Research on Real-time Mining of New Energy Vehicle Fault Diagnosis Data under Cloud Computing

WEI Jia-tong[1],XU Yan[2],KONG Chun-hua [3]

{ouyangmeixue100@163.com[1],xuyan464803398@163.com[2],13596157766@163.com[3]}

(1.Department of Automotive Engineering,Jilin Communications Polytechnic,Changchun 130012,China;

2.Department of Automotive Engineering,Jilin Communications Polytechnic,Changchun 130012,China;

3.Department of Automotive Engineering,Jilin Communications Polytechnic,Changchun 130012,China)

Abstract. In order to effectively improve the semantic retrieval ability and information analysis ability of network database, and improve the accuracy of data mining of new energy vehicle fault diagnosis, a real-time data mining method of new energy vehicle fault diagnosis under cloud computing is proposed. The modern signal mining technology is used to analyze the fault diagnosis data of new energy vehicles, and a data signal analysis model is established. On this basis, the fault diagnosis data of each section of new energy vehicles are segmented matched and filtered, and the characteristic input of the fault diagnosis data of new energy vehicles is obtained, combining the optimal classification surface of the fault diagnosis data characteristics of new energy vehicles and the fault of new energy vehicles The feature vector track of the fault diagnosis data can realize the extraction of the feature value of the fault diagnosis data of new energy vehicles, and complete the real-time mining of the fault diagnosis data of new energy vehicles under the cloud computing. Experimental results show that the proposed method has high accuracy and real-time performance.

Keywords: cloud computing; new energy vehicles; fault diagnosis; data mining;

1 Introduction

The development of information technology has led to the explosive growth of information production scale and transmission speed, which has brought human social life and scientific research into the era of cloud computing. Research shows that the amount of data generated by fault diagnosis of new energy vehicles in recent years is more than the total data created by the whole traffic history, and this trend is still accelerating. In the era of cloud computing, thanks to the continuous progress of information technology, every process of human life can be recorded in the form of data and displayed in different ways. Therefore, these massive data have the characteristics of heterogeneous types, low value density, huge data model and fast propagation speed [1]. Traditional data analysis and processing technologies have been unable to meet the requirements of cloud computing analysis. How to design data mining and analysis technologies that meet the characteristics of cloud computing,

EAI MOBIMEDIA 2020, August 27-28, Harbin, People's Republic of China

DOI 10.4108/eai.27-8-2020.2294585

explore the hidden laws and knowledge in cloud computing, and enhance the timeliness of decision-making services are the key contents of current research in the field of data mining and analysis.

At present, one of the main contradictions of data analysis and data processing is that the increase of data scale and transmission speed cannot be synchronized with the increase of computing power, that is to say, the requirements of analyzing and processing large-scale dynamic fault diagnosis data of new energy vehicles exceed the current computing power, which seriously affects the effectiveness of decision-making services. The traditional methods and models of fault diagnosis data analysis of static new energy vehicles have been unable to meet the processing requirements of large-scale dynamic new energy vehicle fault diagnosis data [2]. In the application of sensor network monitoring, sensors in network nodes transmit large-scale fault diagnosis data of new energy vehicles to the central processor in real time. The monitoring system needs to make timely and accurate statistical analysis on the feedback data of sensors, and make corresponding judgments and decisions. In social network analysis, social network media such as Tweeter, Facebook and so on are available at all times Massive social information release and reception. Among them, public opinion monitoring is an important research content. The real-time analysis of public opinion of social hot spots, key events and issues can accurately grasp the context of the event development, predict the direction of the event development, and timely make reasonable disposal. In the management and analysis of fault diagnosis data of new energy vehicles, the real-time analysis of massive fault diagnosis data of new energy vehicles can quickly and accurately mine out various modes of fault diagnosis of new energy vehicles, and timely discover new energy vehicle faults [3]. In the new energy vehicle market, various activities produce large-scale new energy vehicle fault diagnosis data in real time, and these data have the characteristics of wide coverage, rich content, strong real-time, and most of them are unstructured text data, such as stock index data recording the fluctuation of new energy vehicle stock price, and related financial news information. The new energy vehicle market is driven by information, and the change of fault diagnosis information of new energy vehicles will be reflected in the fluctuation of new energy vehicle market in varying degrees. Therefore, the relevant analysis of large-scale new energy vehicle fault diagnosis data can effectively discover the operation law of new energy vehicle market, and the relevant new energy vehicle fault diagnosis information and new energy The connection between the automobile market.

The fault diagnosis data of new energy vehicles generated in the above application fields have the characteristics of massive, high-speed and dynamic, and the timeliness of the data is strong. It needs the relevant analysis system to make a quick response, accurately mine the hidden laws and knowledge, and make a decision in time [4]. Therefore, according to the characteristics of these application areas, researchers put forward a new data model, the new energy vehicle fault diagnosis data flow model, which can meet the requirements of real-time and efficient data mining tasks.

The fault diagnosis data stream of new energy vehicles is composed of a series of data which arrive in sequence according to the time axis. It can also be regarded as a digital signal string which is encoded and processed in the process of information transmission. It has the dual attributes of time and space. Its spatial attribute is the same as the static new energy vehicle fault diagnosis data set, which can be considered as the physical meaning expressed by the attribute value of the data generated by the system. Its time attribute is mainly reflected in the arrival order of data. Each stream of data has a matching time index, that is, the order of reaching the analysis system, which is the main difference with the static new energy vehicle fault diagnosis data set. Excellent new energy vehicle fault diagnosis data analysis algorithm should be insensitive to the order of the input new energy vehicle fault diagnosis data when analyzing the static new energy vehicle fault diagnosis data set. When processing the new energy vehicle fault diagnosis data flow pattern, the relevant algorithm must consider the time sequence relationship between the data and mine the new energy vehicle fault diagnosis The evolution law of broken data [5]. The data flow model of new energy vehicle fault diagnosis is similar to the time series model, which has time and space attributes, but the time series data is a kind of static data, while the data flow model of new energy vehicle fault diagnosis represents the dynamic data, which has the following characteristics: the arrival sequence of new energy vehicle fault diagnosis data is independent and uncontrolled; the fault diagnosis data of new energy vehicle When it arrives in real time, the analysis system must respond quickly; the scale of fault diagnosis data of new energy vehicles tends to be infinite, and it can not be saved in memory, in principle, it can only read the data once.

The complex pattern mining on the data stream of new energy vehicle fault diagnosis has practical and theoretical significance. First of all, due to the rapid development of Internet and communication technology, a large number of new energy vehicle fault diagnosis data have been automatically generated in the aspect of new energy vehicle fault diagnosis, and data analysis technology has entered the era of cloud computing. Cloud computing has the characteristics of huge data scale, various data types, low data value density and fast data processing speed. Therefore, data flow mining, which can quickly respond to large-scale fault diagnosis data of new energy vehicles, is an important research content of cloud computing analysis. Secondly, the research fields such as intrusion mining, anomaly analysis, trend monitoring, exploratory analysis are very sensitive to the response speed of new energy vehicle fault diagnosis data, which needs real-time complex analysis of the data. Finally, due to the characteristics of dynamic new energy vehicle fault diagnosis data, it is impossible to obtain the overall distribution information of data in the process of data mining. Dynamic data mining is a process from local to global, from details to the whole. Therefore, through the analysis of data flow, we can explore the internal evolution mechanism of the system, as well as the evolution rules of various knowledge or patterns hidden in the data flow law [6].

Therefore, this paper analyzes the data flow of new energy vehicle fault diagnosis, and proposes a real-time mining method of new energy vehicle fault diagnosis data under cloud computing. Combined with modern data processing technology to extract the characteristic value of fault diagnosis data of new energy vehicles, the analytical model of fault diagnosis data signal of new energy vehicles is decomposed through segmented pre whitening processing, and the fault diagnosis data of new energy vehicles is matched and mined, so as to extract the characteristic value of fault diagnosis data of new energy vehicles and complete the fault diagnosis of new energy vehicles under cloud computing Data mining in real time. The experimental results show that the proposed method improves the mining accuracy of fault diagnosis data of new energy vehicles, and achieves the purpose of improving the real-time data mining.

2 Real time mining of fault diagnosis data of new energy vehicles under cloud computing

2.1 Data flow relevance

This paper analyzes the data flow density of new energy vehicle fault diagnosis. After the data flow density of new energy vehicle fault diagnosis is determined, the region is divided according to the data flow density of new energy vehicle fault diagnosis. The correlation between the regional network and the data flow of new energy vehicle fault diagnosis is calculated by using the incremental subspace data mining theory, and the node and time relationship of the data flow are analyzed The number of contacts can accurately determine the nodes of fault diagnosis data flow of new energy vehicles, and complete the real-time mining of fault diagnosis data of new energy vehicles under cloud computing. The specific process is as follows:

Assuming that the number of network types in the cloud computing environment is the number of regions, an undirected ergodic graph $G=(V,E)$ is used to describe the network regions in the whole cloud computing environment, among them, V represents the fault diagnosis node set of new energy vehicles, and E represents the fault diagnosis link set of new energy vehicles, In the cloud computing environment, the network consists of m

regions, and $V=(v_1,v_2,\cdots,v_n)$ is the number of nodes in the new energy vehicle fault diagnosis network region. Then, the data flow density of the regional network new energy vehicle fault diagnosis can be calculated by formula (1):

$$\varphi(k)=\left(\sum_{i=1}^{n}\frac{\left[\frac{M(k)}{N(k)}-\frac{M(p)}{N(p)}\right]}{n}\right)\times\left(\frac{M(k)}{nN(k)}\right)^2 \quad (1)$$

Where, $\varphi(k)$ represents the data flow density of fault diagnosis of new energy vehicles in the k th region, $M(k)$ represents the number of fault diagnosis data flow of new energy vehicles in the k th region, and $N(k)$ represents the number of nodes in the k th region, n represents the number of network loops in the cloud computing environment, i represents the fault diagnosis data flow of new energy vehicles, $M(p)$ represents the number of fault diagnosis data flow of new energy vehicles in the p th region, and $N(p)$ represents the number of nodes in the p th region.

Assuming that $\sigma(i)$ represents the size of the new energy vehicle fault diagnosis data stream i that is known to be mined, and $D(i)$ represents the feature set of the new energy vehicle fault diagnosis data stream i, the correlation calculation formula of the new energy vehicle fault diagnosis data stream i and area k can be obtained as follows:

$$L(k,i)=\exp\left[-\frac{\sqrt{\left(\sum_{j=1}^{m}(\varphi(u)-\varphi(k))\right)}}{1+\sigma(i)}\times\alpha\cdot D(i)\right] \quad (2)$$

In the formula, α represents the correlation degree factor, which is related to the size of the average new energy vehicle fault diagnosis data stream of the regional network, $\varphi(u)$ represents the new energy vehicle fault diagnosis data stream density in the u region, j

represents a new energy vehicle fault diagnosis data stream, and m represents the number of new energy vehicle fault diagnosis data streams in the k region.

After using the above process to obtain the region of the data flow of fault diagnosis of energy vehicles, the data mining method based on multi incremental space is used to mine the loop of the data flow of fault diagnosis of new energy vehicles, and the correlation coefficient between the node and time of the data flow of fault diagnosis of energy vehicles is obtained, so as to accurately determine the node of the data flow of fault diagnosis of new energy vehicles Real time mining of fault diagnosis data of new energy vehicles under cloud computing [7].

2.2 Extract eigenvalues of fault diagnosis data of new energy vehicles

In the process of real-time mining of fault diagnosis data of new energy vehicles, firstly, a new energy vehicle fault diagnosis data signal model is established, and the modern signal mining technology is used to analyze the discrete data of new energy vehicle fault diagnosis data signal. Then, a new energy vehicle fault diagnosis data signal analysis model is established under cloud computing, and the new energy vehicle fault diagnosis data in the network under cloud computing are analyzed The high frequency signal simulation is carried out, and the fault diagnosis data of each new energy vehicle is mined by segmented matching filter. The specific process is as follows:

Firstly, a new energy vehicle fault diagnosis data signal model is constructed, which integrates modern signal mining technology to make the new energy vehicle fault diagnosis data signal discrete data analysis, and establishes the new energy vehicle fault diagnosis data signal analysis model in the cloud computing environment:

$$\begin{cases} H_0 : x_{k+1}(t) = r_{k+1}(t) \\ H_1 : x_{k+1}(t) = s(t-\tau) + r_{k+1}(t) \end{cases} 0 \le t \le T_B \quad (3)$$

Where, $x_{k+1}(t)$ represents the fault diagnosis data signal of new energy vehicles, $x(t)$ represents the real part of the analytical model of the fault diagnosis data signal of new energy vehicles, $x_{k+1}(t)$ represents the LFM signal when the pulse width of the fault diagnosis data of new energy vehicles is T_p, the maximum pulse width is T_B, $s(\cdot)$ represents the high-order cumulant, t represents the sampling time interval of the fault diagnosis data signal of new energy vehicles, and τ represents the time change parameter.

Suppose that the fault diagnosis data of network new energy vehicles in the cloud computing environment is time-varying model AR , p represents the corresponding order of the model, $a_i(t)$ is the model parameter, and formula (4) is used to represent the interference noise $n(t)$ in the process of fault diagnosis data mining of network new energy vehicles

$$n(t) = -\sum_{i=1}^{p} a_i(t)\cdot n(t) + \sigma(t)\cdot w(t) \quad (4)$$

In the formula, $\sigma(t)$ represents the Gaussian asymmetric function. From the above formula, $w(t)$ is the Gaussian white noise with mean value of 0 and variance of 1. In the cloud computing environment, the network reverberation data is obtained by formula (5):

$$r(t) = \alpha n(t) \quad (5)$$

By pre whitening the fault diagnosis data of network new energy vehicles, the analytical model of fault diagnosis data signal of new energy vehicles is decomposed into multiple narrow-band signals, and the matching mining of fault diagnosis data of new energy vehicles is carried out [8], and the interference frequency characteristic formula of fault diagnosis data of new energy vehicles is given by formula (6):

$$f(t) = \frac{1}{2\pi} \times \frac{d\theta(t)}{dt} \cdot \beta \quad (6)$$

In the formula, $\theta(t)$ represents the phase information of fault diagnosis data signal of new energy vehicles, dt represents the interference characteristic amplitude of fault diagnosis data of new energy vehicles, and β represents the segment pre whitening matching parameter. In the process of estimating the spectrum of k segment of fault diagnosis data of new energy vehicles, the reverberation spectrum in k segment is estimated by AR model, and the current fault diagnosis data segment of new energy vehicles is smoothed That is:

$$r_k(t) = -\sum_{i=1}^{p_k} a_{k,i} r(t) + w_k(t) \quad (7)$$

Where, $w_k(t)$ represents the weight of new energy vehicle fault diagnosis data under cloud computing, which represents a Gaussian white noise with variance of $\sigma_{w,k}^2$, $a_{k,i}$ represents the lattice recursive model parameters of the k -th segment of new energy vehicle fault diagnosis data, and p_k represents the corresponding order of Lattice recursive model.

Based on the effective recognition results of new energy vehicle fault diagnosis data, the least square method is used to mine the features of new energy vehicle fault diagnosis data, and the feature input of new energy vehicle fault diagnosis data is obtained. Combined with the optimal classification surface and feature vector track of new energy vehicle fault diagnosis data, the feature value extraction of new energy vehicle fault diagnosis data is realized Based on this, real-time mining of fault diagnosis data of new energy vehicles under cloud computing is completed [9]. The specific process is as follows:

Based on the effective identification results of the fault diagnosis data of new energy vehicles given above, the least square method is used to mine the fault diagnosis data features of new energy vehicles, and the input amount of the fault diagnosis data of new energy vehicles is obtained as follows:

$$E = \sum_{p=1}^{n} E_p = \frac{1}{l} \sum_{p=1}^{n} \sum_{k=1}^{l} \left[r_p(k) - y_p(k) \right]^2 \quad (8)$$

In the formula, l represents the magnitude of fault diagnosis data of new energy vehicles, E_p represents the termination frequency of fault diagnosis data of new energy vehicles, and $r_p(k)$ represents the initial data. According to equation (8), the feature value of fault diagnosis data of new energy vehicles is extracted and the phase space Fourier transform is carried out in the cloud computing environment according to the input value of fault diagnosis data of new energy vehicles. The optimal classification surface of fault diagnosis data of new energy vehicles needs to be calculated. The optimal solution of the problem can be represented by χ^*, and the fault diagnosis of new energy vehicles can be obtained by equation (9) The optimal classification plane of fault data features:

$$H_s(\eta,\psi)=\sum_{m=-\infty}^{\infty} J_m(t)*[f_m+mf_s+\eta]\exp\left[j(m\varphi+\theta+\psi)\right] \quad (9)$$

In the formula, $J_m(t)$ represents the fault diagnosis data of new energy vehicles at time t, f_m represents the decreasing trend function of fault diagnosis data of new energy vehicles, η represents the initial classification time of fault diagnosis data of new energy vehicles, f_s represents the carrier frequency of fault diagnosis data of new energy vehicles, φ represents the inclination angle of fault diagnosis data input of new energy vehicles, θ represents the line of sight angle of data, ψ represents the data output The initial angle of input time, j represents the interval of fault diagnosis data of new energy vehicles. If the characteristic component z of fault diagnosis data of new energy vehicles is set in the radius T region, the optimal classification surface of fault diagnosis data of new energy vehicles under cloud computing should meet the following constraints:

$$j\le \min\left[\left(\frac{T^2}{\Delta^2}\right),n\right]+1 \quad (10)$$

In the formula, Δ represents the change component of the fault diagnosis feature of new energy vehicles, normalizes the above results, transforms the quadratic programming problem in the support vector machine algorithm into a linear equation for solution from another perspective, the time series of the fault diagnosis data of new energy vehicles is $\{x(t_0\_i\Delta t)\}, i=0,1,\cdots N-1$, and the vector track after the quantification of the fault diagnosis data of new energy vehicles is:

$$X=x_n\left[x(t_0),x(t_0+\Delta t),\cdots,x(t_0+(K-1)\Delta t)\right] \quad (11)$$

In the formula, Δt represents the sampling time interval of fault diagnosis data of new energy vehicles, x_n represents the feature directivity of fault diagnosis data of new energy vehicles, $x(t_0)$ represents the vector track after quantization of fault diagnosis data of new energy vehicles at time t_0, $x(t_0+\Delta t)$ represents the vector track after quantization of fault diagnosis data of new energy vehicles at time $t_0+\Delta t$, and K represents the coefficient of non-linear normalized basic parameter.

Through the above processing process, combined with the optimal classification surface of new energy vehicle fault diagnosis data features and the vector track after the new energy vehicle fault diagnosis data quantification under cloud computing, the new energy vehicle fault diagnosis data feature value extraction is realized, which is shown in formula (12)

$$z(r)=X\cdot E*H_s(\eta,\psi) \quad (12)$$

In the formula, X represents the vector trajectory correlation parameter, and regards the above results as the basic fault diagnosis data of new energy vehicles, and combines the deep learning theory to complete the real-time mining of fault diagnosis data of new energy vehicles.

2.3 Real time mining algorithm design under cloud computing

According to the Apriori feature, the selection set is pruned to reduce the support count of the extra selection set. In the phase of parallel operation, according to the load balancing characteristics of Hadoop distributed architecture, the unreasonable data division of fault diagnosis of new energy vehicles can be avoided to the greatest extent.

For the new energy vehicle fault diagnosis data file D, assume its scale is G, transaction number is N, the new energy vehicle fault diagnosis data block size of the new energy vehicle fault diagnosis database is M, Hadoop divides the new energy vehicle fault diagnosis data file D into G / M block new energy vehicle fault diagnosis data. Here, Nmax is used to describe the maximum number of parallel map tasks in Hadoop distributed file system. When G / M > nmax, the 1-term set with support lower than Nminsp is removed, and the fault diagnosis data of new energy vehicles are described in vertical form. The vertical partition method is used for output, and the transaction identifiers of each range are stored in each file, and then the frequent 2-item set is calculated.

Based on the above analysis, a real-time data mining algorithm for fault diagnosis of new energy vehicles based on Hadoop distributed architecture is proposed

Input: new energy vehicle fault diagnosis data file D, minimum support threshold minsup;

Output: new energy vehicle fault diagnosis data set.

Step 1: Organize the fault diagnosis data set G of new energy vehicles into G / M small data blocks. All small data block nodes need to be calculated. The calculation method is based on the fault characteristics of new energy vehicles. The output results of all map tasks are temporary files containing the fault diagnosis data of new energy vehicles;

Step 2: Combine the local output results in the map phase to reduce the data exchange volume of new energy vehicle fault diagnosis between nodes. The format of each line is ＜item，Tid-Set ＞，TidSet is the transaction set containing the project;

Step 3: Merge the temporary files of all calculation nodes into vertical 1-item sets through the reduce function, and calculate the item support degree. If the support degree exceeds the product of N and minsup, it is considered as frequent 1-item set. Split the transaction sets of corresponding items and send them to different types of files. In order to ensure the consistent size of all vertical block data and HDFS data blocks, this section divides the transaction set of all 1-item sets into $\frac{G}{M}$ data blocks, in which the first $\frac{G}{M}$ files cover the transaction volume of $\frac{N}{G/M}$, and the last file covers the transaction volume of $N-N\frac{G/M-1}{G/M}$, so the transaction i is in the j file, where:

$$j=\begin{cases}\dfrac{i}{\left|\dfrac{N}{G/M}\right|}, i \le N\dfrac{G/M-1}{G/M} \\ \dfrac{G}{M}, i > N\dfrac{G/M-1}{G/M}\end{cases} \tag{13}$$

Step4：$m \leftarrow 2$；

Step 5: Divide the transaction set of the project into several data blocks and transfer them to the map process, so that all processes can calculate the support of candidate m-item set in parallel. Through the reduce process, the output results of all map processes are collected together to obtain the global m-item set. According to the idea of block, the global frequent m-item set is sent to each file;

Step6：$m \leftarrow m+1$；

Step 7: Repeat step 5 to step 7 until the fault diagnosis data output of new energy vehicles without more frequent item sets, and the output result is the fault diagnosis data of new energy vehicles.

In the process of parallel mining, the new energy vehicle fault diagnosis data real-time mining algorithm based on Hadoop distributed architecture uses mobile program rather than parallel mining program. At the same time, through step (2), the mining results of all nodes are gathered together, so as to reduce the traffic and communication cost. The proposed method divides the fault diagnosis data set of new energy vehicles into non overlapping data blocks, so that the intersection operation can be executed in parallel, greatly improving the real-time performance of fault diagnosis data mining of new energy vehicles [10].

2.4 Real time mining of fault diagnosis data of new energy vehicles

This section conducts correlation analysis on the fault diagnosis data of new energy vehicles, and mines the fault diagnosis data of new energy vehicles. The flow chart is shown in Fig. 1.

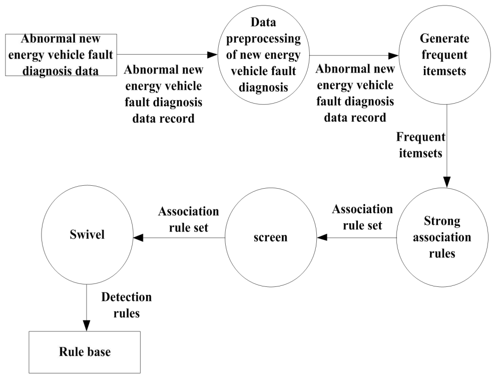

Fig. 1 flow chart of real-time mining of fault diagnosis data of new energy vehicles

The real-time mining process of fault diagnosis data of new energy vehicles is as follows:

1) Preprocessing of fault diagnosis data of new energy vehicles. Before the correlation analysis, the fault diagnosis data of new energy vehicles are processed, and the fault diagnosis data of new energy vehicles are imported into the database for correlation analysis;

2) Correlation analysis. Association analysis is an effective data mining technology for fault diagnosis of new energy vehicles, which is mainly realized through the following two processes:

① Form frequent item set. In this section, the frequent itemsets of new energy vehicle fault diagnosis data are mined by Apriori method, which is an effective method to mine frequent itemsets of Boolean association rules and is widely used.

② Form strong association rules. Strong association rules are association rules that meet the minimum confidence. After obtaining frequent item sets, strong association rules are formed through the following process:

1) For each frequent term set q with cluster label r, form subset $v=(q-r)$ of q.

2) If $\frac{S_N(q)}{S_N(r)} \geq C_{\min}$, strong association rule $q-r \rightarrow r$ will be formed.

Here $S_N(q)$ and $S_N(r)$ are used to describe the support numbers of q and r in turn; $C_{\min}$ is used to describe the minimum confidence level.

Each association rule described a certain attribute of the fault diagnosis data of new energy vehicles. All rules can be regarded as references and applied to the real-time mining of fault diagnosis data of new energy vehicles. In the process of mining, support and confidence are calculated according to the rules, and the similarity of association rules is compared, which can realize the real-time mining of fault diagnosis data of new energy vehicles.

Compare rule $Z_1(A \rightarrow B.S_1C_1)$ with rule $Z_2(A \rightarrow B.S_2C_2)$ by the following formula:

$$w(Z_1, Z_2) = |S_1 - S_2| + |C_1 - C_2| \quad (14)$$

The larger $w(Z_1, Z_2)$ value is, the smaller the similarity between strong association rule Z_1 and Z_2 is. A threshold δ should be set in advance. If $w(Z_1, Z_2) \geq \delta$, the corresponding data is the fault diagnosis data of new energy vehicles.

3 Comparative analysis of experiments

In order to verify the performance of the proposed real-time mining method for fault diagnosis data of new energy vehicles under cloud computing, experiments are carried out. Matlab simulation software is used to design the data mining algorithm of new energy vehicle fault diagnosis. The experimental test data set is from a large-scale cloud storage database. The number of samples is 2500, and the sampling period is T = 0.08s. In the process of data real-time mining, the intensity of interference between network data codes is SNR = 0-24db, and the base frequency of scalar time series is 200Hz. According to the above experimental environment and parameter settings, the simulation analysis of real-time mining method of feature data is carried out, and the time-domain waveform of original new energy vehicle fault diagnosis data flow is obtained as shown in Fig. 2.

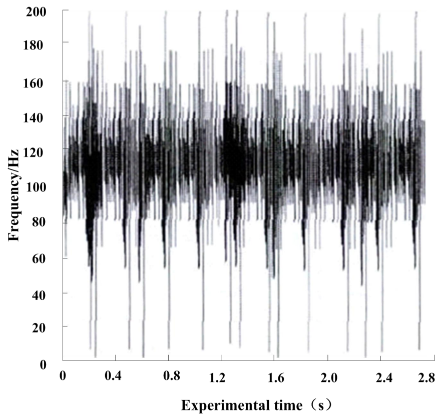

Fig. 2 Time domain waveform of information flow of original new energy vehicle fault diagnosis data

Based on the discrete data feature extraction results of the new energy vehicle fault diagnosis data signal in Fig. 2, the real-time mining performance of different types of new energy vehicle fault diagnosis data sets is measured. New energy vehicle fault diagnosis data set includes three types: data set A (including 602142 pieces of data), data set B (including 495203 pieces of data), data set C (including 584762 pieces of data). Under the condition of false alarm rate $P_f = 0.05$, in the range of -50 ~ 20dB, 500 Monte Carlo tests are used to conduct real-time mining simulation of new energy vehicle fault diagnosis data, and the network under cloud computing is obtained The comparison results of the accurate mining probability (%) of the fault diagnosis data of China new energy vehicle are shown in Fig. 3.

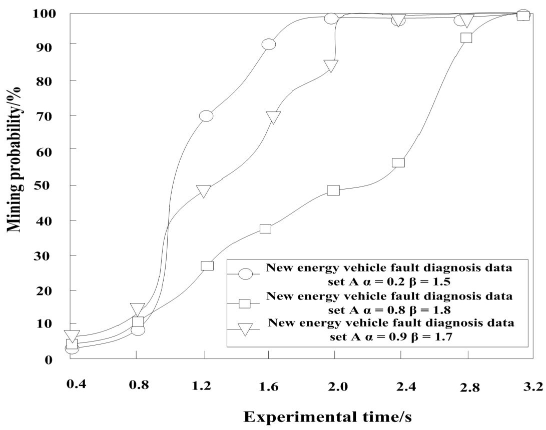

Fig. 3 Comparison of mining probability of different data sets

According to the analysis of Figure 3, the mining probability of fault diagnosis data mining method of new energy vehicles based on segmented pre whitening matching mining under cloud computing is high under different types of data sets, which shows that the method in this paper can reasonably configure parameters α. β, effectively improves the probability of mining, improves the accuracy of the new energy vehicle fault diagnosis data identification significantly, shows the better performance of the new energy vehicle fault diagnosis data mining, improves the semantic retrieval ability and information analysis ability of the network database.

Table 1 shows the real-time mining method of fault diagnosis data of new energy vehicles based on segment pre whitening matching mining under cloud computing, and the mining accuracy results under 500 Monte Carlo experiments on three test data sets.

Table 1 Comparison of mining accuracy on different data sets (%)

Number of experiments	New energy vehicle fault diagnosis data set A	New energy vehicle fault diagnosis data set B	New energy vehicle fault diagnosis data set C
100	92.15	97.27	95.21
200	93.12	96.89	94.89
300	91.02	97.65	93.52
400	90.28	95.41	95.21
500	93.58	96.87	94.21

From the experimental results in Table 1, the accuracy of the data mining method based on segmented pre whitening matching mining in all test sets is higher than 90%. In order to further verify the effectiveness of this method, the mining accuracy of the traditional method and this method is compared and analyzed, and the comparison results are shown in Figure 4.

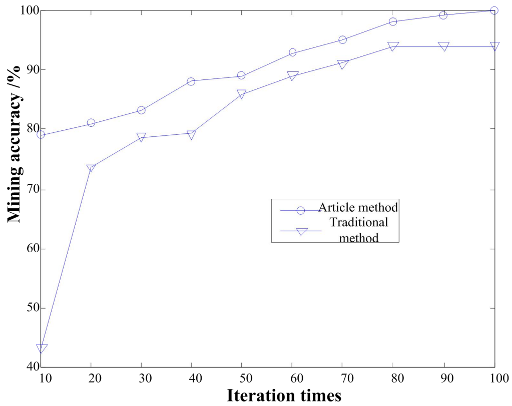

Fig. 4 Mining accuracy of different methods

According to figure 4, the mining accuracy of the method in this paper is mainly due to the traditional method. In the process of segmented matching filter mining for each section of fault diagnosis data of new energy vehicles, the method realizes the fault diagnosis data of new energy vehicles by combining the optimal classification surface of fault diagnosis data characteristics of new energy vehicles and the feature vector track of fault diagnosis data of new energy vehicles The extraction of eigenvalues improves the performance and calculation efficiency of the subsequent new energy vehicle fault diagnosis data classification, and reduces the packet loss rate of non new energy vehicle fault diagnosis data.

4 Concluding remarks

In view of the shortcomings of current data mining methods in dealing with fault diagnosis data of new energy vehicles, a real-time data mining method based on segmented pre whitening matching mining for fault diagnosis data of new energy vehicles in cloud computing environment is proposed, and the real-time performance and accuracy of the proposed method in dealing with fault diagnosis data of new energy vehicles are proved by simulation. In the process of researching the real-time mining method of new energy vehicle fault diagnosis data under cloud computing, the problem of mining time is not considered. In the next research, mining time is taken as the experimental index to improve the effectiveness of real-time mining of new energy vehicle fault diagnosis data.

5 Fund projects

Research and Innovation Fund of Science and Technology Development Center of Ministry of Education—"Beichuang Assistant" Fund Project ：《Research on Application of New Energy Vehicle Training Platform Intelligent Network Technology》（number: 2018A05024）

Reference

[1] Yang, P.L.: Mining method for target feature data in color image database. Journal of Shenyang

University of Technology. Vol. 40, pp. 60-64 (2018)
[2] Liu, B.H. Fu, Z.G. Wang, Y.Z.: Big data mining technology based on parallel algorithm and its application in power plant boiler performance optimization. Journal Of Chinese Society Of Power Engineering. Vol. 38, pp. 431-439 (2018)
[3] Liang, Z.H. Wu, J.H. Xie, Z.L.: Variable frequency room air conditioner operation pattern recognition and data mining. Journal of Mechanical Engineering. Vol. 55, pp.194-202 (2019)
[4] Weng, P.C. Zhang, Y.H. Ma, H.: Research and improvement of outlier data mining technology in Web network. Modern Electronics Technique. Vol. 40, pp. 29-31 (2017)
[5] Zhang, P. Ding, L.Y. Jiang, N.: Data mining algorithm of the automation system of the distribution network based on the support-confidence-lift framework and its application. Electrical Measurement & Instrumentation. Vol. 56, pp. 62-68 (2019)
[6] Yan, L. Qi, B.: Research on big data mining technology of mobile learning system based on Android platform. Modern Electronics Technique. Vol. 40, pp. 142-144 (2017)
[7] Xie, G. Pan, Y.X.: Weak correlation mining technique for laser network failure data. Laser Journal. Vol. 39, pp. 134-138 (2018)
[8] Fei, X.J. Li, H. Tian, G.Z.: Data clustering algorithm based on feature weighting theory. Journal of Shenyang University of Technology. Vol. 40, pp. 77-81 (2018)
[9] Qi, Z.C. Wang, J. Guo, J.C.: Medication regularity of Xin ' an Wang ' s internal medicine in treatment of spleen-stomach diseases: a data mining study based on unblocking yang theory. Journal of Beijing University of Traditional Chinese Medicine. Vol. 42, pp. 691-696 (2019)
[10] Li Z.W. He, L.F. Zhang, X.F.: Research on Application of Correlation Parameter Data of Main Transformer Short Circuit Impedance. Transformer. Vol. 6, pp. 31-36 (2019)