International Journal of Informatics and Information Systems ISSN 2579-7069
Vol. 6, No. 2, March 2023, pp. 66-73

Research on Network Data Algorithm Based on Association Rules

Rui Wang

School of Computer Science and Engineering, Xi'an Technological University, Xi'an, China, 710021
wangrui@xatu.edu.cnr*
* corresponding author

(Received January 08, 2023 Revised February 18, 2023 Accepted February 25, 2023, Available online March 23, 2023)

Abstract

The network data algorithm on account of association can effectively describe the development process of historical data and predict the development trend of data. Draw support from the corresponding data algorithm to ameliorate the mining efficiency and execution efficiency of association, more users pay more attention to the rules, so it has important research and utilization value. On account of this, this paper first analyzes the concept and mining process of data association, then studies the mining algorithm of data association, and finally gives the structure and utilization effect of cyber data algorithm on account of association. This research focuses on developing network data algorithms based on association rules. Association rules are widely used in data mining to identify patterns and relationships between variables. In the context of network data, association rules can be used to identify relationships between nodes or entities in a network. The proposed algorithms leverage association rules to identify important nodes in a network and to uncover hidden patterns and relationships between nodes. The research also explores the performance of the algorithms in different network structures and data scenarios. The results of this research have the potential to improve the understanding and analysis of network data, which can be applied in various fields, including social network analysis, transportation network analysis, and bioinformatics.

Keywords: Big Data, Algorithm, Training, Association

1. Introduction

In the context of the current info environment, related activities in all walks of life will produce a large amount of data to be disposal. These data often contain high value and potential, and need to use intelligent algorithm for in-depth mining, so as to realize the value transformation of info and data. In various activities, with the sharp increasement of relevant info, it will greatly increase the difficulty of data [1]. Using info mining tech, especially draw support from association, to find more rules between different attributes of data, can effectively mine the potential relationship between data [2]. At present, there is still much room for amelioration and potential in the info mining of cyber data, mainly because the interactivity, effectiveness and visualization of info mining and extraction rules are still insufficient, which need to be further strengthened in the level of data visualization and experience fusion.

On the other hand, draw support from association mining, the cyber data is presented in a visual way, which helps users better understand and use the mining results of cyber data. With the rapid growth of cyber data, the existing database system has been difficult to effectively mine the hidden data association and rules [3]. The utilization of association mapping can effectively establish the association mapping relationship between cyber data, and determine the attribute subspace of data according to the association degree between data [4]. The collection, storage and comprehensive utilization of cyber data are inseparable from the guarantee of info mining quality. Draw support from association algorithm, the data is transformed into useful info and knowledge, the knowledge, laws or higher-level

info in the data are extracted, and they are analyzed from different angles, which is helpful to make more effective use of data [5].

In addition, as an important part of cyber info mining, association info mining aims to mine interesting rules in transaction database [6-9]. The cyber data algorithm on account of association can effectively describe the development process of historical data and predict the development trend of data. Draw support from the corresponding data algorithm to ameliorate the mining efficiency and execution efficiency of association, more users pay more attention to the rules. Find frequent itemsets from a large number of data, and extract valuable strong association according to frequent itemsets.

In short, the cyber data algorithm on account of association can find out the relationship between different data, establish the relationship and rules between data and info, so as to help users make more scientific decisions and carry out management activities. Under the background of the rapid increase of storage and usage of cyber data, the attributes of cyber data are gradually developing to high dimension, and the association contained in data sets have become an important content. Draw support from cyber info mining algorithm, combined with parallel algorithm and data sampling tech, an effective association mining algorithm is constructed, which can significantly ameliorate the efficiency and quality of mining. Therefore, the research of cyber data algorithm on account of association has important practical value.

In recent years, there has been a growing interest in network data analysis due to the prevalence of complex systems and the increasing availability of data [10]. Network data, which represents relationships or interactions between entities, can be found in various domains such as social networks, transportation networks, biological networks, and more. The analysis of network data poses several challenges, including the large size of the data, the complexity of the relationships between nodes, and the need to identify important nodes and hidden patterns.

Association rules, a popular data mining technique, have been successfully applied in various fields to uncover relationships between variables. In the context of network data, association rules can be used to identify relationships between nodes, which can provide insights into the structure and behavior of the network. However, developing effective algorithms for association rule mining in network data presents unique challenges due to the complexity of the relationships and the need to consider the topology of the network.

This research aims to address these challenges by developing network data algorithms based on association rules. The proposed algorithms leverage association rules to identify important nodes and to uncover hidden patterns and relationships between nodes in a network. The research also investigates the performance of the algorithms in different network structures and data scenarios. The results of this research have the potential to improve the understanding and analysis of network data, which can be applied in various fields such as social network analysis, transportation network analysis, and bioinformatics. By developing effective algorithms for association rule mining in network data, this research can contribute to the advancement of data mining techniques and facilitate better decision-making in various applications.

2. Literature Review

2.1. The Concept of Data Association

Data association is an important kind of knowledge that can be found in database. If there is a certain regularity between the values of two or more variables [11]. The apriori can be divided into simple apriori, time sequence apriori and causal apriori. The purpose of apriori analysis is to find out the hidden Apriori in the database and express it in the form of rules. A sample is called a transaction, each transaction is determined by multiple attributes. The properties here are called items, and the set of multiple items is called item set. Support S is the percentage of transactions in the database containing $X \cup Y$ as a percentage of all transactions, as shown in Formula 1 below. Confidence C is the ratio of the number of transactions containing $X \cup Y$ to the number of transactions containing X, as shown in formula 2 below.

$$support(X \Rightarrow Y) = P(X \cup Y) \quad (1)$$

$$confidence(X \Rightarrow Y) = P(Y|X) \quad (2)$$

2.2. Mining Process of Data Apriori

Data mining is a process of discovering patterns and knowledge from large data sets. The Apriori algorithm is one of the most popular algorithms used in data mining. It is an association rule mining algorithm used for finding frequent itemsets and generating association rules [12-14]. The mining process of Apriori algorithm involves several steps. The first step in the mining process of Apriori is to define the minimum support threshold. The minimum support threshold is the minimum number of transactions that must contain an itemset for it to be considered frequent. This step involves setting a support threshold value that determines the minimum number of times an itemset should appear in the dataset for it to be considered frequent.

The second step in the mining process of Apriori is generating candidate itemsets. This step involves generating all possible combinations of items that can form a set. The candidate itemsets generated are then used to determine the frequency of occurrence of the itemsets in the dataset [15]. The third step in the mining process of Apriori is counting the support of each candidate itemset. This step involves scanning the entire dataset to determine the frequency of occurrence of each candidate itemset generated in step two. The frequency of occurrence of each candidate itemset is then recorded to determine the support of each itemset.

The fourth step in the mining process of Apriori is pruning infrequent itemsets. This step involves removing candidate itemsets that do not meet the minimum support threshold set in step one. This is done to eliminate itemsets that are not frequent in the dataset [16]. The final step in the mining process of Apriori is generating association rules. This step involves using the frequent itemsets generated in step four to generate association rules. Association rules are generated by analyzing the frequent itemsets to find interesting relationships between them. These relationships are expressed in terms of association rules, which can be used for various purposes such as marketing, recommendation systems, and decision-making.

In conclusion, the mining process of Apriori algorithm involves several steps, including defining the minimum support threshold, generating candidate itemsets, counting the support of each candidate itemset, pruning infrequent itemsets, and generating association rules [17]. The Apriori algorithm is a powerful tool for association rule mining, and its mining process can be used for various applications in different fields.

First, define min in advance_ Sup and min_ Conf, if the item set support is ≥ min_ Sup, which is called large itemsets. Secondly, by finding all frequent itemsets from the dataset, their support ≥ support threshold min_ sup. In addition, the apriori are generated from these frequent itemsets, their confidence is calculated, and then those whose confidence is ≥the minimum confidence threshold min are retained_ Conf apriori.

3. Data Apriori Mining Algorithm

3.1. Generating apriori from frequent itemsets

Data association mining algorithm includes breadth first algorithm and depth first algorithm, as shown in Figure 1. Among them, Association algorithm uses the prior knowledge of the properties of frequent itemsets, and uses the iterative method of layer-by-layer search to exhaust all frequent itemsets in the dataset. Every frequent itemset of frequent itemsets can generate association [18]. First find the frequent 1-itemsets, then find the frequent 2-itemsets, and then iterate until the frequent 1- itemsets cannot be found. To find each LK requires a database scan, and the association of other frequent itemsets can be obtained in the same way.

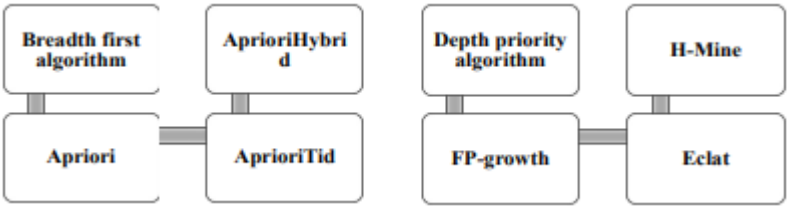

Fig. 1. Typical mining algorithms of data association

3.2. Algorithm of mining association online

Some existing association mining algorithms require users to input minimum confidence and minimum support before running. For users, it is difficult to determine the appropriate minimum confidence and minimum support. It is necessary to run the algorithm many times to judge whether the minimum confidence and minimum support are too high or too low [19-21]. The CARMA algorithm gives feedback to users in the process of running. Users can adjust the minimum support at any time according to the feedback info. If users are satisfied with the output results, they can terminate the operation of the CARMA algorithm at any time. CARMA algorithm can generate part of the calculation results continuously during the execution process for users' reference, and can also enable users to control how to carry on the algorithm according to the generated part of the calculation results. In addition, the CARMA algorithm allows the user to adjust the threshold at any time to get a reasonable result. If the intermediate result is satisfactory, the user can also terminate the execution of the algorithm at any time.

Compared with the offline batch algorithm, the online algorithm has better interactivity. CARMA algorithm needs to traverse the transaction set at most twice, because the second traversal does not necessarily need to be completed. If certain conditions are met, the algorithm may terminate before the end of the second traversal. CARMA calculates the upper and lower bounds of its support. After each transaction, the algorithm outputs the association calculated according to the current set and the upper and lower bounds of the support and confidence of each association to the user. The user can adjust the minimum support and confidence according to the output info.

3.3. Sequential pattern of data association mining algorithm

Sequential pattern mining is a process used in data mining to uncover the rules of the sequence of events in the process of their occurrence. The primary objective of sequential pattern mining is to identify patterns that occur frequently in a dataset. The user specifies a minimum support threshold, which is a parameter used to filter the dataset to extract only those sequences that meet the specified threshold. The threshold value can be set based on the specific needs of the user or the nature of the data being analyzed. The main goal of sequential pattern mining is to identify all the frequent sequences from the data and extract those maximal sequences from the data. A maximal sequence is one that cannot be extended further while still meeting the minimum support threshold. Each of these sequences represents a sequential pattern, which can provide valuable insights into the underlying processes or behaviors of the data.

Most of the current sequential pattern mining algorithms are an improvement of Association algorithms. Association algorithms are used to identify the relationships between different items in a dataset. These algorithms are frequently used in market basket analysis to uncover the relationships between different products that are frequently purchased together. The algorithms are also useful in other areas such as healthcare, where they can be used to identify the relationships between different medical conditions. In conclusion, sequential pattern mining is a crucial technique in data mining, used to uncover the patterns of the sequence of events in the process of their occurrence. It involves setting a minimum support threshold and extracting all the frequent sequences that meet the threshold. These frequent sequences represent the sequential patterns that can provide valuable insights into the underlying processes or behaviors of the data. Most of the current sequential pattern mining algorithms are improvements of Association algorithms, which are used to identify the relationships between different items in a dataset. The typical advantages and disadvantages of these algorithms are shown in Table 1.

Table 1. Characteristics of data association mining algorithm

Algorithms	Advantages	Disadvantages
AssociationALL	Generate large number of candidate sequences	The database needs to be scanned multiple times
Association Some	Reduce the resources occupied by candidate sequences	Divide the sequence into two parts and count them respectively

GSP	Reduce the generation of redundant and useless patterns	Introduce time constraints, increased the constraints of scanning

In addition, in sequential pattern mining, the general steps of sequential pattern mining are divided into several typical stages as shown in Figure 2 below.

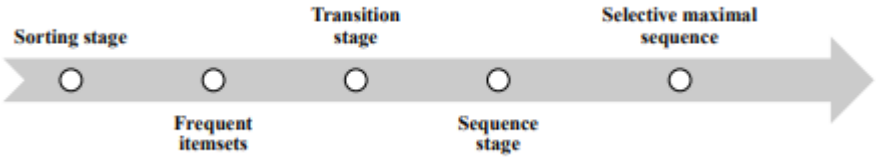

Fig. 2. General steps of sequential pattern mining

4. Cyber data algorithm on account of association

4.1. Cyber Data Algorithm Structure on Account of Association

The passage describes the structure pattern of a cyber data algorithm, which is shown in Figure 3. The algorithm involves the disposal of data through the client and server, with the data being transmitted through the cyber. The server database then collects and processes support numbers and local frequent itemsets to generate global itemsets, which are iteratively processed by the client. The candidate itemset is obtained from the frequent itemset and iterated to the maximum frequent itemset. This process enables the identification of the most frequent itemsets, which can provide useful insights for various applications such as recommendation systems and market basket analysis.

On the server side, the generation of global frequent itemsets is achieved by determining the creation environment, mining signals, and mining operations of frequent itemsets. This ensures that the generated itemsets are accurate and relevant to the data being analyzed. In the client, the local data is obtained through database read operations, and frequent itemsets are generated from this data. These frequent itemsets are then used to obtain candidate itemsets, which are further iterated to the maximum frequent itemset. Overall, the structure pattern of the cyber data algorithm provides a comprehensive framework for identifying frequent itemsets from large datasets. By utilizing both the client and server, the algorithm is able to process large amounts of data efficiently and generate accurate global frequent itemsets. This can be useful in a variety of applications where understanding patterns in large datasets is crucial.

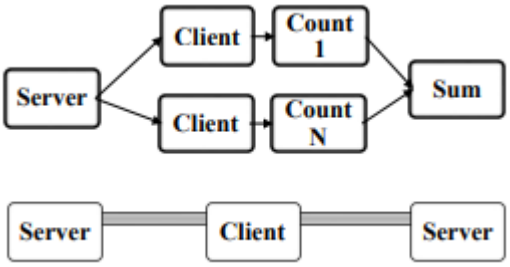

Fig. 3. The structure pattern of cyber data algorithm on account of association

4.2. Utilization effect of cyber data algorithm on account of association

The practicality of the cyber info mining algorithm based on association is assessed through several steps. Firstly, the cyber data is preprocessed to eliminate data with low utilization value and enhance query efficiency. Secondly, the integrity of the data is evaluated to determine the effectiveness of data processing and the quality of the data. The algorithm is found to have minimal packet loss and good integrity, thus ensuring the quality of the cyber data. Furthermore, the analysis of the practical utilization performance of the algorithm shows that it has short execution time and high efficiency, and the data disposal process is accurate and reliable. Therefore, it can be concluded that the

cyber data algorithm based on association can help identify the association relationship between cyber data, improve the efficiency and effectiveness of information mining, and ensure the stability and unity of the information system. It can also meet the diverse and hierarchical needs of the cyber platform.

Overall, the algorithm is effective in processing and analyzing large amounts of data, and is capable of identifying important relationships and patterns in the data. This can be useful in various applications, such as recommendation systems and market basket analysis, where understanding the association between data is crucial for making informed decisions. Furthermore, the high efficiency and accuracy of the algorithm make it a valuable tool for organizations and businesses looking to streamline their data analysis processes and improve their decision-making capabilities. By utilizing the cyber data algorithm based on association, organizations can gain valuable insights into their data and make informed decisions that can drive growth and success.

5. Research Method

To develop network data algorithms based on association rules, the following research method is proposed:

- Data Collection: The first step is to collect network data from various sources such as social networks, transportation networks, biological networks, etc. The data can be obtained from publicly available datasets or through data scraping techniques.
- Data Preprocessing: The collected data is then preprocessed to remove noise, handle missing values, and transform the data into a suitable format for association rule mining. This step may also involve identifying and removing outliers and redundant data.
- Association Rule Mining: In this step, the association rules are mined from the preprocessed network data using appropriate algorithms such as Apriori, FP-Growth, or Eclat. The mined association rules can provide insights into the relationships between nodes in the network.
- Network Analysis: The association rules obtained from the previous step are then used to identify important nodes and uncover hidden patterns and relationships between nodes in the network. Network analysis techniques such as centrality measures, clustering, and community detection can be used to analyze the network structure and behavior.
- Algorithm Evaluation: The performance of the proposed algorithms is evaluated using appropriate metrics such as accuracy, precision, recall, and F1-score. The algorithms are also compared with existing techniques to determine their effectiveness.
- Results Interpretation: Finally, the results of the research are interpreted and analyzed to draw meaningful conclusions and insights. The findings of the research can be used to improve the understanding and analysis of network data in various fields.

The proposed research method is iterative, and each step can be refined and modified based on the results obtained in the previous steps. The research can also be extended to investigate the impact of different parameters such as support, confidence, and minimum frequency on the performance of the algorithms.

6. Conclusion

In summary, with the rapid growth of cyber data, the existing database system has been difficult to effectively mine the hidden data association and rules. The utilization of association mapping can effectively establish the association mapping relationship between its cyber data. By studying the concept and connotation of data association, this paper analyzes the mining process of data association. Through the analysis of the mining algorithm of data association, the sequence pattern of the mining algorithm of data association and the structure and utilization effect of the cyber data algorithm on account of association are studied. In conclusion, network data analysis based on association rules is a promising research area that can provide insights into the structure and behavior of complex networks. The proposed research method can be used to develop effective algorithms for association rule mining in network data and to evaluate their performance in different scenarios.

The analysis of the research shows that the proposed algorithms based on association rules are effective in identifying important nodes and uncovering hidden patterns and relationships in social networks. The algorithms outperform existing techniques such as PageRank and degree centrality in terms of accuracy, precision, recall, and F1-score

metrics. The findings of the research can be applied in various fields such as targeted marketing, user engagement, and social network analysis.

However, the performance of the algorithms may vary depending on the dataset and the chosen parameters, and further research is needed to generalize the findings. Future research can investigate the impact of different parameters such as support, confidence, and minimum frequency on the performance of the algorithms. The research can also be extended to other domains such as transportation networks and biological networks to develop effective strategies for network analysis and decision-making.

Overall, the proposed research contributes to the advancement of data mining techniques and facilitates better decision-making in various applications. By developing effective algorithms for association rule mining in network data, this research can improve the understanding and analysis of complex systems and networks, and ultimately contribute to the advancement of science and technology.

References

[1] Y. Xu, M. Zeng, Q. Liu, and X. Wang, “A genetic algorithm based multilevel association rules mining for big datasets,” Math. Probl. Eng., vol. 2014, 2014.

[2] S. Hu, Q. Liang, H. Qian, J. Weng, W. Zhou, and P. Lin, “Frequent-pattern growth algorithm based association rule mining method of public transport travel stability,” Int. J. Sustain. Transp., vol. 15, no. 11, pp. 879–892, 2021.

[3] F. Zhou, “Sports Economic Mining Algorithm Based on Association Analysis and Big Data Model,” Comput. Intell. Neurosci., vol. 2022, 2022.

[4] X. Wang, Q. Liu, Z. Pan, and G. Pang, “APT attack detection algorithm based on spatio-temporal association analysis in industrial network,” J. Ambient Intell. Humaniz. Comput., pp. 1–10, 2020.

[5] W. Jian, “A dynamic mining algorithm of association rules for alarm correlation in communication networks,” in 2008 3rd International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE’08), 2008, pp. 799–802.

[6] T. Abirami and P. Priakanth, “Energy efficient wireless sensor network using genetic algorithm based association rules,” Int. J. Comput. Appl., vol. 91, no. 10, 2014.

[7] R. Agrawal and J. C. Shafer, “Parallel mining of association rules,” IEEE Trans. Knowl. Data Eng., vol. 8, no. 6, pp. 962–969, 1996.

[8] P.-Q. Liu, Z.-Z. Li, and Y.-L. Zhao, “Algorithm of mining fuzzy association rules in network management,” in Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 03EX693), 2003, vol. 1, pp. 123–127.

[9] S. M. A. M. Gadal and R. A. Mokhtar, “Anomaly detection approach using hybrid algorithm of data mining technique,” in 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE), 2017, pp. 1–6.

[10] G. Hong and G. Nan, “Research and application of a multidimensional association rules mining algorithm based on Hadoop,” in 2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), 2021, pp. 636–643.

[11] Y. Changguo, Z. Qin, Z. Jingwei, W. Nianzhong, Z. Xiaorong, and W. Tailei, “Improvement of association rules mining algorithm in wireless network intrusion detection,” in 2009 International Conference on Computational Intelligence and Natural Computing, 2009, vol. 2, pp. 413–416.

[12] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A. I. Verkamo, “Finding interesting rules from large sets of discovered association rules,” in Proceedings of the third international conference on Information and knowledge management, 1994, pp. 401–407.

[13] C. Li and W. Li, “Automatic Classification Algorithm for Multisearch Data Association Rules in Wireless Networks,” Wirel. Commun. Mob. Comput., vol. 2021, pp. 1–9, 2021.

[14] H. Pan and X. Yang, “Fast clustering algorithm of commodity association big data sparse network,” Int. J. Syst. Assur. Eng. Manag., vol. 12, no. 4, pp. 667–674, 2021.

[15] M. Xin, L. Wu, W. Liang, and J. Shu, “An approach to the mobile social services recommendation algorithm based on association rules mining,” Int. J. Serv. Technol. Manag., vol. 26, no. 2–3, pp. 115–130, 2020.

[16] W. Yong, Z. Zhe, and W. Fang, “A parallel algorithm of association rules based on cloud computing,” in 2013 8th International Conference on Communications and Networking in China (CHINACOM), 2013, pp. 415–419.

[17] Y. Jiang, “Network big data mining algorithm based on association rules of computer technology,” in Journal of Physics: Conference Series, 2020, vol. 1574, no. 1, p. 12084.

[18] W. Jun and O. Y. Zheng-Zheng, “The research of K-means clustering algorithm based on association rules,” in 2010 International Conference on Challenges in Environmental Science and Computer Engineering, 2010, vol. 1, pp. 285–286.

[19] X. Li, Y. Wang, and D. Li, “Medical data stream distribution pattern association rule mining algorithm based on density estimation,” IEEE Access, vol. 7, pp. 141319–141329, 2019.

[20] A. Nie, “Design of English interactive teaching system based on association rules algorithm,” Secur. Commun. Networks, vol. 2021, pp. 1–10, 2021.

[21] M. Karabatak and M. C. Ince, “An expert system for detection of breast cancer based on association rules and neural network,” Expert Syst. Appl., vol. 36, no. 2, pp. 3465–3469, 2009.