Research on the New Path of Internet of Things Data Mining under the Background of Cloud Computing

Li Zhang1,*

1Chongqing Vocational Institute of Safety & Technology, Wanzhou, Chongqing, 404100, China

*Corresponding author e-mail: zhangli@cqvist.net

Abstract. In the big data era, data mining technology has been widely used in various fields. The data and information on the Internet are extremely large, so this brings new challenges to data mining. This article mainly analyzes the IoT data mining model, which is based on cloud computing.

Keywords: Internet of Things, Cloud Computing, Data Mining

1. Introduction

With the continuous development of computers, the Internet of Things technology has begun to be applied to real life. At the same time, it also makes the Internet of Things technology the mainstream development trend. In this context, cloud computing provides more possibilities for the development of data mining technology in the Internet of Things. Cloud computing not only has a broader level of IT foundation, but also has a good level of data mining integration. Therefore, it is necessary to strengthen the application of cloud computing in the data mining of the Internet of Things, and it will lead the trend of social development.

2. The key issues of IoT data mining

2.1. The characteristics of data in the Internet of Things system

(1) The amount of data is large. Each Internet of Things system has thousands or even more sensing devices, and these sensing devices continuously transmit collected data to the data center. The data center not only needs to store the collected data currently received, but also needs to save historical data to support object status tracking, data statistical analysis and data mining. Therefore, the first key problem faced by data mining tasks in IoT systems is the large amount of data [1].
(2) The data type is complex. There are many types of objects monitored by the Internet of Things system, including transportation, biology, forests, and buildings. The information collected by different monitoring objects is different. For example, the traffic system needs to collect video information, and the medical monitoring system needs to collect physiological information such as pulse and blood pressure, as well as medical stereoscopic information. It can be seen that the types of data collected by the IoT system are complex, including text types, image types, and video types.

(3) The data is heterogeneous [2].

(4) Highly dynamic. With the increase of sensor nodes, the data collected by them must be inserted into the database. Similarly, when a sensor node is removed from the Internet of Things, the database should no longer record the data collected by the sensor node.

(5) Temporal and spatial characteristics. The sensor terminals of the IoT system are distributed in different regions, and the data collected by each sensor terminal reflects the status and other information of the monitored object at that moment. Perceived data is meaningful only in a specific time and a specific space. If it is not in this place or after this time, the meaning of the data may not be significant [3].

(6) Incompleteness. The sensor terminals work without manual monitoring. Each sensor terminal may be attacked by natural or human factors at any time, including lightning damage, artificial malicious damage, etc., resulting in incomplete data reception of the sensor terminal. On the other hand, although sensing terminals can be widely deployed in different geographic locations, they still cannot cover every corner. Therefore, incomplete spatial data collection is also one of the characteristics of IoT system data.

2.2. Characteristics of data mining technology

The characteristics of data mining technology are shown in Figure 1.

![Figure 1. Characteristics of data mining technology](image)

2.3. Requirements for data mining in the Internet of Things
(1) Real-time and efficient data mining. Any control terminal needs to analyze the environment in real time and make correct decisions. Therefore, real-time and high-efficiency are one of the most critical requirements for data mining in IoT systems.

(2) Distributed data mining. The Internet of Things computing devices and data are naturally distributed, and distributed parallel data mining has to be adopted.

(3) Data quality control. The storage and management of multi-source, multi-modal, multimedia, and multi-format data is an important guarantee for controlling data quality and obtaining real results [4].

(4) Decision control. The mined patterns, rules, and characteristic indicators are used for prediction, decision-making and control.

(5) Excavation task. Mainly include data extraction, classification prediction, clustering, association rule discovery, etc.

2.4. Challenges of data mining in the Internet of Things environment

(1) The choice of data mining algorithm. Choose the appropriate algorithm, and adopt the appropriate parallel strategy, and then the parallel efficiency can be improved. Therefore, the design of the algorithm becomes very important, the adjustment of the parameters becomes indispensable, and the adjustment of the parameters directly affects the final result.

(2) Uncertainty. First of all, the description of data mining tasks is uncertain, and data collection and preprocessing are also uncertain. Secondly, there are uncertainties in data mining methods and results. Finally, because the final mining target that each user pays attention to is different, this leads to uncertainty in the evaluation of the mining results. Uncertainty is the biggest challenge faced by data mining in IoT systems [5].

(3) Credibility and security. Doing data mining in a cloud computing environment will lead to data mining cloud service software credibility issues. The first is the correctness of the service and the security of the service; the second is the quality of the service, which is measured by three aspects: availability, reliability and high performance.

3. IoT data mining model based on cloud computing

The data mining model architecture of the Internet of Things based on cloud computing is divided into five layers, and the specific model structure is shown in Figure 2 [6].
4. Analysis of IoT data mining technology based on cloud computing platform

4.1. IoT perception layer

The perception layer of the Internet of Things realizes the perception function. The nodes realize data collection through sensors, cameras and other related equipment. The collected data needs to rely on the network communication equipment of the perception layer of the Internet of Things for centralized processing, and the required data is transferred to each node, and then transferred to the data processing center of the cloud computing platform through the transmission layer after centralized storage, realizing the functions of the entire perception layer [7].

4.2. IoT transport layer

This is the intermediate link of all data transmission, which covers devices and technologies such as sensors and wireless networks. Through the connection of multiple network devices, a highly efficient and seamless data transmission system is formed, which can transmit the data collected by the perception layer through the network. It is transmitted to the data processing center, thereby achieving the goal of all-round intercommunication and interconnection. Analyzing the actual work content, the point is to network monitoring and processing equipment with multiple attributes to achieve transmission efficiency and disseminate data information between devices and nodes [8].
4.3. Data layer

This layer is the core link. The Internet of Things itself has certain heterogeneity and mass characteristics. Therefore, the ability to store, process and analyze all the data collected by the device in the data layer is the focus. The data layer internally covers two main parts: data source conversion and storage. The data source conversion refers to the transformation of the heterogeneous data of the Internet of Things, and the storage refers to the HDFS system in the platform built by Hadoop. Distributed storage, so a large amount of data can be reliably stored in each data node.

In a specific environment, the same target will also select different data types for performance. Based on this data source conversion function is mainly to maintain performance Data integrity, while avoiding heterogeneous Internet of Things data from being damaged due to other uncertain factors in the transformation, thereby achieving the purpose of ensuring the reliability. The value of data source conversion in the entire system is mainly as the role of the connection line between the data layer and the perception layer [9]. Through the decoding and conversion of the data packet, the data of different attributes is converted into the required data type, and at the same time, it is distributed The method is stored in the data processing center.

4.4. Data mining service layer

This layer mainly includes the following modules. Data preparation module: This module covers the data situation, transformation, data regulation and other links. Data mining engine module: This module covers the data mining algorithm set, pattern evaluation and other links. User module: This module covers the visualization technology of data mining knowledge. The data mining engine module has features such as metaphysics, association, clustering, trend analysis, deviation analysis, and similarity analysis. The core link that provides the above-mentioned functions is the multiple functional algorithms of the algorithm set in the data mining module. The data mining algorithm in the Hadoop platform needs to be adjusted to a certain degree, that is, to achieve Parallel processing of algorithms [10].

5. Conclusion

In summary, the development of economic has also promoted the advancement of information technology. Therefore, the Internet of Things technology came into being and achieved unprecedented development. In the face of the current growing trend of massive heterogeneous data in the Internet of Things, the emergence of related data mining technologies has also been affected to a certain extent. For this reason, it is very necessary to form data mining technology, which is based on cloud computing.

References

[1] WANG Ruigang. Research on Main Characteristics and Basic Theory of Internet of Things [J]. Computer Science, 2012, 39(6A): 201-203.

[2] LI Zhining. Research Progress of Internet of Things Technology [J]. Computer Measurement and Control, 2012, 20(6): 1445-1448.

[3] ZHANG Hu. Construction of Data Mining Mode for Internet of Things Based on Cloud Computing [J]. Wireless Interconnection Technology, 2016, 2 (20) : 50-51.
[4] CHEN Dafeng. Research on Key Technologies of Data Mining for Internet of Things Based on Cloud Computing [J]. China New Technology and New Products, 2014, 5 (23) : 20.

[5] Li Li, Zhang Yuzhou, Jiang Keqin. An Improved Data Mining Algorithm for Internet of Things Based on Cloud Platform [J]. Journal of Anqing Normal University: Natural Science Edition, 2014, 7 (2) : 37-40.

[6] 51:2642.. Research on Data Mining Model of Internet of Things Based on Cloud Computing [J]. Software, 2017: 237-240.

[7] Bu Fanyu, Wang Xin, Zhang Qingchen. Data mining model of Internet of Things based on cloud computing [J]. Computer and Information Technology, 2012: 53-56.

[8] TANG Yongfeng. Research on Data Mining of Internet of Things Based on Cloud Computing Platform [J]. Computer Knowledge and Technology, 2017

[9] HUANG Xin. Research on Data Mining of Internet of Things Based on Cloud Computing Platform [J]. Shanxi Youth, 2017.

[10] CHU Cuixia. Research on Data Mining of Internet of Things Based on Cloud Computing Platform [J]. Digital Technology and Application, 2015: 95.