Application of big data collection-analysis-visualization in the teaching process of colleges and universities under the background of the epidemic

Jin Zhang1*

1 College of Computer Science and Design, Guangdong Mechanical & Electrical Polytechnic, Guangzhou, 510700, P.R.China

*Corresponding author’s e-mail: zhangjin0023@gdmec.edu.cn (ORCID: 0000-0002-4495-0003)

Abstract. With the advent of the era of big data, new vision and broad ideas have been provided for college teaching. Using big data-related knowledge to study the teaching system of universities can not only promote the implementation of big data teaching in universities, but also effectively improve the quality of university teaching. This article discusses the technology related to the teaching big data cloud platform. This article introduces the role and application of the association rule algorithm in the process of big data mining, and according to the actual situation, studies the use of this algorithm in the process of big data mining in colleges and universities. In order to further reflect the advantages of its algorithm, it is compared with other algorithms, which proves that the algorithm in this paper has higher execution efficiency and accuracy.

1. Introduction

With the rise of big data applications, automated teaching in colleges and universities becomes possible. Big data can provide a large amount of data support for teaching, making teaching more scientific and fair [1]. Teaching under the drive of big data can provide more timely feedback on the results of teaching. Big data has challenged the current teaching system in terms of technology and system. Therefore, it is necessary to explore the construction of a university teaching system that meets the requirements of the big data era.

Due to the rapid development of network services such as social networking sites and e-commerce, the scale of network services and network information of various universities has increased dramatically, which has brought greater challenges to the processing of big data. At the same time, the financial industry, telecommunications, medical treatment, retail industry and other fields will also generate a large amount of data [2]. The Open University of the United Kingdom has published the Open Learn project and created an open teaching resource website that allows learners to obtain learning resources for free [3]. The website is mainly composed of two parts: learning space and experimental space. The learning space can provide up to 5400 hours of learning content for learners who visit the website; the experimental space is open to teaching staff, which includes a learning inventory of 8,100 hours. In addition, learners and educators can also learn and communicate through this website. The Open University of the United Kingdom cooperates with Yale University, Stanford University and other universities in the United States to upload learning resources to i-tune U channel to benefit learners around the world [4]. The sharing of teaching resources is not limited to a single country, but also...
extends to cooperation and sharing between countries, such as the German, Austrian, and Swiss Library Association organizations that exist in Europe. The alliance was established in 2000 and was first initiated by Germany, and then Austria and Switzerland also participated in the formation of the alliance to jointly study the archiving and storage of books under scientific development and the future open library joint mechanism.

Cloud computing is an on-demand, convenient, and pay-as-you-go network access mode. It can share and quickly provide configurable computing resources such as servers, networks, application networks, and storage. The manager’s investment time and management cost are very low, and the desired effect can be obtained. The characteristics and advantages of cloud computing make it very suitable for education. Cloud computing not only makes school data storage reliable and safe, but also makes data sharing more convenient. Old school computers, students’ entry-level laptops, and even smart phones can quickly and easily access cloud services through the web front end. With the improvement of resource utilization and management level, the school can also greatly save the cost of information infrastructure construction [5]. Facing hundreds of institutions of higher learning across the country, how to make big data provide them with useful value information is an important research direction [6]. At present, relying only on the employment information or the announcement of course results on the school's webpage obviously cannot fully utilize the powerful information mining capabilities of big data. The core of the big data cloud platform architecture for university teaching management is the Hadoop platform, which is also the focus of our work. Hadoop is an open source distributed computing framework that can be used to build a big data cloud platform for university teaching management. This article explains the problem of big data association rule algorithm and proposes a general solution, and further we obtain the applicable big data association rule algorithm data set. According to the Apriori algorithm and the improved association rule algorithm, the performance of this type of algorithm is further analyzed. The results show that the association rule algorithm of this article is maintained under the premise of maintaining the accuracy of the algorithm. When dealing with big data, the effect is obviously better than other algorithms for processing big data. Then, with the background knowledge mastered, the feasibility study of the big data mining algorithm in this article is carried out in the application prospects of universities.

2. Architecture of a big data cloud platform for university teaching management under the background of the epidemic

2.1. Hadoop-based university teaching management big data cloud platform architecture

Figure 1 illustrates the hierarchical architecture of the university teaching management big data cloud platform. The resource layer includes server resources, network resources, storage resources, and physical resources. The platform layer includes middleware services, database services, and the application layer includes enterprise application services and personal application services.

![Figure 1. Overall architecture of the big data cloud platform for university teaching management](image)

It consists of a management node and a large number of data storage nodes. Hadoop Distributed File System (HDFS) manages and stores all files on the big data platform. These files are responsible for storing and managing data in the Hadoop cluster. The key element in Hadoop is the engine responsible for data-parallel calculation. It is mainly composed of Job Trackers nodes and Task Trackers nodes. For external client clients, the distributed file system HDFS can create, delete, move or rename files, but the structure of the distributed file system HDFS is about a group of specific nodes, which include
a Name for managing data, its location in HDFS is mainly responsible for providing metadata services, and the data node in HDFS is responsible for providing storage space. However, there is only one Name Node, which is prone to single-point failure errors. A large amount of file data is divided into data blocks and stored in HDFS. A large number of Data Nodes are responsible for storing these data blocks. The block is generally 64MB in size, and the client decides the number of data blocks copied when the file is created.

2.2. Architecture of big data cloud platform for university teaching management

The logical topology of the operating mode of the proposed system is shown in Figure 2. It is a software framework that can be used to process large-scale data in parallel. The key idea of the mapping-reduction model Map Reduce is the principle of proximity, computing the closest data storage node. In the face of all massive large-scale data processing technologies, because the cost of using the network for data transmission is greater than the cost of computing nodes overhead, so the data-parallel processing model moves the data calculation processing task to the node where the data is stored, staggering the data transmission overhead, and reducing the overall data calculation overhead.

Student self-evaluation refers to students’ understanding of the quality of self-learning, that is, students’ self-knowledge of the learning process. Self-evaluation helps students understand the problems in the learning process, thereby improving their own learning methods and improving the quality of learning. Students conduct self-inspection, summary and evaluation based on evaluation indicators. In order to recognize your own strengths and weaknesses, you give full play to your subjective initiative in the future learning process, and promote your own learning progress. Students can self-evaluate their class status by logging into their own evaluation account.

Group evaluation is to divide the students into groups equally, and the group members use one-to-one evaluation methods to evaluate according to the indicators. This can not only stimulate the enthusiasm of students to learn better, and promote mutual learning between students, but also learn from the learning methods of others and improve their learning methods. Therefore, taking the form of group evaluation can collect more information about students' development, changes and progress in the learning process. The teacher enters the group information in advance, and the students log in to the evaluation account to evaluate other group members in the same group.

Students and teachers are the direct contacts for the success or failure of teaching, and students’ evaluations of teachers are the most convincing. Therefore, students’ evaluation activities cannot be ignored. When students evaluate teaching, we must be clear that the student is the subject and the teacher is the object, the student is the subject, and the main action is used for the teacher. Students can evaluate the teacher's class by logging into their own evaluation account.

Peer evaluation is a kind of understanding of teachers' teaching quality by peer teachers. In the evaluation process, peer teachers are the subject and the teacher being evaluated is the object. In peer evaluation, peer teachers cannot rely on subjective experience to evaluate, let alone personal feelings and other non-teaching factors to evaluate, thus ignoring specific teaching classroom surveys. Peer teachers can evaluate teachers' teaching in the way of auditing.
3. Performance analysis and application of association rules for college teaching big data

3.1. Performance analysis of association rule algorithms for big data

The processing, perception, storage, service and acquisition of "big data" are all facing huge challenges. How to acquire "big data" and perform dynamic and efficient processing has become a technical difficulty. Because "big data" has the characteristics of untrustworthiness, heterogeneity and non-structure, effective management and in-depth analysis and research of "big data" need to solve a series of important issues such as the mining of "big data" for analysis and processing problem. It is very difficult to use a unified analysis model to process and analyze big data, because traditional data mining algorithms cannot directly analyze and obtain the potential value of big data. In order to make full use of data resources and dig out the potential value of big data, it is necessary to find the most suitable and effective mining method to analyze and process big data. There are many methods in this area, but they cannot effectively solve the problem of big data mining. Therefore, research on association rule algorithms that are both effective and low-complexity is one of the key technical problems we face when we process big data.

It can be seen from Figure 3 that the running time of the Apriori algorithm and the algorithm proposed in this article is constantly increasing when the minimum support remains unchanged while the average length of transactions, the total number of projects, and the number of transactions have increased.
Figure 3. The first set of experimental results

The main reason why the association rule algorithm of big data is more efficient than the Apriori algorithm is that the association rule algorithm of big data proposes a new storage structure. When using this storage structure to count candidate item sets, only the ordered list is required. From another two sets of experiments, it can be seen that the data source used in each experiment is the same, but due to the continuous decrease of the minimum support, the candidate item set increases, and the time consumption remains basically unchanged, while the Apriori algorithm has greatly increased.

3.2. Application of big data association rule algorithms to universities under the background of the epidemic

In previous education, teachers made teaching policies based on their own teaching experience. Now, we can conduct "big data" association rule mining on the teaching content of teachers, and through in-depth analysis of this massive teaching content information, we can understand which content students are more interested in and have insufficient mastery of teaching content. The teacher will improve the teaching plan in future teaching to determine the new teaching content. This will not only promote teachers to reform teaching methods, but also stimulate students' interest in learning, help students better tap their potential abilities, and create more and better value for society.

The reason why big data is so valued by people is that it is possible to predict the probability of something in the future through the combination of big data and mathematical thinking mode. Why WeChat can recommend friends we may know, why every time you log in to Amazon again, it will recommend books you like to read, why every time you shop online, it will be based on your previous browsing history to recommend you the products you love with the right price range, and the personalized advertising breaks in the customized video website. These massive data provide a solid foundation for accurate forecasting. As university students establish more and more relationships, the connections between them are getting closer, and the amount of data that the university system can receive is increasing. The same technology can be used to predict the source of enrolment and the future development of the university. The experimental results of the number of association rules between courses are shown in Figure 4.
As shown in Figure 4, they respectively represent the invalid association rules mined by the algorithm, the association rules that are not of interest to users, and the strong association rules. According to the analysis of the number of associations in the above figure, we can see that according to the inference algorithm and course-related attributes (time), uninteresting rules are deleted, and some association rules are deleted for validity. Invalid association rules and association rules that are not of interest to users have a larger proportion. You need to delete the above rules to further better analyze the rules that users care about.

In order to maximize the value of the association rule algorithm in the process of big data mining, it is necessary to fully understand the relationship between big data and its association rule algorithm in the process of big data mining, and then carefully analyze and study the definition of mining and the goal of mining. For example, in the big data of college informatization, suitable packages are formulated for different objects to facilitate use. Making full use of the association rule mining of big data can not only improve education management, but also contribute to the further enhancement of the implementation of digital education. Using the association rule algorithm of big data in teaching evaluation can relate its internal connections from different factors such as teacher teaching, the interaction between students and teachers, the use of multimedia courseware in teachers' classes, and the teaching environment of teachers. It can provide more support information for the teaching management department to make the next decision. Through big data, we can see the enthusiasm of students in class by understanding the interaction between students and teachers, so as to provide more valuable teaching programs for different students.

Based on "big data", the overall research direction of the internal structural characteristics of complex systems is likely to become a new way of thinking and approach to studying complex systems. The research of big data is different from traditional research methods. It is to perform related search, clustering, classification and other analysis and merging operations for huge data. Therefore, the research and analysis of big data have statistical characteristics. Statistics are concerned with whether the data are related. The so-called "correlation" refers to the law of a certain correlation between the values of two or more variables.

The data network often implies the commonality between the data and the characteristics of the integrity of the entire network, and big data exists in the form of such a complex linked data network, so we must make full use of the internal existence value of big data. It is necessary to conduct an in-depth analysis of the network behind the big data. The network also has its own average path length, degree analysis, betweenness and many other attribute descriptors and unique properties. These parameters and properties may be able to successfully describe the network behind the "big data".

![Figure 4. Number of association rules](image-url)
4. Conclusion
In the context of the epidemic, the distributed file system HDFS based on the open-source framework for big data processing uses multiple copies to implement redundant data storage and realize the secure storage of data in Hadoop. At the same time, the key component of the parallel computing model (mapping-protocol) Map/Reduce provides the capability of data-parallel processing. The most critical components of the open-source distributed framework Hadoop are Map Reduce and HDFS. The idea of Map Reduce can be understood as decomposing tasks and summarizing results. We experiment and compare the application of association rule algorithms in big data mining, fully understand big data and its association rule algorithms, and clearly understand and define the goals of mining. It also explains the use of big data in the informatization of colleges and universities, and makes suitable packages for different objects to facilitate use. Finally, a few points that should be paid attention to when using big data are briefly explained.

References
[1] Tan, X.(2019) Research on College English Writing Teaching under the Background of Big Data: Taking Leshan Normal University as an Example. Theory and Practice in Language Studies, 9(1):60-66.
[2] Cui, L., Liao, Y., Wang, Y., et al.(2019) College Mathematics Teaching Method Based on Big Data. International Journal of Emerging Technologies in Learning (iJET), 14(13): 47-58.
[3] Jiang, D., Wang, Y., Lv, Z., et al.(2019)Big Data Analysis Based Network Behavior Insight of Cellular Networks for Industry 4.0 Applications. IEEE Transactions on Industrial Informatics, 16(2): 1310-1320.
[4] Wen, J., Zhang, W., Shu, W.(2019) A Cognitive Learning Model in Distance Education of Higher Education Institutions Based on Chaos Optimization in Big Data Environment. The Journal of Supercomputing, 75(2):719-731.
[5] Jan, B., Farman, H., Khan, M., et al.(2019) Designing a Smart Transportation System: An Internet of Things and Big Data Approach. IEEE Wireless Communications, 26(4): 73-79.
[6] Regan, P. M., Jesse, J.(2019) Ethical Challenges of Edtech, Big Data and Personalized Learning: Twenty-First Century Student Sorting and Tracking. Ethics and Information Technology, 21(3) 167-179.