Research on Data Mining of the Internet of Things Based on Cloud Computing Platform

Wenqing Zhang
Foshan Polytechnic. 528000

Abstract: Based on the development of society and the progress of information technology, China's information industry has made great progress and has gradually become an important pillar of national economic development. In this context, the gradual integration of information technology had promoted the construction of the Internet of Things system, so as to promote the human life developed in the direction of modernization intelligently. At present, in the process of forming the development of the Internet of Things the first need to fully tap the data, which thus provide users with better service, for the development of large-scale development of the Internet. This paper analyzes the meaning of Internet of things, and discusses the characteristics of Internet of things and data mining, hoping to promote the improvement on the Internet of Things system in China, and thus promote the realization of higher efficiency.

1. Introduction
Based on the development and progress of the times, China's Internet of Things system has been established and perfected, and has promoted the development of our residents' lives in the direction of intelligence and modernization. At present, in order to ensure the sustainable development of Internet of Things system, China's technical staff has strengthened the rational use of the cloud computing platform, so as to ensure the steady development of data mining work. At present, the traditional mining algorithm can not effectively identify and quantify the abnormal data onto cloud computing environment, and increase the error of data mining, which hinders the sustainable development of Internet of Things. Based on this, this paper focuses on the analysis of Internet of things based on cloud computing data mining.

2. the Definition of the Internet of Things
On the definition of the Internet of Things, the academic community have existed a variety of different types of interpretation, as mentioned above, with the popular discourse to explain the Internet of things that is the link between the objects, of course, this explanation is too simple, it can not summarize the nature of things. On the concept of things, there are three kinds of argument are the most popular.

Massachusetts Institute of Technology, the relevant research center of 1999 has defined on the concept of things. The research center pointed out that the Internet of Things is essentially through the RFID and other information sensing equipment and to connect the Internet items, and do the intelligent identification and management on the implementation of these items. In other words, the essence of things is the Internet as the center, and rely on RFID technology for the development of information carrier.

In addition, the EU as an important economic entity in the world, the concept of the Internet of Things has put forward their own views, the EU thinks that the Internet is part of the future of the
Internet, and it has physical properties, anthropomorphic properties and other characteristics.

In addition to the definitions given by the two organizations mentioned above, the International Telecommunication Union also interpreted the definition of Internet of Things in 2005. It believes that the Internet of Things is based on the global positioning system and other information sensing equipment, with the help of network protocols, linked the items and items, and implemented the intelligent monitoring and management model.

By summing up the above three concepts we can see that the Internet of Things as a carrier of information, the foundation and the core of its development is the Internet, and its development process needs to use other information technology, and then do the Intelligent monitoring, tracking and management on the information of the items. On the concept of things, I believe that it is through the information sensing equipment and the Internet as the basis, linked the physical, information space, and through the use of other information technology means to achieve the digital link between the items.

3. Cloud Computing Technology Overview

The so-called cloud computing, refers to the calculation method based on Internet technology. The technology in the implementation process is often providing users with the relevant computing services through the Internet heterogeneous, autonomous services. As all the service process and the results are on the Internet chivalry, so called this calculation method as cloud computing technology. All in all, cloud computing refers to the use of the Internet to provide users with services to dynamically scalable virtualization of the computing model.

4. The Data Layer of the Internet of things

As an important part of the data mining platform of the Internet of Things, the construction and improvement in / to / on the data layer can adapt to all kinds of data heterogeneity and mass, so as to bring the data storage and processing efficiency of the Internet of things to improve the data mining and sexual enhancement.

At present, the technical staff in the data layer construction process, mainly divided into two important modules, respectively; data source conversion module, distributed storage module. The so-called data source conversion module, which in the process of running the object can be all kinds of heterogeneous data in the effective conversion. Under normal circumstances, the technical staff with Hadoop platform file system for specific operations, and the use of distributed way to store mass data.

In the process of running the Internet of Things, we tend to use different data types of object representation, and sometimes we will use different data to represent. Based on this, the technical staff need to use the data source converter, the heterogeneity of things on the issue of a comprehensive summary of the data storage to ensure the integrity of the upgrade, and promote the smooth development of data mining work.

In addition, the emergence of data sources conversion module is often the equivalent of data layer and the interface between the sensing layer, which can guarantee the smooth development of data decoding and other work, and to protect the integrity of the data. Since the data converter can convert different types of data onto PML data onto the running process, the file type distributed among each Name Node is PML type data. The so-called PML, refers to the use of a common way of contempt for natural objects, the technology in the implementation process is often required to create XML language. At present, the main reason for the technical staff to proceed with PML are to provide detailed information about the items and to facilitate the exchange of items. The Internet of things still is information collection process, for example, the Internet of Things in the process of running the data has achieved the collection of data, and then to carry out the transmission, and finally the use of PML for modeling work. Under normal circumstances, the model of information mainly contains information on the various types of information, and the summary of these informations is in order to achieve a comprehensive description of the information on the items.
5. Data Mining Service Layer

In general, the data mining service layer in the process of running mainly divided into three modules, namely: data preparation module, data mining engine module, user module.

Among them, the data preparation module in the process of running to help achieve the data clean-up and change, and data mining engine module in the process of running mainly contains data mining algorithm set, the model evaluation and other content, and the user module in the process of running In the realization of the data mining knowledge visualization of the representation. In fact, based on the difference between / in the actual data mining type, the data mining engine module in the implementation process helps to promote the evolution analysis, deviation analysis, similarity analysis and other classification work carried out.

As the main platform for the interaction between the data mining platform and the personnel, the user module has strong friendliness in the design and application process, and the user can also carry out the data mining task through the interface operation. In recent years, in order to further to enhance the portability of the cloud computing platform, the staff in the process of processing to increase the related open interface, so that make the third-party can use the Internet Data Mining platform features, in order to make the application of things more abundant.

6. Decision Tree Classification Algorithm

In order to further to promote the steady development of data mining work based on cloud computing platforms, technicians need to choose a reasonable data classification algorithm for a large number of data analysis, classification processing. At present, the most common data classification algorithm is the decision tree classification algorithm. At present, the decision tree taxonomy is called the greedy algorithm in the process of implementing the operation. The algorithm in the process of data classification process is often taken from top to bottom of the division of the specific operation. The related data analysis practice shows that the classification algorithm is often able to derive the classification rule of decision tree in the process of running the chaotic data resources. In fact, this classification algorithm is based on the instance of the relevant data classification and processing. Because of its practical application in the process of noise data processing has a good robustness, and thus gradually become the field in the process of data classification process, the most common adopted algorithm.

In the process of building the decision tree algorithm, each node represents a test of a property, and the division system represents the data test output. In the process of analyzing unknown data samples, the method of comparing the attribute values of the samples of the decision trees is often taken. In order to more intuitive to understand the decision tree algorithm formation, the operation process, the author carried out the relevant algorithm flow chart, the specific content is shown in Figure 1.

![Decision Tree Diagram](image)

Figure 1. The specific content

In fact, the traditional model of decision tree algorithm is C4.5 algorithm, which has the advantages of simple rules and convenient operation, but with the increasing popularity and application of computer
technology in recent years, the data needs to be processed is gradually increased, while C4.5 algorithm can only deal with the smaller amount of memory data, in the process of sorting large quantities of data often appear the situation that run the algorithm blocked but can not continue to run the . And this situation also led to the C4.5 algorithm that can not adapt to the current needs of the classification of data processing, and gradually withdraw from the data classification processing work. However, with the relevant personnel in combination with the needs of the development of the times, the C4.5 algorithm has been improved and broken on the basis of the original basis, and promoted the SLIQ algorithm with the C4.5 algorithm and the birth of the SPRINT algorithm. The improved decision tree algorithm adapts to the need of large data mining, and promotes the sorting and processing of data.

7. Conclusion:
In order to further promotes the efficiency of China's Internet of Things and improve the quality of life, to enhance the quality of life of our residents and the level of China's enterprises have to strengthen the data resources for the excavation and use. This paper focuses on the connotation of the Internet of things in China, and analyzes the connotation of cloud computing and data mining. I believe that with the implementation of relevant measures in place and the development of technology, China's Internet of Things technology will be developed by leaps and bounds, and thus satisfied with the needs of social life, in order to achieve social life in the direction of intelligent and modern.

References:
[1] Chu Cui-xia. Study on Data Mining of Internet of Things Based on Cloud Computing Platform [J]. Chinese Journal of Digital Technology & Applications, 2015, (1): 85.
[2] Cao Jian-chun, Zeng Yun. Design of Data Mining Model for Cloud Computing Internet of Things [J]. Journal of Ship Science and Technology, 2015, (6): 228-231.
[3] Cui Wen-bin, Wen Fu-jiang, Mu Shao-min, HAO Qing-bo. Have Support Vector Machine Based on Hadoop [J]. Journal of Computer Research and Development, 2014, (S2): 116-121.
[4] Lin Xin. Research on the construction of large data mining platform based on cloud computing [J]. Jiangdong Industrial Technology, 2015, (71): 104.
[5] Chen Da-feng. Study on key technology of data mining in Internet of things based on cloud computing [J]. New Technology & New Products, 2014, (23): 20.
[6] Zhang Yi, CUI Xiao-yan. Study on Data Mining of Internet of Things Based on Cloud Computing Platform [J], 2014, (1): 108-111.
[7] Cao Yunqin. Study on coal mine safety dynamic diagnosis system based on cloud computing and Internet of things [J]. Coal Science and Technology, 2016 (7).023.
[8] Chen Junli. Study on data mining of Internet of things based on cloud computing platform [J]. New China Communications, 2016, (21): 74-75.
[9] Tang Yongfeng. Study on data mining of Internet of things based on cloud computing platform [J]. Computer Knowledge and Technology, 2017, (7): 218-219. [10] Yu Yang. Construction of cloud computing based on the Internet of things operating platform analysis and research [J]. Computer Knowledge and Technology, 2013, (35): 8189-8191.