Project Management Model of Internet of Things Application System Based on Association Rule Mining Algorithm

Yajuan Zhang, Ru JING*, Haifeng LU, Jiajing CAI

Hainan Vocational University of Science and Technology, Haikou, Hainan 571126, China

305969190@qq.com

*Corresponding Author

Keywords: Association rule mining algorithm, Internet of things application system, Project management

Abstract: With the rapid development of a series of technologies such as Internet, Internet of things and information technology, today's society has entered the era of big data. In order to efficiently process these diversified massive data, we need to use the operation and storage functions provided by various distributed platforms to feed back the results to the above applications through the management, processing, analysis and optimization of data from different sources, which will create great economic and social value. Association rule is an important method and technology in data mining. Cluster analysis, predictive modeling, association analysis, anomaly detection. To get valuable information from complex data, we need to deeply study and flexibly use data mining algorithms. Data mining is a key part of the intelligent application of the Internet of things. Applying it to the field of the Internet of things can promote the improvement of intelligent application level and optimize user experience. Facing the massive data information generated under the background of in-depth application of the Internet of things. Based on the characteristics of association rules, this paper constructs an association rule tree, and studies the search strategy, classification strategy and parameter selection of the tree.

1. Introduction

Internet of things technology has gradually extended from the research stage to real life. More and more Internet of things applications appear in people's vision. More and more companies have joined the wave of Internet of things development. Internet of things technology and products have developed rapidly. There is no doubt that Internet of things will become another new information hotspot after the Internet [1]. Various industries related to people's clothing, food, housing and transportation have realized the wide application of Internet of things technology. Under the action of a large number of sensor nodes, all kinds of Internet of things application systems can exchange information and realize communication [2]. The universal application of information management system, database system and Internet of things has gradually increased people's ability to generate and collect data in the process of daily life and management [3]. Data mining technology is highly comprehensive. It involves interdisciplinary knowledge such as database technology, artificial intelligence technology, statistics, pattern recognition, information retrieval, computer network and application, computer software, hardware and operating system [4].

With the passage of time and the deepening of application, the scale and dimension of the generated data grow rapidly. The traditional data analysis method only counts the data, which can not achieve the in-depth acquisition of data value. In order to make better use of the generated large amount of data and meet the extraction of the information contained in the data, data mining technology is produced [5]. With the wide application of data mining technology in various fields, the academic and Industry Research on data mining technology is more in-depth, and the mining algorithm is also gradually improved. Data mining has become a research hotspot in various fields [6]. With the rapid development of science and technology, especially the development of computer technology, it has made unparalleled contributions to improving life and production [7]. With the

  DOI: 10.25236/iiiece.2022.034

continuous updating of information technology, people's lifestyle and rhythm have gradually changed. People have more and more demand for knowledge and information, and the amount of data is almost growing at a geometric speed [8]. Therefore, many industries begin to pay attention to the update of information technology in order to obtain and store data more quickly and efficiently. Association rule mining has become one of the most widely used data mining methods because of its simple principle, easy operation process and providing users with satisfactory mining results [9]. A large amount of data is generated in the process of human interaction with sensors, and there are many valuable information in such data. After in-depth mining and analysis, we can promote the research and development of intelligent applications [10].

2. Association Rule Mining

2.1 Basic Concepts of Association Rule Mining

If there is some regularity between the values of two or more variables, it is called correlation [11]. Association rules are to find the correlation of different items in the same event, such as the correlation of different goods purchased in a purchase activity [12]. Association analysis is data mining using association rules. In the research field of data mining, the research on association analysis is carried out in depth. The purpose of association analysis is to mine the relationship between data and automatically detect the hidden patterns that have not been found before [13]. The maximum data item set is used to generate the required association rules, the choice of rules is determined according to the minimum confidence specified by the user, and finally the strong association rules are obtained. The basic model of association rule mining is shown in Figure 1:

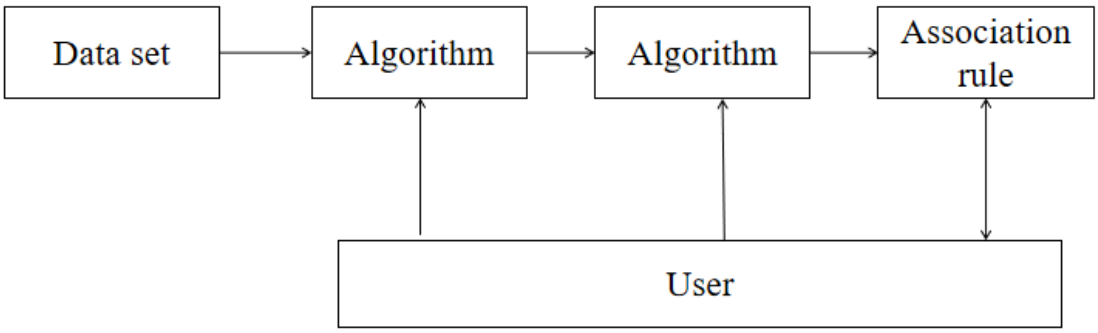

Fig.1 Basic Model of Association Rule Mining

Association model mainly describes the closeness or relationship of a group of data items. Correlation can be divided into simple correlation, temporal correlation and causal correlation. Relationships or rules are always described with some minimum confidence level. The confidence level measures the strength of association rules. According to the types of values processed in association rules, association rules can be divided into Boolean association rules and quantitative association rules. Association rule mining describes the dependency between data. Because it is widely used, simple and practical, it has become an important content in the research of mining algorithms. In order to improve the effective efficiency and application scope of association mining algorithm, especially to keep up with the development of cloud era, researchers optimize and improve the algorithm from all aspects and angles.

Association is the correlation or interdependence between two or more transactions. In practice, not all the association rules mined are meaningful, and there are many useless or meaningless association rules. Association rule mining algorithm is a classic algorithm in data mining. Many scholars have done a lot of research on this algorithm. According to the different results of association analysis, association rules can be divided into three mining types: frequent closed itemset mining, maximum frequent itemset mining and complete frequent itemset mining. Because the mining results of complete frequent itemsets will produce a large number of redundant frequent itemsets, and the mining results of maximum frequent itemsets can summarize all frequent itemsets. The traditional data mining algorithm is improved to significantly improve the running speed of the

algorithm. By studying the traditional classical data mining algorithm, an efficient distributed data mining algorithm suitable for big data is improved and proposed to mine the potential value of data more effectively.

2.2 Mining Goal and Mining Process of Association Rules

Association rule mining is mainly to find some hidden correlation or law between one or some attributes of data from a large amount of data, that is, association rules, and provide action guidance for practice. Numerical and Boolean association rules. The processing object of numerical association rules is numerical data. It is necessary to segment the data, divide the data into different intervals, and take an interval as an attribute. This section presents the theoretical framework of association rule mining algorithm based on metadata integration, which is convenient for readers to understand the specific steps of the algorithm more clearly. The framework is shown in Figure 2:

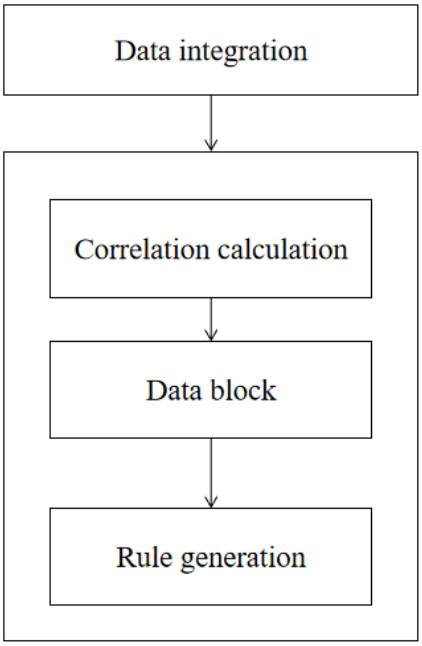

Fig.2 Framework of Association Rule Mining Algorithm Based on Metadata Integration

In addition, such rules can also have category variables, and a category can be used as an attribute; Single layer and multi-layer association rules. Single layer association rules do not consider that data has different levels in real life, and ignore the attributes of different abstraction layers; One dimensional and multi-dimensional association rules. One dimensional association rules only involve one dimension of processing data, which reveals the relationship between one-dimensional attributes of processing data. Association rule mining is the process of refining data into information, condensing information into knowledge, and then applying knowledge to practice, that is “data information knowledge practice”, and finally to extract knowledge from data to guide practice. By analyzing consumers' shopping habits to find consumers' purchase patterns, so as to guide supermarkets in shelf design, commodity promotion and so on; Analyze the drug sales of pharmacies in a region, predict infectious diseases, and prevent them in advance; Predict the stock market situation by analyzing the stock market; By analyzing the user's friend relationship data, you can recommend friends to users, making the user experience more humanized. The research of association rule mining mostly focuses on the static database, that is, the database with unchanged data records. By comparing and analyzing the relationship between data schema and ontology concept, the relationship between model and local ontology can be obtained. Figure 3 describes the relationship between local data source and local ontology.

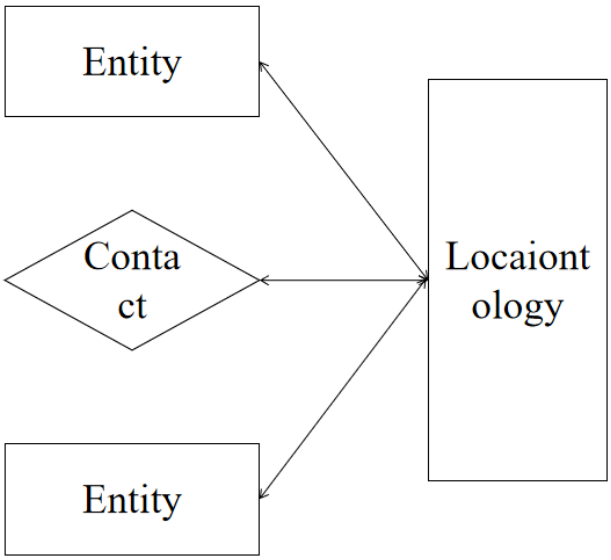

Fig.3 Relationship between Local Data Source and Local Ontology

However, in practical applications, no database is static. With the increase of people's operations on the database, the data records of the database will also change, which is a dynamic rather than static process. In order to meet the needs of users, mine the latest association rules reflecting the current database state and provide reliable data basis for users' decision-making, an effective association rule update algorithm is needed. The update algorithm is used to find effective knowledge on the basis of the old data that has been mined and the new data that has not been mined. It can not only effectively solve the heterogeneous problems in syntax and semantics, but also obtain the information related to local data sources through metadata, so as to reduce the integration cost and improve the integration efficiency.

3. Internet of Things Application System Management

3.1 Application of Data Mining Technology in Internet of Things

People's life and work have been improved due to the birth of the Internet of things. At the same time, the application of the Internet of things in product quality supervision and management also shows considerable results. Since there are still some deficiencies in China's quality supervision system, and incidents of unqualified drugs, unqualified projects and food additives are common, the government quality supervision department can complete the construction of the traceability system through the application of Internet of things technology. Association rule mining is to find interesting knowledge between data items and itemsets from a large amount of data

The support and confidence of association rules are two measures of rule interest. The relevant definitions of association rule mining are as follows: Definition 1. Support s is the probability or percentage of all transactions in D of the union of article a and article B, $P(A \cup B)$ which is the probability, that is:

$$S(A \Rightarrow B) = P(A \cup B) \quad (1)$$

Definition 2: the occurrence frequency of itemsets refers to the number of transactions containing itemsets, which is called support count C ', also known as absolute support, that is:

$$C(A \Rightarrow B) = P(A|B) \quad (2)$$

Definition 3: confidence C is the probability or percentage of article B under the premise of article a, $P(B|A)$ which is the conditional probability, that is:

$$C(A \Rightarrow B) = P(B|A) = \frac{S(A \cup B)}{S(A)} = \frac{S - C(A \cup B)}{S - C(A)} \quad (3)$$

A large amount of data will be generated in the operation of the quality supervision system.

When looking for links affecting quality, data mining technology can reduce time-consuming, save resources, find problems in time and effectively optimize internal quality management. The continuous popularization of Internet of things in logistics enterprises has reduced certain costs and promoted the improvement of economic benefits of enterprises.

By using the Internet of things, logistics enterprises can accurately identify and track each commodity in real time under the application of wireless data communication technologies such as Internet and RFID. Store various data in the database, mine the potential information of logistics transportation, warehousing and management under the application of data mining technology, and provide the necessary reference information for the management of enterprise managers. In this way, we can save a lot of costs and optimize the service efficiency to a certain extent. For each enterprise, the problem of cost consumption has attracted much attention. The small labor and materialized labor currency in product time occupation or space movement is the cost, which has a complex production process and is closely related to all operation links of the enterprise. Through the application of data mining technology, information can be provided for the rational allocation of enterprise resources, which can save a lot of costs and obtain greater benefits on the basis of reducing the total cost. Active service refers to intelligent information service, which actively provides information services that can meet customer needs; Passive service means to find the potential demand information service based on the customer demand information and the application of data mining technology. Active services can provide relevant data for passive services and promote the use of data mining; Passive service can provide information reference for active service and optimize enterprise service projects. Therefore, there is a complementary and interdependent relationship between active and passive services.

3.2 Data Mining of Internet of Things Based on Cloud Computing

The platform data mining selects the Internet of things dataset as an example, and selects Hadoop, the current research hotspot, as the basic platform. The role of the sensing layer is mainly to arrange a large number of acquisition nodes in the target area. These nodes collect IOT data through sensors, cameras or other instruments. These data will have communication in the IOT sensing layer, that is, there are wireless sensor networks, which gather data to the aggregation nodes through these networks, Then the data is summarized and stored, and finally transmitted to the cloud platform data center through the transport layer. It can flexibly and quickly transmit the perceived data to the cloud computing data center to realize more comprehensive interconnection; Carry out networked data transmission of various monitoring equipment to realize networked high-speed data transmission of monitoring equipment in the Internet of things. The data layer is very important for the whole Internet of things data mining platform. Because we have mentioned the heterogeneous and massive characteristics of Internet of things data, the data storage and processing in the data layer, such as and solving the Internet of things, determine the feasibility and performance of the Internet of things data mining platform.

In the Internet of things, different objects will be represented by different data types, and even the same objects will be represented by different data. Therefore, the role of data source converter is mainly to solve the data heterogeneity of the Internet of things. It can not only ensure the integrity of data storage, but also ensure the smooth progress of data mining, The data source conversion module is equivalent to the interface between the data layer and each monitoring device in the perception layer, and completes the packet decoding and finally makes the distributed storage module store effective and complete data according to the corresponding data model. The data mining service layer mainly includes data preparation module, data mining engine module and user module. The data preparation module mainly includes data cleaning, transformation, data protocol, etc; The data mining engine module mainly includes data mining algorithm set, pattern evaluation, etc; Users can carry out data mining tasks through interface operation, and can get understandable knowledge. In order to enhance the portability of the platform, an open interface is added to the bottom module of user service, so that the third party can call the functions of the Internet of things data mining platform and enrich the applications of the Internet of things.

4. Conclusions

Data mining technology is a young and promising research field. The strong driving force of business interests will continue to promote its development. On the basis of summarizing the research progress of data mining and association rule mining. In order to better deal with large-scale data, improve the effectiveness of association rule mining algorithm and solve the problem of memory self limitation, the data is preprocessed. It includes attribute correlation calculation and data segmentation based on metadata. Based on association rule mining, this paper studies the problems existing in association rule mining in distributed system, improves the distributed mining algorithm, designs an efficient and available algorithm, and effectively solves the bottleneck of distributed data mining. Using the characteristics of association rule mining, a classification tree is constructed based on association rules, and the performance of the tree is analyzed and verified. Finally, the landslide monitoring data mining is used to carry out experiments on data extraction and cleaning. Compared with the practical environment, the reliability of association rule mining is verified. New data mining methods and models come out every year, and people's research on it is becoming more and more extensive and in-depth. Therefore, in terms of the application of the Internet of things, how to realize the effective utilization of such massive data information has become an important problem. Because data mining technology can effectively solve this problem, it must be widely used in the Internet of things.

Acknowledgement

Exploration and Practice of the reform of the Curriculum system of Internet of Things Engineering based on project-based hierarchical and progressive approach. Project No. : HKKG2021-01.

References

[1] Gu Weijie, Wang Xiaofeng. Research on rule system of coal mine Internet of Things based on improved Apriori algorithm . Coal mine machinery, vol.37, no.1, pp.3, 2016.

[2] Luo Haiyang. Research and application of association rule mining algorithm based on MapReduce . Information and Communication, no.2, pp.3, 2019.

[3] Luo Li. Research on high-frequency data item mining algorithm of network information big data based on improved BigFIM algorithm . Journal of Laser, vol.37, no.7, pp.6, 2016.

[4] Li Chao, Zhao Shuliang, Zhao Junpeng, et al. Scaling up algorithm of multi-scale association rules . Computer Science, vol.44, no.8, pp.5, 2017.

[5] Ma Lihong, Qin Dan, Lin Qiang, et al. Intelligent extraction of key indicators of distribution network project based on NLP method . Automation and Instrumentation,no. 4, pp.5, 2019.

[6] Bai Miaoqing. Research on Big Data Mining in Radio and Television in \ “internet plus \” Era . Internet of Things Technology, vol.8, no.10, pp.2, 2018.

[7] Wu Haozhong, Qian Xuezhong. Improvement of Apriori algorithm using Boolean mapping matrix . Fujian Computer, vol.36, no.3, pp.4, 2020.

[8] Jia Lili. Research on fragmented knowledge mining and intelligent reasoning methods . Science and Technology Communication, vol.12, no.2, pp.3, 2020.

[9] Sun Xiuchun, Lian Xinyu. Research and practice of innovation and entrepreneurship education model of Internet of Things major in higher vocational colleges based on big data analysis . Human resources development, no.4, pp.1, 2017.

[10] Bai Yuxiu, Zheng Huanhuan. Research on the application of mobile GPRS service association rules based on CRISP-DM model . Internet of Things Technology, vol.7, no.3, pp.4, 2017.

[11] Zhao Xuejian, Zhixin Sun, Yuan Yuan, et al. An improved Apriori algorithm for orthogonal linked list storage . Miniature microcomputer system, vol.37, no.10, pp.5, 2016.

[12] Li Dechen, Lu Yifan, Zhao Xuejian. A frequent itemset mining algorithm based on predictive filtering . Computer Technology and Development, vol.28, no.5, pp.4, 2018.

[13] Qin Ming, Cai Ming. Semantic annotation of images based on classification fusion and association rule mining . Computer Engineering and Science, vol.40, no.5, pp.7,2018.