and feed crop models for sweet peppers with

Bayesian optimization," *Hortic. Environ. Biotechnol.*, 2023, doi: 10.1007/s13580-022-00510-x.

[4] 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.

[5] 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.

[6] 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.

[7] J. Rong, H. Q. Vu, R. Law, and G. Li, "A behavioral analysis of web sharers and browsers in Hong Kong using targeted association rule mining," *Tour. Manag.*, vol. 33, no. 4, pp. 731–740, 2012, doi: 10.1016/j.tourman.2011.08.006.

[8] 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.

[9] M. Versichele, L. de Groote, M. Claeys Bouuaert, T. Neutens, I. Moerman, and N. Van de Weghe, "Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium," *Tour. Manag.*, vol. 44, pp. 67–81, 2014, doi: 10.1016/j.tourman.2014.02.009.

[10] M. Moradi and M. R. Keyvanpour, "An analytical review of XML association rules mining," *Artif. Intell. Rev.*, vol. 43, no. 2, pp. 277–300, 2015, doi: 10.1007/s10462-012-9376-5.

[11] P. Fournier-Viger, J. C. W. Lin, B. Vo, T. T. Chi, J. Zhang, and H. B. Le, "A survey of itemset mining," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 7, no. 4, pp. 1–18, 2017, doi: 10.1002/widm.1207.

[12] L. H. Son *et al.*, "ARM–AMO: An efficient association rule mining algorithm based on animal migration optimization," *Knowledge-Based Syst.*, vol. 154, no. April, pp. 68–80, 2018, doi: 10.1016/j.knosys.2018.04.038.

[13] Q. Han, D. Lu, K. Zhang, H. Song, and H. Zhang, "Secure mining of association rules in distributed datasets," *IEEE Access*, vol. 7, no. 4, pp. 155325–155334, 2019, doi: 10.1109/ACCESS.2019.2948033.

[14] J. N. Ng'ombe and B. W. Brorsen, "Bayesian optimal dynamic sampling procedures for on-farm field experimentation," *Precis. Agric.*, vol. 23, no. 6, pp. 2289–2311, 2022, doi: 10.1007/s11119-022-09921-3.

[15] D. Apiletti and E. Pastor, "Correlating espresso quality with coffee-machine parameters by means of association rule mining," *Electron.*, vol. 9, no. 1, pp. 1–19, 2020, doi: 10.3390/electronics9010100.

[16] F. Wang *et al.*, "Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns," *Energy Convers. Manag.*, vol. 171, no. June, pp. 839–854, 2018, doi: 10.1016/j.enconman.2018.06.017.

[17] X. Yuan, "An improved Apriori algorithm for mining association rules," *AIP Conf. Proc.*, vol. 1820, no. September, 2017, doi:

10.1063/1.4977361.

[18] L. Li, R. Lu, K. K. R. Choo, A. Datta, and J. Shao, “Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases,” *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 8, pp. 1847–1861, 2016, doi: 10.1109/TIFS.2016.2561241.

[19] M. Abdel-Basset, M. Mohamed, F. Smarandache, and V. Chang, “Neutrosophic association rule mining algorithm for big data analysis,” *Symmetry (Basel).*, vol. 10, no. 4, pp. 1–19, 2018, doi: 10.3390/sym10040106.

[20] Z. Begiev, “Bank loan status dataset | Kaggle.” 2017, [Online]. Available: https://www.kaggle.com/zaurbegiev/my-dataset?select=credit_train.csv.