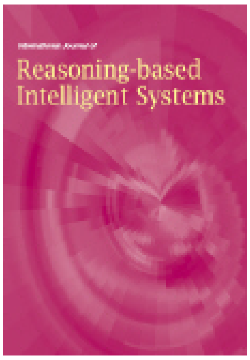


International Journal of Reasoning-based Intelligent Systems

ISSN online: 1755-0564 - ISSN print: 1755-0556
https://www.inderscience.com/ijris

An optimisation of mobile terminal data mining method based on internet of things

Yi Wang

DOI: 10.1504/IJRIS.2023.10053466

Article History:

Received:	12 May 2022
Accepted:	22 November 2022
Published online:	19 March 2024

An optimisation of mobile terminal data mining method based on internet of things

Yi Wang

Modern Education Technology Centre,
Shijiazhuang University of Applied Technology,
Shijiazhuang 050081, China
Email: lindawangyi@163.com

Abstract: In this paper, the optimisation of mobile terminal data mining method based on internet of things (IoT) is studied. Firstly, a framework for mobile terminal data mining optimisation is constructed, and mobile terminal data is collected by the mobile agent wireless sensor data acquisition technology. Then, the collected data are clustered by the chaotic search particle swarm K-means algorithm, and the clustered data are transmitted to the abnormal access detection module of mobile terminal users. The access detection module finally completes the mining of abnormal access behaviours of mobile terminal users by detecting the abnormal characteristics of user access behaviours, determining the abnormal type and checking the abnormal evolution. The experimental results show that the energy consumption of this method does not exceed 4J in a noisy environment, and this method is low in the data mining energy consumption and high in the accuracy.

Keywords: internet of things; IoT; mobile terminal; data mining; data acquisition; data clustering; abnormality detection.

Reference to this paper should be made as follows: Wang, Y. (2024) 'An optimisation of mobile terminal data mining method based on internet of things', *Int. J. Reasoning-based Intelligent Systems*, Vol. 16, No. 1, pp.58–65.

Biographical notes: Yi Wang received her Master's degree in Computer Application Technology from Shijiazhuang Tiedao University in 2008. She is currently an Associate Professor in Modern Education Technology Centre of Shijiazhuang University of Applied Technology. Her research interests include internet of things technology, database technology and higher vocational education.

1 Introduction

In actual work and life, in order to solve some business problems or achieve goals, people often need to find the data information or knowledge they need in some complicated data. However, this useful information is not readily available, and it is often hidden in the massive datasets with a lot of noise, randomness, fuzziness and even incompleteness (Wang et al., 2021). At this time, people need to take effective measures or methods to strip out this useful information, and data mining can just achieve the above process. In the research of data mining methods, Abnormality detection is a very important content of data mining. Abnormality detection is widely used in all fields of our life and work. Therefore, it is very necessary to study a reasonable and efficient data mining method to meet people's demand for anomaly data mining (Jian et al., 2019).

Mobile terminals are widely used in various fields of life and work because of their powerful storage function, computing power and rich communication methods. The access behaviours of mobile terminal users are not all legal and reasonable, and some users' access behaviours will be abnormal. If the abnormal behaviours are not identified in time and treated with corresponding measures, it is likely that individual businesses can't be completed or enterprises will suffer huge economic losses. Therefore, in recent years, many scholars at home and abroad have spent a lot of energy on the research of abnormal data mining of mobile terminals. For example, Chen and Zhang (2020) studied the data mining model of mobile terminal users' preference based on interval constraint in view of the fact that there are many personalised demands of mobile central segment users at present, which leads to data loss easily, low mining accuracy and low efficiency when they continue data mining. The method first analyses the user's preferences sensitive characteristics, and then USES the preference degree of cost matrix to fill such preference formation of deflection, build node class label distribution rule, finally using the adaptive adjustment of favour sensitive degrees to generate the optimal decision tree, accurately judge the user preference information, according to the constraint characteristics build interval sets, on the basis of the interaction between the user information to establish the user's preference vector By using user preference dot product mining to calculate user preference differences, the mobile terminal user preference data mining model is constructed. Experimental results show that this method improves the validity of data application and the accuracy of data mining to a certain extent, but it takes a long time to excavate due to the complexity of calculation. Mao (2019) proposed a data mining method based on fuzzy association rules to solve the problems of high computational complexity and long time consuming of traditional data mining methods. The method for mining association rules based on fuzzy theory, first implementation of fuzzy association rules, and then based on this, will be blurred mass data processing, and fuzzy database, and the classifying data in the database to gather and continuous attribute discretisation, determine the frequent association rules of fuzzy database, so as to realise the huge amounts of databased on fuzzy association rules mining method. This method can reduce the time and memory consumption in data mining, but the data validity is not improved. Feng et al. (2022) studied the method of travellers' emotion perception based on social media data mining and further applied the data mining model in view of the difference of users' needs. This method is mainly based on the Sina Weibo data released by tourists from 2017 to 2019. BERT model is used to conduct text analysis on the weibo content, explore the temporal and spatial distribution law of tourists' emotions and the emotional characteristics of tourists under different themes, and analyse the related factors leading to tourists' negative emotions (Feng et al., 2022). This method improves the data validity through in-depth mining of mobile terminal data, but it also has the problem of time-consuming and labour-intensive data mining. Zhu (2020) proposed A spatio-temporal feature mining algorithm based on pattern growth multiple minimum support for the low clustering efficiency of moving objects in the convergence mode mining of the internet of things (IoT). Based on the temporal characteristics of user trajectories, this method mines frequent and asynchronous periodic temporal and spatial motion patterns. First, the location sequence is modelled and time information is added to the model. Then, the mining algorithm of asynchronous periodic sequence pattern is adopted. The algorithm is based on multiple minimum support for pattern growth. The sequential patterns of asynchronous cycles are deeply recursively mined based on multiple minimum supports. Finally, the method is verified and evaluated by Gowalla dataset. This method has certain applicability, but its operation rate is slow in the application process.

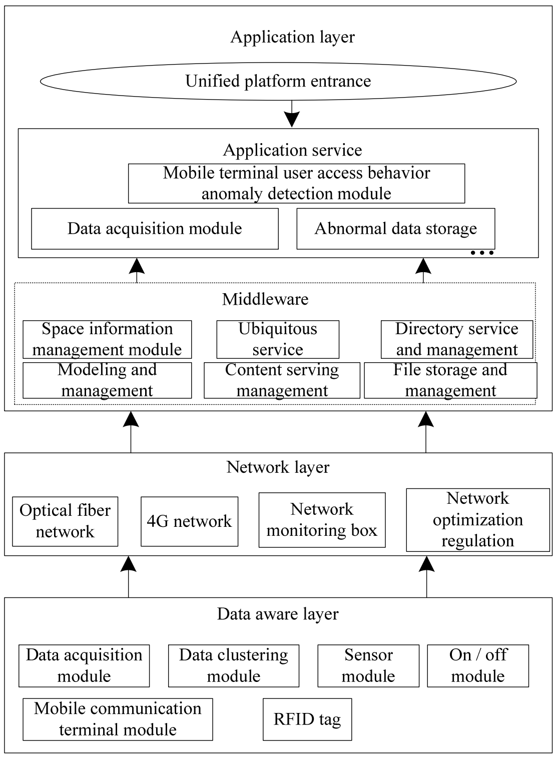
Therefore, a mobile terminal data mining method based on the IoT is proposed in this paper, which can better meet the actual data mining needs. In this study, the optimisation framework of mobile terminal data mining was firstly established based on IoT technology, and the mobile agent wireless sensor data acquisition technology was used to complete the extraction of mobile terminal data, and then the chaotic search particle swarm k-means algorithm was used for clustering. Applying the improved PSO algorithm to K-means clustering can find a better initial centre for it and improve its clustering effect. The clustered data is imported into the abnormal access detection module, and the abnormal data is detected to judge the abnormal type, so as to realise the mining of user abnormal access behaviour. Finally, the paper proves the advance of this method through experiments, hoping to provide reference for mobile user data mining.

2 Optimisation design of mobile terminal data mining method

2.1 *Design of the architecture for mobile terminal data mining*

The IoT and cloud computing technology, and various sensors and various data acquisition sensing technologies are used to accurately collect data information, and the information is transmitted to the data centre through the communication network. The advantages of cloud computing in data processing are given full play to place data on a suitable server, effectively mine data information, and obtain required resources that are of great significance to people's life and work. The overall architecture for mobile terminal data mining optimisation platform based on IoT designed in this paper is shown in Figure 1.

Figure 1 Architecture of data mining optimisation platform

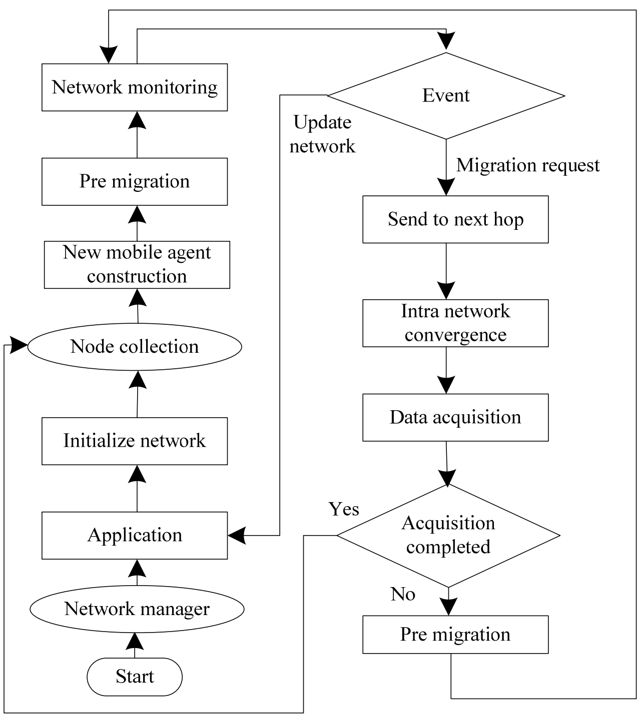


As the first layer of the overall architecture for mobile terminal data mining optimisation based on the IoT, the data perception layer consists of data acquisition module, data clustering module and communication terminal module. These modules can effectively collect and cluster all kinds of data of mobile terminals. In order to obtain more comprehensive and complete data, in the actual data acquisition process of mobile terminals, a large number of sensor nodes can be arranged in the data acquisition area, which can not only obtain comprehensive and complete data, but also significantly improve the efficiency of data acquisition.

The network layer is in the middle of the overall architecture for mobile terminal data mining based on the IoT, and its main responsibility is to connect the data sensing layer with the application service layer. The mobile terminal data collected by the data sensing layer can be transmitted to the application service layer through the optical fibre network and the 4G network.

The application layer is the access port of IoT and users, and the efficient utilisation of IoT is realised by this layer. The addition of cloud platform also makes the management of various application services more efficient. It has powerful data storage and processing functions, and users can access the platform portal according to actual needs, and use various applications that meet users' needs. Therefore, it has absolute advantages in resource utilisation.

Figure 2 Data acquisition process by sensor with the agent technology



2.2 Data acquisition of mobile terminals

After the data mining platform is established, it is necessary to collect the data of mobile terminals. In order to reduce the loss rate of data acquisition, this paper uses mobile agent technology to collect data. When an agent encounters a failure point in the process of migrating to a node, it can independently adjust the migration route and find a new route to migrate, thus avoiding this failure point, which can effectively reduce the probability of data loss in the process of data acquisition. The data acquisition process of mobile terminals by the agent technology is shown in Figure 2.

As shown in Figure 2, after the network manager performs the network initialisation operation according to the user's needs, a new mobile Agent will be built through the collection node and it will be sent out. The sent mobile Agent will acquire the data of all sensors one by one along the migration path. It should be noted that, in general, the sensor data acquisition network based on Agent technology contains a monitoring mechanism, whose main function is to sense the change of the routing network and the migration demand of mobile Agent in real-time, so as to achieve the purpose of timely adjusting the migration path of mobile Agent and completing sensor data acquisition. When the data collection task of mobile terminals is finished, the mobile Agent will bring the collected data back to the collection point.

2.3 Clustering of mobile terminal data

The sensors used to collect data contain a lot of redundant data. In order to reduce the complexity of subsequent data processing, the data clustering module of the data awareness layer is used to perform IPK-means data clustering operation on the mobile terminal data collected by sensors (Jiang et al., 2019; Zamzami and Bougila, 2020; Sheng et al., 2020). Because k-means algorithm cannot find a good initial centre, it cannot cluster the data well, which affects the data mining effect. Based on this, the engineering problem is transformed into a mathematical optimisation problem to improve its clustering ability. Based on this, the improved particle swarm optimisation algorithm is used to optimise k-means clustering algorithm. The chaos search process was added to PSO to increase the diversity of PSO in the late PSO iteration, and a dynamic adjustment factor formula was given to make the adjustment factor correlated with the fitness value of the particle in the updating process. Finally, the improved PSO algorithm is applied to k-means clustering to find a better initial centre for it (Veit and Jensen, 2019). The data clustering process can be summarised as follows:

1 Initialise the particle population. After random selection of a mobile terminal data as a cluster centre, the other cluster centres are selected in turn according to the description in the principle of maximum distance. After numerous operations, initial particles can be obtained. The solution process of distance in the principle of maximum distance can be realised by the following formula:

$$dis\left(z_j, E\right) = \sum_{i=1}^{|E|} dis\left(z_j, E\right) \tag{1}$$

where the cluster centre set and the current cluster number are represented by E and $|E|$, respectively, and data of the j^{th} mobile terminal in the mobile terminal dataset is represented by z_j.

2 Based on the selected cluster centre, the cluster dataset is segmented in each initial particle according to the relevant description of the minimum distance, and then

the fitness of the initial particle is calculated to obtain the individual extremum of each particle and the particle position of individual extremum, as well as the global maximum value and its particle position (Xu and Li, 2019). The fitness function can be solved by the calculation formula of DBI index, which can be expressed as follows:

$$fit = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq 1}\left(\frac{\bar{E}_i + \bar{E}_j}{\left\| w_i - w_j \right\|_2}\right) \tag{2}$$

where the average distance and inertia weight of cluster are represented by $\left\| w_i - w_j \right\|_2$ and k, respectively;

Euclidean distance and cluster population are represented by and respectively.

3 The dynamic factor is adjusted and the particle velocity and position are updated.

4 The particles after the update operation are taken as the new clustering centre, and the clustering operation is performed once again on the mobile terminal data. The fitness of each particle is solved.

5 The particles whose extremums have never changed and those that can't become the global optimum in the particle swarm are eliminated, and the generation operation of the equivalent particles at the global optimum position is completed with the help of chaotic search.

6 *thre* is used as a threshold, *var* as the fitness variance of the current particle population, and *l* as the maximum iteration number. When $var < thre$ or $var \geq l$ is satisfied, turn to the next step, otherwise, turn back to step (3). The fitness variance of the current particle population can be solved by equation (3).

$$\mathrm{var} = \frac{1}{n}\sum_{i=1}^{n}\left(f_i - f_{avg}\right)^2 \tag{3}$$

The fitness and average fitness of population i are f_i and f_{avg} respectively.

7 The best position obtained by the particle swarm optimisation algorithm is used as the initial centre to construct the category matrix, and the iterative operation is not performed when constructing the category matrix (Huang et al., 2021).

8 The cluster dataset is divided again according to the relevant regulations in the principle of the nearest distance, and the category matrix is updated. The update of the category matrix can be expressed by the following formula:

$$u_{ij} = \begin{cases} 1, & z_i \in E_j \\ 0, & z_i \notin E_j \end{cases} \tag{4}$$

9 If there is no change in the category matrix, the iteration is finished, otherwise, step (10) is executed.

10 The average value of each cluster is solved, and this average value is taken as the cluster centre to complete the next round of clustering. In this process, empty clusters are deleted directly once they occur to make the cluster satisfy.

11 If the number of iterations at this time is greater than or equal to the set maximum number of iterations, then the clustering process is finished, and the result can be output and the clustering index can be solved; otherwise, it should turn back to step (8). The solution process of index cluster index can be expressed by the following formula:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq 1}\left(\frac{\bar{E}_i + \bar{E}_j}{\left\| w_i - w_j \right\|_2}\right) \tag{5}$$

where the average distance of clusters is represented by $\bar{E}_i$; Euclidean distance and cluster population are represented by $\left\| w_i - w_j \right\|_2$ and k respectively.

2.4 Mining of abnormal behaviour of users

After acquisition of the processed data, the user abnormal behaviour detection module of mobile terminals in the application service layer is used to mine the data to detect the abnormal behaviours of users. By detecting the abnormal features of user access behaviours, determining the abnormal type and checking the abnormal evolution, the detection of abnormal user access behaviours at mobile terminals can be completed. In essence, detecting abnormal features means taking some measures to show the historical track of mobile terminal users' visits in the form of spatio-temporal co-occurrence areas, and completing the mining of mobile terminal users' behaviours and group structure based on the obtained spatio-temporal co-occurrence areas.

Determining the abnormal type needs to be based on the abnormal characteristic value of the mobile terminal users, and done with the abnormal behaviour detection model (Wang et al., 2018). Because the random forest algorithm has the advantages of high data training efficiency, high classification accuracy and no over-fitting, this paper builds a mobile terminal user abnormal behaviour detection model based on the random forest algorithm. The main process can be summarised as follows:

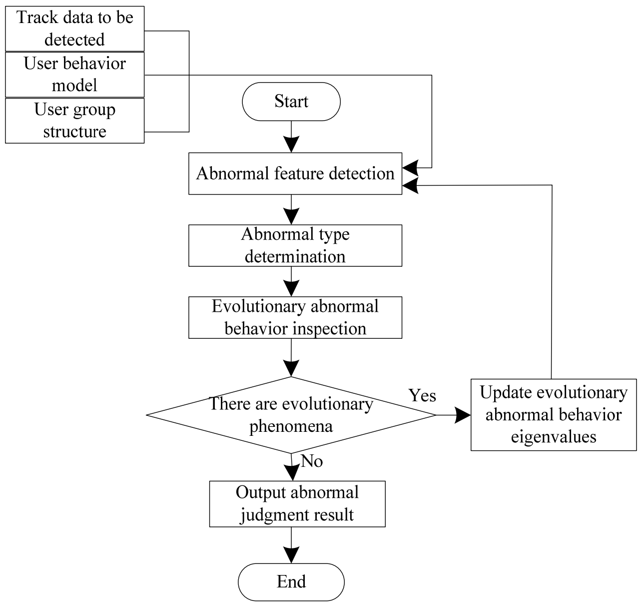
1 According to description in the *Boostrap* method, the sample subset extraction operation is performed on the mobile terminal data training sample set. It should be noted that the sample subset should be extracted in a way that can be put back.

2 Taking the maximum growth as the tree construction rule, the *Cart* tree construction operation is performed on each sample subset.

3 The *RF* classification model is constructed by the generated trees *Cart*, and the classification process is finally completed by model voting.

To check the evolution of abnormal behaviours, it is necessary to judge whether the abnormal eigenvalues of mobile terminal users are consistent with the abnormal results obtained (Jia, 2021). In order to achieve a better detection effect of abnormal behaviours of mobile terminal customers, abnormal feature detection, abnormal type determination and evolutionary abnormal behaviour inspection are indispensable (Abbasi et al., 2018). The specific mining process of abnormal behaviours of mobile terminal users is shown in Figure 3.

Figure 3 Mining process of abnormal behaviours of mobile terminal users



3 Experiment and analysis

The management network mobile terminal of a large power company in D city was taken as the experimental object. The network mobile terminal was a kind of intelligent terminal equipment closely connected with the IoT. Taking Windows 10 system with 8 GB of running memory as the operating system, with the help of Matlab platform, the data collection and analysis of the method proposed in this paper were completed, and the abnormal behaviours of users were detected to verify the feasibility and effectiveness of this method in data mining.

3.1 Feasibility experiment

After design of the method, the feasibility experiment of the method proposed in this paper was carried out to test its ability to detect users' abnormal behaviours, data clustering effect and data collection effect, and the energy consumption of the above operations were recorded.

The method proposed in this paper was used to detect abnormal user behaviours among 10 mobile terminal data samples containing abnormal data. The results of abnormal user behaviour detection are shown in Table 1.

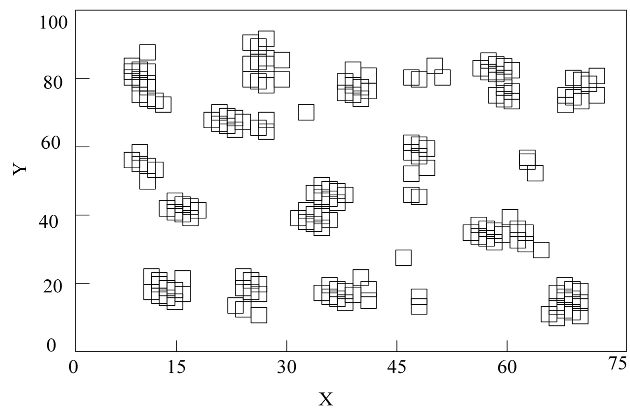
Table 1 Detection effect of abnormal user behaviour of mobile terminals

Number	*Actual abnormal condition*	*This method is used to detect abnormal conditions*	*Exception type*
1	Y	Y	Individual abnormality
2	Y	Y	Group anomaly
3	N	N	No abnormality
4	Y	Y	Individual abnormality
5	Y	Y	Spatiotemporal anomaly
6	Y	Y	Event exception
7	Y	Y	Event exception
8	N	N	No abnormality
9	Y	Y	Group anomaly
10	Y	Y	Individual abnormality
11	N	N	No abnormality

As can be seen from Table 1, the abnormality detection of mobile terminal users' behaviours can be realised by the method proposed in this paper, and the detection results are consistent with the actual situation, suggesting that this method can effectively mine users' abnormal behaviours and better meet the actual work needs.

In the process of mining data of mobile terminals, the collected data of mobile terminals is complex, and there are many duplicate and redundant data, which will reduce the efficiency of data mining. Therefore, in order to improve the efficiency of data mining of mobile terminals and quickly obtain useful information, the method proposed in this paper was applied to cluster 100 collected data of mobile terminals. The clustering effect is shown in Figure 4.

Figure 4 Data clustering effect



In Figure 4, the collected mobile terminal data is successfully divided into 18 categories, and only 2 data are not clustered. The results imply that the method proposed in this paper has a good clustering effect when clustering data of mobile terminals and can effectively reduce the complexity of data processing and improve the efficiency of data mining of mobile terminals.

In order to verify the advantages of this method in data acquisition, the data acquisition results of mobile terminals under different noise interference environments are drawn, and the data acquisition results obtained by this method are shown in Figure 5.

Figure 5 Data acquisition effect

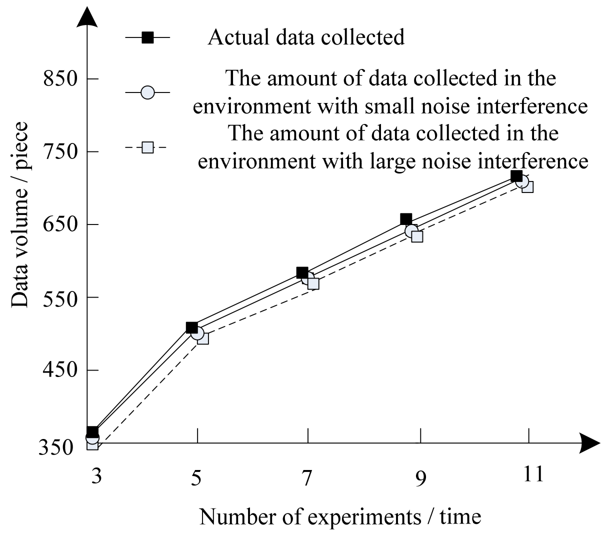
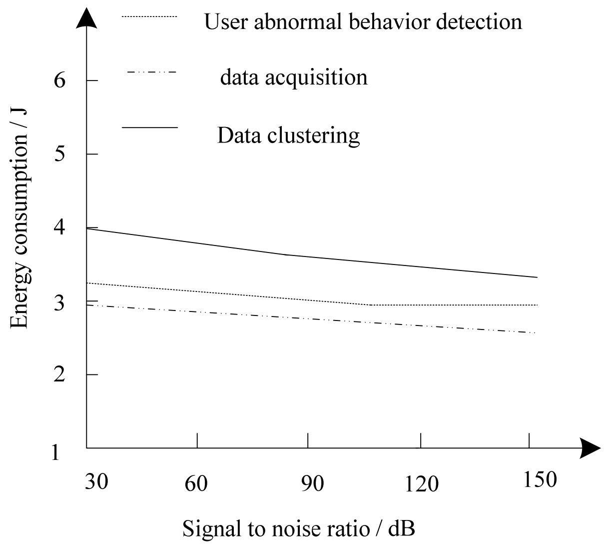


Figure 6 Energy consumption of data mining



In Figure 5, the data of mobile terminals collected by this method is close to the actual data in each experiment, and the data volume curves almost coincide under any degrees of noise environment. The experiment proves that this method can achieve accurate data collection of mobile terminals, which is conducive to the subsequent work of mining abnormal behaviours of mobile terminal users.

The application of this method in data mining of mobile terminals is involved in data collection, data clustering and abnormal behaviour detection of mobile terminal users, and its energy consumption is shown in Figure 6.

It can be seen from Figure 6 that when this method is used to mine mobile terminal data, the energy consumed in data collection, data clustering and abnormal behaviour detection of users is low, and even in a noisy environment, the energy consumed does not exceed 4J. The results imply that this method only needs a little energy consumption to complete the data mining of mobile terminals, and it has obvious advantages in the actual data mining work.

Figure 7 Abnormal access distribution of mobile terminal users, (a) actual distribution (b) distribution with the method proposed in this paper (c) distribution with the method proposed in Chen and Zhang (2020) (d) distribution with the method proposed in Mao (2019)

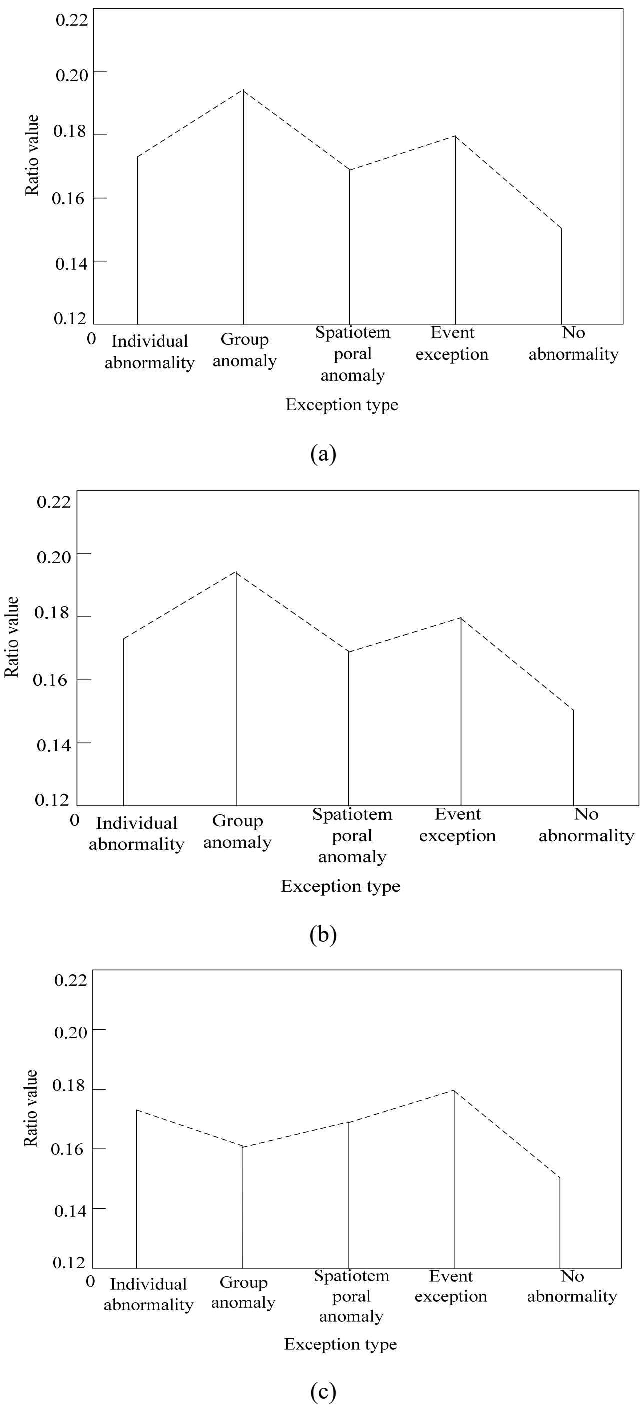
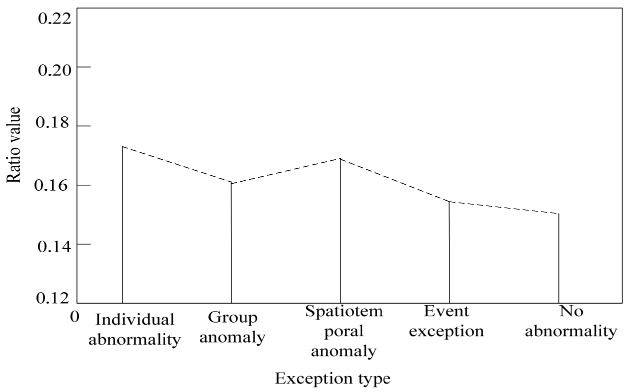


Figure 7 Abnormal access distribution of mobile terminal users, (a) actual distribution (b) distribution with the method proposed in this paper (c) distribution with the method proposed in Chen and Zhang (2020) (d) distribution with the method proposed in Mao (2019) (continued)



(d)

3.2 Comparative experiment

One hundred twenty experiments have been carried out on the method proposed in this paper, the mobile terminal data mining method based on interval constraint proposed in Chen and Zhang (2020) and the mobile terminal data mining method based on fuzzy association rules proposed in Mao (2019) respectively to compare the abnormal distribution effect of mobile terminal users and judge whether it is consistent with the actual situation. Through this process, the accuracy of data mining of the three methods is judged. The distribution of abnormal access types of mobile terminal users on the samples obtained is shown in Figure 7.

From the analysis of Figure 7, it can be seen that the abnormal distribution effect map of mobile terminal users obtained by the method proposed in this paper is completely consistent with the actual abnormal distribution, while that obtained by the other two methods is obviously inconsistent with the actual abnormal distribution. The experiment verifies that the method proposed in this paper is more advantageous than the other two methods in mining mobile terminal data, and can better meet the needs of data mining in practical work.

4 Conclusions

Based on the poor effect of mobile terminal data mining, this paper proposes a mobile terminal data mining method based on the IoT. This method uses the IoT technology to complete the construction of mobile terminal data mining optimisation framework, and uses the mobile agent wireless sensor data acquisition technology to complete the extraction of mobile terminal data, and then implements clustering based on chaotic search particle swarm optimisation k-means algorithm to realise abnormal data detection and user abnormal access behaviour mining. This method can improve the data collection effect and meet the actual demand of mobile terminal data mining. Experiments prove the advanced nature of this method, and the conclusions are as follows:

1 The mobile terminal data collected by the method proposed in this paper are close to the actual data, no matter in the environment with high noise interference or low noise interference, and the data curve almost coincides with the increase of the number of experiments.

2 When clustering mobile terminal data, the clustering effect of this method is relatively good. In the experiment, 100 data are divided into 18 categories, and only 2 data are not clustered, which has little impact on the results and can effectively reduce the complexity of data processing.

3 When this method is used for mobile terminal data mining, users consume low energy in data collection, data clustering and abnormal behaviour detection, even in a noisy environment, the energy consumption is less than 4J, so this method has low consumption.

4 In comparison experiment, three methods are used to detect data at the same time, but only the abnormal data type distribution detected by the method proposed in this paper is completely consistent with the actual distribution, which is superior to other comparison methods, and the mined data has a certain effectiveness.

It is hoped that this paper can provide references for the research of mobile user data mining methods.

References

Abbasi, B., Calder, J. and Oberman, A.M. (2018) 'Anomaly detection and classification for streaming data using PDES', *SIAM Journal on Applied Mathematics*, Vol. 78, No. 2, pp.921–941.

Chen, Q. and Zhang, C. (2020) 'Data mining model of mobile terminal user preferences under interval constraints', *Computer Simulation*, Vol. 37, No. 11, pp.320–323.

Feng, Z., Peng, X. and Wu, Y. (2022) 'Touristemotion perception based on social media data mining', *Geography and Geographic Information Science*, Vol. 38, No. 1, pp.31–36.

Huang, S., Kang, Z., Xu, Z. and Liu, Q. (2021) 'Robust deep k–means: an effective and simple method for data clustering', *Pattern Recognition*, Vol. 117, No. 8, p.107996.

Jia, X.Q. (2021) 'Time series data mining algorithm based on multi? Objective decision', *Electronic Design Engineering*, Vol. 29, No. 17, p.5.

Jian, C., Yang, M. and Zhang, M. (2019) 'Mobile terminal gesture trajectory recognition based on improved LSTM model', *IET Image Processing*, Vol. 13, No. 11, p.143.

Jiang, Y., Bi, A., Xia, K., Xue, J. and Qian, P. (2019) 'Exemplar-based data stream clustering toward internet of things', *The Journal of Supercomputing*, Vol. 76, No. 7, pp.376–391.

Mao, X.J. (2019) 'Research on massive data mining method based on fuzzy association rules', *Microelectronics and Computer*, Vol. 35, No. 2, pp.89–93

Sheng, W., Wang, X., Wang, Z., Li, Q. and Chen, S. (2020) 'A differential evolution algorithm with adaptive niching and k-means operation for data clustering', *IEEE Transactions on Cybernetics*, Vol. 31, No. 99, pp.1–15.

Veit, S. and Jensen, O.N. (2019) 'Vsclust: feature-based variance-sensitive clustering of omics data', *Bioinformatics* (Oxford, England), Vol. 34, No. 17, pp.2965–2972.

Wang, J., Jia, S., Zhao, H., Xu, J. and Lin, C. (2018) 'Internet anomaly detection based on complex network path', *IEICE Transactions on Communications*, Vol. 101, No. 12, pp.2397–2408.

Wang, L., Lin, B., Chen, R. and Lu, K.H. (2021) 'Using data mining methods to develop manufacturing production rule in iot environment', *The Journal of Supercomputing*, Vol. 78, No. 3, pp.4526–4549.

Xu, W. and Li, J. (2019) 'An improved algorithm for clustering uncertain traffic data streams based on hadoop platform', *International Journal of Modern Physics B*, Vol. 33, No. 19, p.134.

Zamzami, N. and Bougila, N. (2020) 'High-dimensional count data clustering based on an exponential approximation to the multinomial beta-liouville distribution', *Information Sciences*, Vol. 524, No. 8, pp.1256–1269.

Zhu, A. (2020) 'Spatiotemporal feature mining algorithm based on multiple minimum supports of pattern growth in internet of things', *The Journal of Supercomputing*, Vol. 76, No. 12, pp.9755–9771.