Java Web Cloud Data Analysis and Application Based on Spark Machine Learning Algorithm

Pengfei Che*

School of Information Engineering, Xi'an University, No.1 Keji Sixth Road, Xi'an, 710065

*Corresponding author e-mail: cargeapu@163.com

Abstract. The rise of information technology has promoted the rapid production of all kinds of information data, with the rapid growth of information data, the problem of data scalability began to appear, the system of data system time and space complexity will be more and more large, which brings great difficulties to the processing of information data. Java web cloud data based on Spark machine learning algorithm provides a new way for information data processing. The purpose of this paper is to analyze Java web cloud data, explore its application, and realize deep data mining and fast processing. This paper gives an overview and analysis of Java web cloud data and Spark machine learning algorithm. It attempts to build a Java web cloud data mining system with the help of Spark machine learning algorithm. First, it explains the basic design principle of the system, then points out the construction framework of the system, and finally carries out a simulation experiment on the system. Experimental results show that the application of Java web cloud data in data mining system is feasible and scientific, which not only changes the way of data storage, but also improves the overall efficiency of data processing.

Key words: Spark Machine Learning Algorithm, Java web Cloud Data, Data Mining, Analysis of the Application

1. Introduction

At present, with the rapid development of social and Internet technology, the demand for information processing is expanding. Taking China's e-commerce as an example, the data information of e-commerce presents a linear rising trend every year, and the complexity of data information is also increasing. Based on this situation, traditional information processing methods can no longer meet the needs of complex mass information processing. Java web cloud data can realize the storage and processing of massive complex data, greatly improving the speed and accuracy of data processing, and providing a new path for data processing. It is of great significance to analyze Java web cloud data and
study its application for future information processing.

Spark machine learning algorithm is a common big data processing algorithm, which has been deeply studied by scholars at home and abroad in recent years. In [1], the author proposed a nonlinear machine learning algorithm under Spark distributed memory environment, and optimized it in data compression, loading, bias and other aspects. Under this optimization, the calculation practice and accuracy of the algorithm were greatly improved. In [2], the author proposed a hybrid recommendation algorithm based on Spark machine learning algorithm based on the research of traditional recommendation algorithm, and tried to apply the algorithm to various types of e-commerce websites. In [3], the author emphatically analyzes the application value of Spark machine learning algorithm in data classification. Experiments prove that this algorithm can optimize data classification standards, improve data classification accuracy, predict future data classification, and propose several specific application measures.

Compared with the traditional data storage and processing technology, Javaweb has obvious advantages and has gradually become the focus of scholars at home and abroad. In [4], the author deeply analyzes the framework structure and principle of Java web, and puts forward the handling method of Java web error exception. In [5], the author explored the application of cloud storage in Java web technology, explained the development environment and storage requirements of cloud storage, and designed and realized the basic functions of the system such as user management and file management with the help of Java web technology. In [6], the author combines cloud computing with Java web technology to build the relevant cloud storage system, and verifies the feasibility of the system with the help of simulation experiments.

To sum up, the current researches are basically independent researches on Spark machine learning algorithm and Java web technology. Few researches use Spark machine learning algorithm to analyze and explore the application of Java web cloud data [7]. To make up for the gap in terms of theory, this paper Java web cloud data and Spark machine learning algorithms has carried on the summary of related analysis, using the Spark machine learning algorithms attempt to establish the Java web cloud data mining system, first of all, the basic design principle of the system is illustrated, and then points out the frame of the system, finally the system simulation experiment [8-9]. The research in this paper not only promotes the application range of Java web cloud data, but also lays a theoretical foundation for the related research in the future.

2. Method

2.1. Java web cloud data

Java web is the sum of the Web technology problems that can be solved on the basis of Java. Since the web has an obvious hierarchical structure and the server client is separated from each other, it is more suitable to develop and design the Internet web with the help of Java language [10]. The first to invent the Java language is James green, the language to achieve the design of realistic thinking and programming ideas, incisively and vividly give play to the idea of object-oriented. Traditional development is process-oriented in thinking, and there is an essential difference between the two. The encapsulation and inheritance of Java language are relatively good, which greatly reduces the coupling between different modules. This programming philosophy has greatly improved the program's
scalability and maintainability. Generally speaking, Java web is based on MVC design framework as its framework structure [11-12]. At present, the Java language and the web have produced a relatively harmonious chemistry, in which also produced a lot of development tools and ideas for the web. The above advantages meet the requirements of cloud storage system, such as device interconnection and device increase and decrease demand, Java web cloud storage system is mainly used for data storage, namely Java web cloud data. Java web cloud data has obvious advantages, which can realize fast and deep mining of data, so as to help users quickly find targeted products and customers according to their own needs [13]. Java web cloud data is characterized by complexity and diversity. At present, Java web cloud data is mainly applied in the field of data mining.

2.2. Spark machine learning algorithm
Spark is a big data processing engine that can process all kinds of data in parallel under the function of distributed data sets. The advantages of Spark in memory computing greatly improve the speed of data processing, about 100 times faster than traditional data processing engines. The computing framework adopted by Spark machine learning algorithm is distributed. The specific computing environment includes YARN, Apache Mesos, EC2, etc. Elastic distributed data set is the core of this algorithm. Therefore, the characteristics of Spark machine learning algorithm are consistent with the characteristics of Java web cloud data, so Spark machine learning algorithm can be used to process cloud data. The key to Spark's machine learning algorithm is data compression. Data compression requires first determining the type of data set, then converting according to the specific type, and finally data compression. Data compression usually relies on the dimensionality reduction formula of SVD. The specific formula is as follows:

\[ A_{m \times n} \approx U_{m \times k} \sum_{k=1}^{k \times n} V_{k \times n} \quad (1) \]

Through equation (1), the dimension of the data is reduced. When the result of a certain value pair of the data produces a slight image, the characteristic value can be eliminated. In the equation, k represents the sequence of values. After data compression is completed, sampling processing is required. The specific calculation formula is as follows:

\[ k(x_1, x_d) = \left( \frac{3}{4} \right)^d \frac{1}{B_1 B_2 \ldots B_d} \prod_{1 \leq i \leq d} \left(1 - \left(\frac{x_i}{B_i}\right)^2\right) \quad (2) \]

In formula (2), Bi represents the bandwidth with the sequence of eigenvalues of I, d represents the number of eigenvalues, and xi represents the eigenvalues with the sequence of I.

3. Overall framework of Java web cloud data mining system
As mentioned above, Java web cloud data is mainly applied in the field of data mining. Existing research data show that the application of Java web cloud data in data mining system not only promotes the modularization of data, but also improves the extensibility of data, and most importantly, the maintenance of data is realized. By referring to relevant materials, this paper applies this data to the data mining system with the help of Spark machine learning algorithm. Under the data mining system, users can quickly find targeted products and customers according to their own needs. Based on the trend of deep computing of Internet data, the Java web cloud data mining system proposed in this paper has realized the combination of deep computing and wide computing characteristics, and mined and processed data on the Windows platform with the help of B/S, Struts2 framework and DM
and OLAP data analysis technology of Eclipse development and integration application. In general, Javaweb cloud data mining system builds a four-tier architecture based on the WEB, and builds a system service module on truts2 and spring framework. The analysis process of data mining under the general framework mainly includes two parts, namely data training and data prediction.

4. Discuss

4.1. Specific design of java web cloud data mining system
(1) Architecture
Already mentioned above, Java web cloud data mining system based on WEB built four layers architecture, the architecture of the specific as follows: the first layer architecture is the client's WEB browser, Internet users to use version 5.5 or more search browser to be able to access, represented by Google bearing the core IE browser can; The second layer is the Web server, which USES Windows xp for its operating system and Tomcat6.02 for its Java server. The third layer system is the application server, which has a wide range of types, including various Web Service services and data mining engines. The fourth layer USES an OLAP server that provides cube data through MySQL and Analysis services.

(2) Cloud data storage
The computing framework of Spark machine learning algorithm is distributed, so the cloud data of this system is stored in a distributed way, which is mainly composed of HDFS and HBase. System users upload semi-structured data to HDFS. HBase module is mainly used for storing real-time data. The data of this module is constantly updated with the change of time, and it can access and read and write cloud data at any time, so as to realize the rapid mining and display of real-time cloud data. Cloud data storage module is the core of statistical test data services, the service function of diversity, including the characteristics of the data extraction, data statistical analysis, data association, etc., in statistical test has been completed, can use Spark machine learning packages are distributed as, at a deeper level training and projections for cloud data, finally to obtain result set HBase module, can realize the cloud data real-time visualization, ensure the timely grasp of cloud data.

4.2. Application testing of java web cloud data
In order to ensure the accuracy of Java web cloud data application, it is necessary to carry out feasibility test for the cloud data mining system designed in this paper. This paper USES matalb software to carry out simulation application test. The specific test data are shown in figure 1 and figure 1 below.

| Table 1. Operation data of Java web cloud data mining system |
|------------------|------------------|------------------|
| Run program     | Data             | Composite scores |
| Processing time | 5.12s            | 21.56%           |
| Mining rate     | 98.72%           | 26.99%           |
| Accuracy        | 97.91%           | 30.90%           |
| Error rate      | 2.09%            | 20.54%           |

*Data came from experimental analysis results
Figure 1. Comparison between Java web cloud data mining and traditional data mining

The data in table 1 shows that the data processing time of Java web cloud data mining system proposed in this paper is only 5.12s, and 98.72% data information can be mined in a short time. The accuracy rate of data processing is 97.91%, and the error rate is only 2.09%. This shows that the system can realize the efficient processing of data in a short time, taking into account the comprehensiveness and accuracy of data processing. The data in figure 2 shows that compared with traditional data mining technologies, Java web cloud data mining is far more efficient and accurate in data mining, exceeding about 40%. To sum up, the Java web cloud data mining system designed in this paper is feasible and scientific, which indicates that Java web cloud data can be applied in the field of data mining.

5. Conclusion

In this paper, Java web cloud data based on Spark machine learning algorithm is analyzed in detail, and its application is studied. Finally, a Java web cloud data mining system is built on the basis of Spark machine learning algorithm, realizing the application of Java web cloud data in the field of data mining. The research shows that Java web cloud data mining system has incomparable advantages over traditional data mining system, which not only realizes the storage and processing of massive data, but also greatly improves the convenience and accuracy of data storage.

References

[1] Xu xin-rui, meng caixia, zhou wen. A recommendation algorithm based on Spark aging collaborative filtering[J]. Computer technology and development, 2017, 11(6):48-55.

[2] He qing, li ning, luo wenjuan. Overview of machine learning algorithms based on big data[J]. Pattern recognition and artificial intelligence, 2017, 27(4): 327-336.
[3] Wang jialin, duan zhihua. Mechanism analysis and performance tuning of Spark kernel[J]. Machinery industry, 2017, 23(5):117.

[4] Zhou wei. Research and implementation of MVC design pattern of computer JavaWeb software[J]. Computer nerd, 2017, 13(12):33-57.

[5] Yang ning, liu danjun. JavaWeb application development research based on JFinal framework[J]. Computer knowledge and technology, 2017, 10(7):1440-1443.

[6] Zhang xuemin. A brief discussion on the role of MVC design pattern in JavaWeb[J]. Journal of science and technology, 2018, 13(22): 58-59.

[7] Yang guoxun. Research and application of JavaWeb application framework[J]. Science and technology communication, 2017, 9(21): 90-91.

[8] Gao jin. Analysis of Java SSH framework in Java web[J]. Journal of science and technology, 2018, 13(23): 48.

[9] Ole Marius Hoel Rindal, Trine M. Seeberg, Johannes Tjønnås. Automatic Classification of Sub-Techniques in Classical Cross-Country Skiing Using a Machine Learning Algorithm on Micro-Sensor Data[J]. Sensors, 2017, 18(1):75.

[10] Feras Hatib, Zhongping Jian, Sai Buddi. Machine-learning Algorithm to Predict Hypotension Based on High-fidelity Arterial Pressure Waveform Analysis[J]. Anesthesiology, 2018, 129(4):1.

[11] Mahdi Nasiri, Behrouz Minaei, Zeinab Sharifi. Adjusting data sparsity problem using linear algebra and machine learning algorithm[J]. Applied Soft Computing, 2017, 13(2):61.

[12] George S Fanourgakis, Konstantinos Gkagkas, Emmanuel Tylianakis. A Robust Machine Learning Algorithm for the Prediction of Methane Adsorption in Nanoporous Materials[J]. The Journal of Physical Chemistry A, 2019, 123(28):113.

[13] Christopher McMaster, David Liew, Claire Keith. Correction to: A Machine-Learning Algorithm to Optimise Automated Adverse Drug Reaction Detection from Clinical Coding[J]. Drug Safety, 2019, 42(6):24-27.