ISSN 2616-5775 Vol. 6, Issue 8: 82-88, DOI: 10.25236/AJCIS.2023.060810

Artificial Intelligence Algorithm and Device for Big Data Processing of the IoT System

Guangjian Ma

Henan Province IUR Artificial Intelligence Research Institute, Zhengzhou, Henan, 450046, China
lilili@ncwu.edu.cn

Abstract: *With the rapid development of the Internet of Things (IoT), there are more and more big data generated in the IoT system, which requires effective processing and analysis. Traditional data processing methods cannot meet the processing needs of big data in the IoT system, so it is necessary to study new big data processing technologies in the IoT system. This paper has proposed a big data algorithm, which uses data mining technology in big data to process sensor and device data. In the data pre-processing stage, the algorithm and device use data cleansing and other technologies to ensure data quality and reliability. In the feature extraction and selection stage, the algorithm and device adopt adaptive feature extraction and selection techniques to extract key features of the data and reduce the dimensionality and complexity of the data. In the experiment, this article tested and evaluated the algorithm to verify its performance. The experimental results showed that the F1 value of the model established in this study was 0.87, and the training time was the shortest, only 9 seconds. This algorithm and device can effectively improve the efficiency of data processing and analysis, as well as improve the accuracy and reliability of data processing. Compared with traditional data processing methods, this algorithm and device have better performance and application prospects. The algorithm and device also have good robustness and scalability, and can adapt to different data processing and analysis needs. The algorithm based on big data mining technology is an effective big data processing technology of the IoT system, which can improve the efficiency of data processing and analysis, and improve the accuracy and reliability of data processing.*

Keywords: *Internet of Things System, Big Data Technology, Artificial Intelligence, Data Mining Technology*

1. Introduction

In recent years, more and more devices, sensors, and terminal devices have been connected to the Internet, generating a large amount of data. These data not only have diversity and complexity, but also have the characteristics of high speed and large capacity. Traditional data processing and analysis methods can no longer meet the requirements of real-time and accuracy. Therefore, how to deal with and analyze these big data has become an important issue in the research of IoT system. In order to solve this problem, this paper proposes a big data processing device of the IoT system based on artificial intelligence algorithm.

In the past decades, big data processing and artificial intelligence technology have been widely used in academia and industry. In the IoT system, big data processing and artificial intelligence technology are also widely studied and applied. Some scholars have proposed artificial intelligence algorithms such as data mining, machine learning, and deep learning to solve the big data processing problem in the IoT system. For example, the scholar Aryal A confirmed the main research trends of big data analysis and the IoT through empirical analysis, and deeply understood the changes in understanding and implementation of disruptive technologies, especially big data analysis and the IoT over time [1]. The scholar Daissaoui A and others summarized the work related to the IoT and big data analysis in intelligent buildings, aiming at improving productivity, safety and implementation quality, sustainable development, collaboration and sustainable smart city solutions [2]. The scholar Wan S proposed that large amount of data and wide coverage have always been the key issues in big data processing of the IoT, and designed a three-layer online data processing network based on MEC technology. The results could help big data processing of the IoT in the future [3]. The scholar Kumari A proposed that with the exponential growth of raw data generated by different sensors, actuators and mobile devices, data analysis is becoming a challenging problem, and analyzed and summarized how to deal with the

ISSN 2616-5775 Vol. 6, Issue 8: 82-88, DOI: 10.25236/AJCIS.2023.060810

security stream data generated by different devices. It includes the main threats and risks introduced in big data processing in the IoT environment [4]. However, these methods still have some problems, such as high algorithm complexity and poor model interpretability. Therefore, this paper aims to propose a simple, efficient and accurate artificial intelligence algorithm to process big data in the IoT system.

The research method of this paper is based on the artificial intelligence algorithm of data preprocessing, feature extraction and supervised learning, and combined with hardware devices to achieve real-time processing of big data of the IoT system. The significance of this study is to provide a new method to solve the big data processing problem in the IoT system. This method is not only efficient and accurate, but also can be applied to a variety of different IoT system scenarios. In addition, this study can also provide reference for optimizing the performance of IoT systems.

2. Related Technologies of the IoT and Big Data

2.1 IoT System

The IoT system is a network system that connects items through the Internet [5-6]. It connects various intelligent devices, sensors, and terminal devices to the Internet to achieve interconnection and information sharing between devices, thereby achieving functions such as interconnection between items, data collection and processing, intelligent decision-making and control [7-8]. The IoT system can be applied in various fields such as smart home, intelligent transportation, intelligent healthcare, intelligent manufacturing, etc. It can improve the collaboration and intelligence level between devices, improve resource utilization efficiency and production efficiency, and bring more convenience and benefits to people's production and life. The IoT structure is shown in Figure 1.

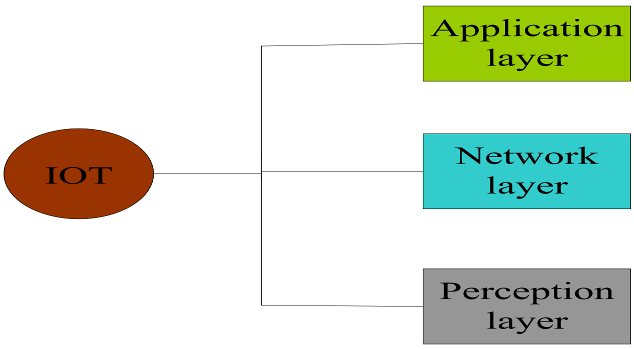

Figure 1: Structure of the IoT

2.2 Overview of Big Data Processing Technology

Big data processing technology refers to the process of processing, storing and analyzing data with a large number, high speed, diversity and complexity by adopting various technologies and methods [9-10]. Big data processing usually requires the use of distributed computing, cloud computing, machine learning and other technologies and algorithms to process data [11-12], and efficient storage and computing equipment to support the speed and scale of data processing. In big data processing, data usually needs to go through multiple stages such as cleaning, preprocessing, feature extraction, modeling and evaluation [13-14] to obtain effective results and insights. Big data processing has a wide range of applications, involving finance, medical care, energy, transportation, e-commerce, the IoT and other fields. Big data processing is of great significance in improving the competitiveness and innovation ability of enterprises, optimizing resource allocation and decision-making analysis, improving user experience and service quality.

2.3 Related Artificial Intelligence Algorithms and Technologies

Artificial intelligence algorithms and technologies are important tools and methods in big data processing, which can help process and analyze large amounts of data with complexity and diversity. The following are commonly used artificial intelligence algorithms and technologies:

(1) Machine learning: Machine learning is a data-based automatic learning algorithm [15-16] that achieves data processing and analysis through pattern recognition and classification of data.

ISSN 2616-5775 Vol. 6, Issue 8: 82-88, DOI: 10.25236/AJCIS.2023.060810

(2) Deep learning: Deep learning is an algorithm based on multi-layer neural networks [17-18], which can process a large amount of unstructured data and automatically extract features from the data.

(3) Data mining: Data mining is a technique that extracts useful information from data. It can help people discover patterns and patterns in data.

(4) Natural language processing: Natural language processing is a technology based on artificial intelligence and linguistics [19-20], which can help computers understand and process human language. Natural language processing can be applied to text analysis, emotion analysis, machine translation, speech recognition and so on.

(5) Reinforcement learning: Reinforcement learning is an algorithm based on trial and error learning, which tries and adjusts constantly to obtain the optimal behavior strategy. It can be applied to intelligent control, game playing, robot control, etc.

These artificial intelligence algorithms and technologies can be combined and applied to achieve more efficient and accurate data analysis and decision-making in big data processing. The algorithm is shown in Figure 2.

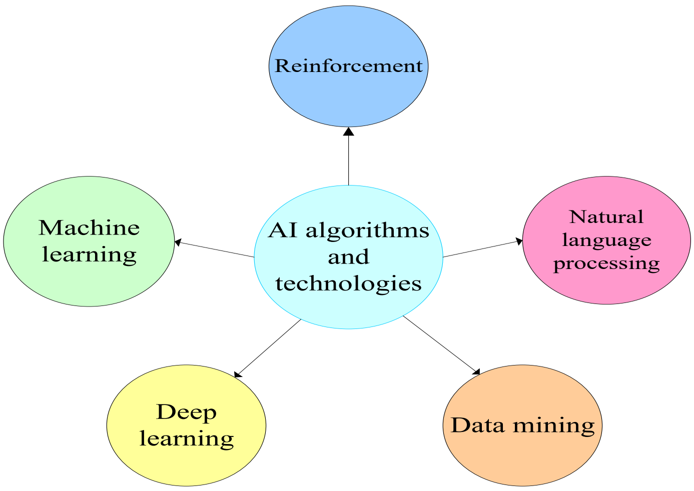

Figure 2: Algorithms and technologies

3. Big Data Mining Algorithm Based on Artificial Intelligence Technology

3.1 Model Construction

The process of building a data mining model usually includes several steps: firstly, before conducting data mining, it is necessary to clarify the mining objectives; secondly, it is necessary to preprocess the data; furthermore, feature extraction is performed to select suitable features for modeling and analysis. Feature extraction can use statistical methods, machine learning methods, etc. Then algorithms and models are selected, which require selecting appropriate data mining algorithms and models for modeling and analysis. Common data mining algorithms include decision trees, support vector machines, neural networks, and naive Bayes. Then there is the establishment and evaluation of the model. After selecting the appropriate algorithm and model, it is necessary to establish and evaluate the model. Establishing a model involves two processes, training the model and verifying the model. Evaluating the model includes indicators such as accuracy, recall, F1 value, etc. Finally, there is model optimization and application. After establishing and evaluating the model, it is necessary to optimize and apply the model. The optimization model can adopt parameter adjustment, feature selection and other methods, and the application model can be used for actual data analysis and decision support. Figure 3 shows the process of model construction.

ISSN 2616-5775 Vol. 6, Issue 8: 82-88, DOI: 10.25236/AJCIS.2023.060810

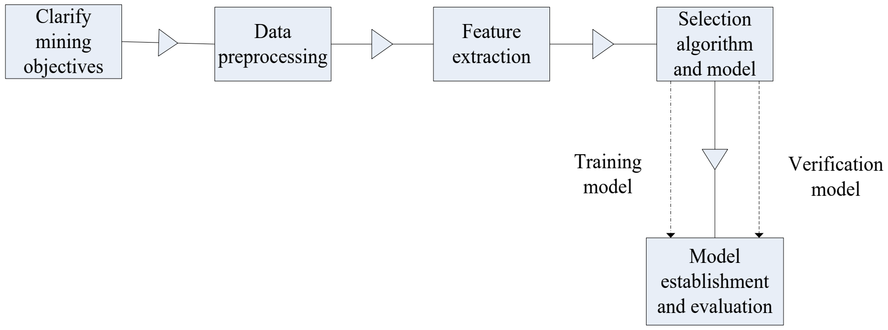

Figure 3: Process of model construction

3.2 Data Detection

In the feature detection of IoT data, the algorithms and models involved are relatively complex, and different algorithms and models have their own specific formulas and mathematical methods. Although the data model constructed earlier can determine the content and scope of data mining, it cannot guarantee the interrelationships between patterns, and therefore cannot obtain accurate data mining results. Therefore, in order to obtain more accurate data mining results, it is necessary to use feature extraction algorithms to analyze the characteristics of these data and extract the values of the data from their characteristics. Assuming that the big data mining set is *A* and the dataset dimension is *a*, the attribute set *V* of big data can be obtained according to the attributes of big data.

In this article, data detection can be achieved through time series analysis. The commonly used indicators in time series analysis include mean, variance, autocorrelation coefficient, partial autocorrelation coefficient, etc. Among them, autocorrelation coefficient and partial autocorrelation coefficient can be used to analyze the correlation and lag of time series data. Firstly, the data mining subspace is defined as Z, where all classes belong to A, and the attribute set of the data in subspace Z is encapsulated. From this, it can be calculated that the address of the data is (0, *Z*), which is in a non-normal distribution state. A random data is separated from the subspace, and its random probability is $I_a(0, Z)$. In this set, the central position of the data attributes in this subspace is 0, so the probability calculation of the subspace is shown in formula 1 (in formula 1 *a* is the distance).

$$A_s = \frac{1}{I_a(0, Z)} \tag{1}$$

Because most data in the IoT are non-normal distribution, the standard distance and density of physical quantity data must be used to describe the characteristics of discrete data, such as formula 2.

$$\lambda = \frac{I_a}{\sigma(0, Z)} \tag{2}$$

Based on discrete features, *K* can collect the distribution of dispersed data in the IoT. To obtain the above data values, the detection method of information entropy is used. Based on *x*, the measured data set *O* is analyzed and the data information in the measured data set *O* is obtained, as shown in formula 3.

$$N(\mathrm{o}) = -\sum_{o \in O} x(o) \ln x(o) \tag{3}$$

In the feature detection of IoT data, it is usually necessary to use technologies such as data mining and machine learning to achieve it. The commonly used implementation methods include the following steps:

(1) Data collection: collecting and obtaining data from different data sources, which can be done manually or automatically.

(2) Feature extraction: performing feature extraction on IoT data, including time series analysis, frequency domain analysis, wavelet analysis, etc., to extract key features of the data.

(3) Feature selection and dimension reduction: selecting and reducing the dimensions of extracted features to reduce the dimensions and complexity of data and improve the efficiency of data processing

and analysis.

(4) Selection of algorithm and models: selecting appropriate machine learning algorithms and models for modeling and analysis.

(5) Training model and prediction: training and predicting the model after selecting the appropriate algorithm and model. The training model includes data partition, model training and model verification. The prediction model can be used for the detection and analysis of actual data.

(6) Model evaluation and optimization: evaluating and optimizing the model after establishing and predicting the model. The evaluation model includes the accuracy, recall rate, F1 value and other indicators of the model. The optimization model can adopt parameter adjustment, feature selection and other methods.

Overall, IoT data feature detection requires the use of various data mining and machine learning technologies to process and analyze data, which requires comprehensive consideration and practical operation in multiple steps to achieve better detection results and application value. The specific situation is shown in Table 1.

Table 1: Feature detection

Feature	Value Range	Detection Method
Mean	Any real number	Calculate the mean of data
Standard Deviation	Real number greater than or equal to 0	Calculate the standard deviation of data
Skewness	Any real number	Calculate the skewness of data
Kurtosis	Any real number	Calculate the kurtosis of data

3.3 Data Processing

The standardization processing of feature data is to convert different feature data into data with the same size and distribution for comparison and analysis. The commonly used methods include Robust standardization, Log conversion, and Power conversion. Robust standardization subtracts the median from the data and divides it by Interquartile range, making the data robust and not affected by outlier. Log conversion is a logarithmic conversion of data, which can compress the value range of data and make the data more consistent with normal distribution. Power transformation can make the data more consistent with Normal distribution and enhance the characteristics of the data. This article calculates the average variance $D\delta$, as shown in formula 4.

$$\delta_{\mathrm{i}} = \frac{\delta - \overline{\delta}}{D\delta} \tag{4}$$

On this basis, the anti-interference ability of big data mining algorithm is enhanced by normalizing the feature data.

4. Experimental Design of Algorithms and Devices

4.1 Obtaining Datasets

In order to test the performance of the artificial intelligence algorithm for big data processing of the IoT system, the simulation test method was used to test the application effect of the algorithm. This article extracted 4620 two-dimensional data from the database, processed them, and placed them in a dataset. Afterwards, the data was classified into 18 different groups, with a fuzzy coefficient of 2.1 set for each group and 3 set as the debugging parameter for the transfer point.

4.2 Performance Indicators

The index used in this experiment design is mainly F1 performance index, which is an important index in the big data mining technology of the IoT. Its accuracy usually depends on its accuracy and recovery rate. Accuracy refers to the proportion of accuracy indicators in the data processing process. The recovery rate refers to the quantity of data with specific types of characteristics, as shown in Table 2.

ISSN 2616-5775 Vol. 6, Issue 8: 82-88, DOI: 10.25236/AJCIS.2023.060810

Table 2: Performance indicators

Number	Indicators
I	Accuracy
II	Recall
III	F1 value
IV	Training time

4.3 Experimental Results

The following is an experimental result based on a binary classification problem, which tests the validation rate, recall rate, F1 value, and training time of four performance indicators. The model established in this article is Model 1, while the other three models are relatively common in the market. The results are shown in Figure 4.

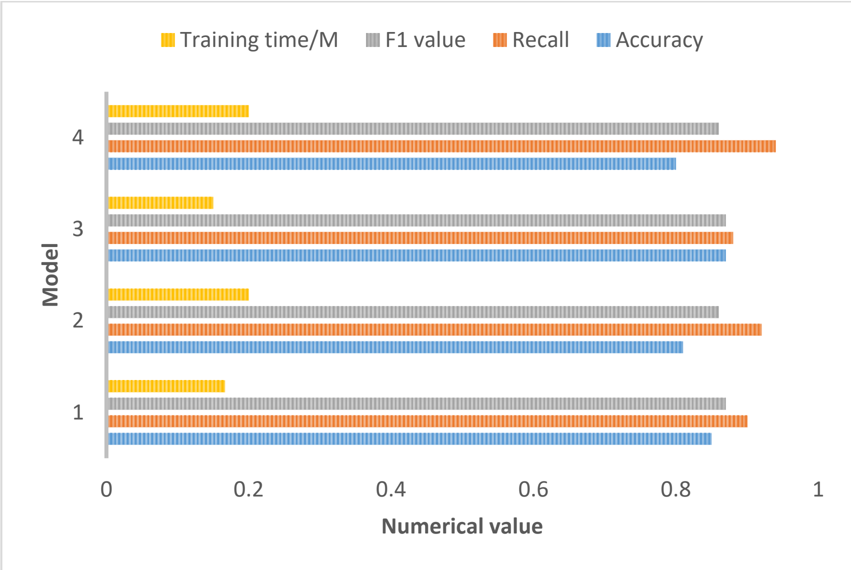

Figure 4: Test results

In the above experiment, four different models were used to test a binary classification problem, and the accuracy, recovery rate, average variance, and training time of each model were calculated. The experimental results showed that Model 1 studied in this article has the best performance with an F1 value of 0.87, while Model 2 and Model 4 have slightly poorer performance with an F1 value of 0.86. In terms of training time, Model 1 has the shortest training time of only 9 seconds, while Model 4 has the longest training time of 15 seconds. Overall, Model 1 performs the best in terms of performance and training time, and can be used as the optimal model for this binary classification problem.

5. Conclusions

After the processing of big data in the IoT system, this paper proposes an artificial intelligence algorithm and device of data mining technology to process big data in the IoT system. This algorithm and device can effectively improve the efficiency of data processing and analysis, as well as improve the accuracy and reliability of data processing.

Specifically, in the data preprocessing stage, this algorithm is used to process sensor data and device data in IoT systems to ensure data quality and reliability. In the feature extraction and selection stage, the algorithm and device adopt adaptive feature extraction and selection techniques to extract key features of the data and reduce the dimensionality and complexity of the data. The algorithm and device are tested and evaluated in the experiment. The results indicate that the algorithm and device can effectively improve the efficiency of data processing and analysis, as well as improve the accuracy and reliability of data processing. At the same time, the algorithm and device have good robustness and scalability, which can adapt to different data processing and analysis needs. The experiment verifies that the artificial intelligence algorithm and device based on data mining technology is an effective big data processing technology of the IoT system, with good performance and application prospects, and

ISSN 2616-5775 Vol. 6, Issue 8: 82-88, DOI: 10.25236/AJCIS.2023.060810

can be widely used in big data processing and analysis in the IoT system.

References

[1] Aryal A, Liao Y, Nattuthurai P, et al. The emerging big data analytics and IoT in supply chain management: a systematic review. Supply Chain Management: An International Journal, 2020, 25(2): 141-156

[2] Daissaoui A, Boulmakoul A, Karim L, et al. IoT and big data analytics for smart buildings: A survey. Procedia computer science, 2020, 170(2): 161-168

[3] Wan S, Lu J, Fan P, et al. Toward big data processing in IoT: Path planning and resource management of UAV base stations in mobile-edge computing system. IEEE Internet of Things Journal, 2019, 7(7): 5995-6009

[4] Kumari A, Tanwar S, Tyagi S, et al. Verification and validation techniques for streaming big data analytics in internet of things environment. IET Networks, 2019, 8(3): 155-163

[5] Nguyen D C, Ding M, Pathirana P N, et al. 6G Internet of Things: A comprehensive survey. IEEE Internet of Things Journal, 2021, 9(1): 359-383

[6] Dai H N, Zheng Z, Zhang Y. Blockchain for Internet of Things: A survey. IEEE Internet of Things Journal, 2019, 6(5): 8076-8094

[7] Loftus T J, Tighe P J, Filiberto A C, et al. Artificial intelligence and surgical decision-making. JAMA surgery, 2020, 155(2): 148-158

[8] Shrestha Y R, Ben-Menahem S M, Von Krogh G. Organizational decision-making structures in the age of artificial intelligence. California management review, 2019, 61(4): 66-83

[9] Wang J, Yang Y, Wang T, et al. Big data service architecture: a survey. Journal of Internet Technology, 2020, 21(2): 393-405

[10] Singh N. Big data technology: developments in current research and emerging landscape. Enterprise Information Systems, 2019, 13(6): 801-831

[11] Wang S, Cao J, Philip S Y. Deep learning for spatio-temporal data mining: A survey. IEEE transactions on knowledge and data engineering, 2020, 34(8): 3681-3700

[12] Sunhare P, Chowdhary R R, Chattopadhyay M K. Internet of things and data mining: An application oriented survey. Journal of King Saud University-Computer and Information Sciences, 2022, 34(6): 3569-3590

[13] Li D, Deng L, Lee M, et al. IoT data feature extraction and intrusion detection system for smart cities based on deep migration learning. International journal of information management, 2019, 49(2): 533-545

[14] Zebari R, Abdulazeez A, Zeebaree D, et al. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. Journal of Applied Science and Technology Trends, 2020, 1(2): 56-70

[15] Liu B, Ding M, Shaham S, et al. When machine learning meets privacy: A survey and outlook. ACM Computing Surveys (CSUR), 2021, 54(2): 1-36

[16] Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electronic Markets, 2021, 31(3): 685-695

[17] Chen J, Ran X. Deep learning with edge computing: A review. Proceedings of the IEEE, 2019, 107(8): 1655-1674.

[18] Bengio Y, Lecun Y, Hinton G. Deep learning for AI. Communications of the ACM, 2021, 64(7): 58-65

[19] Qiu X, Sun T, Xu Y, et al. Pre-trained models for natural language processing: A survey. Science China Technological Sciences, 2020, 63(10): 1872-1897

[20] Galassi A, Lippi M, Torroni P. Attention in natural language processing. IEEE transactions on neural networks and learning systems, 2020, 32(10): 4291-4308