Mining and Analyzing Behavior Patterns of Smart Home Users Based on Cloud Computing

Xing-hua Lu[(⊠)], Chang-shen Mo, Xiao-qi Wang, and Qing-qing Ma

Huali College Guangdong University of Technology, Guangzhou 511325, China
`luxinghua5454565@l63.com`

Abstract. Aiming at the problem of poor robustness of traditional user behavior pattern mining analysis method, a cloud computing-based intelligent home user behavior pattern mining analysis method is designed. Intelligent household IoT from using cloud computing method and data mining the user behavior patterns, establish a two-layer neural network level of data is divided into 2 kinds, the user behavior mode by setting the input weight vector calculation after classifying data correlation between user behavior model, using Apriori algorithm, input minimum support and minimum confidence, on the basis of analyzing the correlation between data, and establish the user behavior mode decision tree, on the basis of complete analysis of cloud computing smart home user behavior patterns mining method design. Through the comparison experiment with the traditional method, it is concluded that the designed mining analysis method based on cloud computing has higher robustness, the proposed cloud computing-based intelligent home user behavior pattern mining method has good application space.

Keywords: Cloud computing · Smart home · User behavior pattern

1 Introduction

Smart home is a residential platform, using integrated wiring technology, network communication technology, security technology, automatic control technology, audio and video technology to integrate home life related facilities, and build an efficient management system for residential facilities and family schedules. Unlike traditional home systems, smart home systems based on IoT technology are systems that use communication networks to connect critical appliances and services and allow remote control, monitoring or access. According to the three-layer architecture of the Internet of Things, the first layer of the Internet of Things is a perceptual extension system, which is mainly responsible for collecting information and control equipment. The second layer is a heterogeneous converged ubiquitous communication network, which is mainly responsible for the interconnection and storage of heterogeneous networks. Processing; the third layer is the application and service layer, mainly responsible for data monitoring, data analysis, etc.

Literature [3] based on the Internet of Things and cloud computing, research on smart home security access control technology, through analysis of existing home appliance control systems, propose improvements in remote control and information security, and apply the improved solution to home appliance control systems This system enables access based on webpage and Android smartphone control mode, and can wirelessly remotely control household appliances such as color TVs, air conditioners, and fans. Through the communication interface provided by the system, the scene mode control with security alarms and environmental monitoring in the home is realized. However, the robustness of this method is poor.

The method of reference [3] divides the behavior pattern of mobile users into three grades according to the degree of interest, namely, low interest, ZTE interest and pleasure, retaining the behavior pattern with high interest, eliminating the behavior pattern with low interest, and then mining the association rule of behavior pattern with higher interest degree by using data mining method. Use these rules to make certain recommended services to users in order to provide better service for mobile users. Although the method can classify user behavior patterns according to grade, it is not suitable for smart home users.

Through the mining and analysis of user behavior patterns, the home system can better understand the user's living habits, and design the electrical switch time and the electrical start-up amount according to the user's habits, so as to simplify the user's life, environmental protection and low power consumption. Cloud computing is a kind of distributed computing. It refers to the decomposition of huge data calculation processing programs into countless small programs through the network "cloud". Then, through the system composed of multiple servers, processing and analyzing these small programs to obtain the results. And return to the user. The related performance of cloud computing can meet the mining and analysis of intelligent user behavior patterns. Based on the above analysis, this paper studies the cloud computing-based smart home user behavior pattern mining analysis method, the following is the research content.

2   Cloud Computing-Based Smart Home User Behavior Pattern Mining Analysis Method

2.1   Cloud Computing Mining User Behavior Pattern Data

In the cloud computing environment, the user behavior pattern is mined. User functions are mined using Map and Reduce functions of the MapReduce parallel programming model. The main idea of MapReduce processing data is to divide a relatively large data set into smaller data sets, hand a small data set to each node in the cluster for calculation, and then aggregate the key-value pairs with the same key value. Then, it is handed over to different nodes for calculation and the final result is obtained [3]. The MapReduce mining data process diagram is shown below.
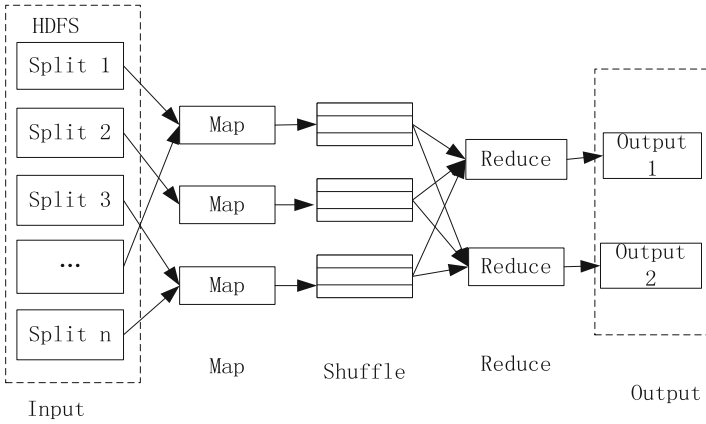
Fig. 1. Data mining process diagram

In the mining of smart home user behavior pattern data, each node calculates data from the smart home Internet of Things according to the data clustering principle. The calculation formula of each node, that is, the smart home user behavior pattern similarity matrix is as follows.

$$Q_{mn} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m1} & q_{m2} & \cdots & q_{mn} \end{bmatrix} \tag{1}$$

In the formula (1), q_{mn} represents the similarity between the smart home user behavior pattern data m and the smart home user behavior pattern data template n. The similarity matrix between the data of the two smart home user behavior patterns is as shown in the following figure [4].

$$R_{mn} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{mn} \end{bmatrix} \tag{2}$$

In formula (2), R_{mn} represents the similarity between the two smart home user behavior pattern data of mode data m and mode data n.

According to the process diagram shown in Fig. 1, the user behavior pattern data in the smart home is extracted from the Internet of Things. Store the mined user behavior pattern data in HDFS. HDFS is able to split large file data into many small blocks of data. Modify the size of the block as needed. The default size is 64 MB. HDFS follows

the three-way replication system, ensuring that files residing in HDFS remain unchanged in three different node assignments, as shown in the following figure [5] (Fig. 2).

NameNode

metadata

DateNode

| 2 | 4 |
| 1 | 5 |

| 5 | 2 |
| 3 | |

	3
4	
	1
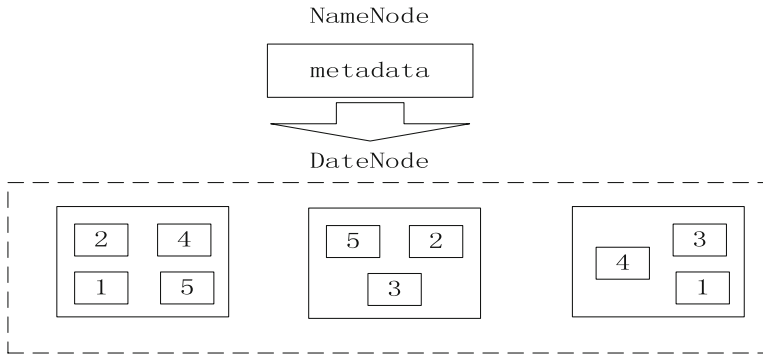
Fig. 2. HDFS block multi-copy storage

Storing with multiple copies allows you to read data from any node that has a block of data, thereby increasing the speed at which data is read. At the same time, it is also beneficial to compare between different copies to ensure the accuracy of the data. When a DataNode is down or the data disappears, it can be retrieved from other copies. When HDFS stores new files, it calculates its own checksum for the new file and keeps the checksum in its own namespace to hide the file independently. The checksum is sent to the client together with the data. After receiving the data, the client performs local check. If the locally calculated checksum is different from the checksum in the data, the data is sent incorrectly and needs to be retrieved from other nodes. This data. After mining the smart home user behavior pattern data using cloud computing, the user behavior patterns are classified for subsequent analysis and processing.

2.2   User Behavior Pattern Classification

In a smart home system, there are many variations in user behavior. In view of the classification of smart home users, they can be divided into entertainment, security, health and other categories. Considering the security of smart homes, this section divides all the user behavior patterns of cloud data storage into two types, which can be directly executed. And cannot be executed directly. The smart home user behavior mode that can be directly executed means that the user behavior action can be controlled by the smart home gateway and directly executed by the smart home terminal, without first confirming whether the event is executed by the user. The smart home behavior mode that cannot be directly executed means that this type of user action may affect the daily life of the user after execution, and may even affect the user's life safety. Such behavior requires the cloud to first alert the user through the smart home gateway after matching the user behavior. After the user selects, the smart home system will perform related events.

Set the primary input data set containing n data to $X = (x_1, x_2, \cdots, x_n)$, where $x_1, x_2, \cdots, x_{20}$ is used for primary mode identification and $x_{21}, x_{22}, \cdots, x_n$ is used for secondary mode classification. In these n items, if an item matches, the item is set to 1, and vice versa. With the use of users, the user's behavior pattern will continue to increase, and the primary and secondary mode classification modes are dynamically added in the smart home system [7].

After receiving the data, the cloud queries the data warehouse for the mode set of the user, and performs pattern matching on the data from the mode set. Finally, after competition, a pattern with the highest matching rate is obtained. The following figure shows a flow chart for classifying user behavior patterns using a classification algorithm (Fig. 3).
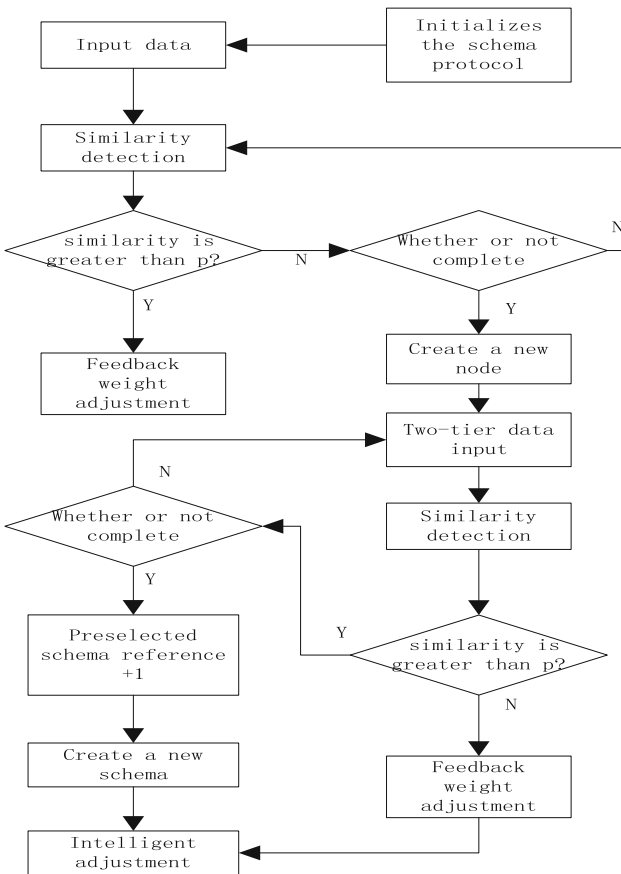

Fig. 3. User behavior pattern classification flow chart

First, initialize the weight vector: Since the system has not been trained yet, there can be only one neuron at the output, and there is no input from the outside world at this time, so the input weight vector is set as follows:

$$b_{ij}(0) = \frac{1}{1+n} \qquad (3)$$

In formula (3), $b_{ij}(0)$ is the input weight vector, and the value is 90. Calculate data similarity and match the data [8]. Taking the similarity threshold of 0.8, the similarity between the data is calculated according to the similarity calculation formula shown in the following equation.

$$sim(u, v) = \frac{u \cdot v}{\sqrt{u^2}\sqrt{v^2}} \qquad (4)$$

In formula (4), u is the matching template data in the database, v is the input data of the neural network, and $sim(u, v)$ is the degree of similarity. $sim(u, v)$ The smaller the value, the more similar the two data are [9]. In order to make the similarity ratio closer to 1 in the subsequent similar matching, the parameters are adjusted so that the output node can be better fused with the data. If the two data are not similar, that is, no suitable pattern matching is found in the data warehouse, then another level node is added to the neural network to make the large class match successfully.

The second layer of neural network pattern matching will be performed after matching to the large class. The overall process is similar to the first layer, except that after a matching pattern is not matched, a pre-selected mode is set. After the number of occurrences of this mode exceeds the threshold, the system will officially set it to the matching mode. After classifying user behavior, calculate the degree of association between user behaviors so that smart home adjustment control services.

2.3 User Behavior Correlation Calculation

What smart homes need is to discover user habits and provide the right services at the right time. Association rule mining is to discover the relationship between attributes in the database, and measure the relevance with support and credibility to meet the requirements.

The data set of the association rule mining is recorded as D, $D = \{t_1, t_2, \cdots, t_k, \ldots, t_n\}$, in the data set $t_k = \{i_1, i_2, \cdots, i_m, \ldots, i_p\}$. $t_k(k = 1, 2, \cdots, n)$ is called transaction, and $i_m(i = 1, 2, \ldots, p)$ is called item. Let $I = \{i_1, i_2, \cdots, i_m\}$ be the set of all the items in D and any subset W of I be called the set of items of D. The number of transactions in the data set D containing the item set W is referred to as the support number of the item set W, and is denoted as σ_w. The support of item set X is denoted as s(W), and the calculation formula is as follows.

$$s(W) = \frac{\sigma_w}{|D|} \times 100\% \qquad (5)$$

In formula (5), $|D|$ is the number of transactions of transaction set D [10]. If $s(W)$ is not less than the minimum support specified by the user, then $s(W)$ is called a frequent item set. If W and V are item sets, and $W \cap V = \Phi$, implied formula $W \Rightarrow V$ is called an association rule, and W and V are respectively referred to as premise and conclusion of association rule $W \Rightarrow V$. The support level of item set $W \cap V = \Phi$ is called the support degree of association rule $W \Rightarrow V$, and is recorded as $s(W \Rightarrow V)$, namely:

$$s(W \Rightarrow V) = s(W \cup V) \tag{6}$$

The confidence level of association rule $W \Rightarrow V$ is recorded as $c(W \Rightarrow V)$, that is,

$$c(W \Rightarrow V) = \frac{s(W \cup V)}{s(W)} \times 100\% \tag{7}$$

After calculating the relevance degree of the smart home user behavior according to the above formula, the classified smart home user behavior data is subdivided again in the same category according to the degree of association.

According to the association rule, the threshold is set to 0.6. In the same user behavior pattern data category, all user behavior pattern data whose association value is greater than the threshold is grouped into one set. User behavior pattern data with a degree of association below the threshold is split into another collection. Until the smart home user behavior data in all categories is subdivided. Use algorithms to analyze smart home user behavior data.

2.4   Analyze User Behavior Pattern Data

Use the Apriori algorithm to analyze smart home users using home device activity associations. First, find all the frequency sets. These frequency sets appear at least at the same frequency as the predefined minimum support. Secondly, strong correlation rules are generated by the frequency set. These rules must meet the minimum support and minimum confidence. Third Use the frequency set found in the first step to generate the desired rule, and generate all the rules that only contain the set items, where each rule has only one right part, and the definition of the middle rule is used here. Once these rules are generated, only those rules that are greater than the minimum confidence given by the user are retained.

Before the Apriori algorithm starts running, input two parameters of minimum support $s(w)$ and minimum confidence $c(w \Rightarrow v)$, and initialize the array data and the vertical array two data structures. Scan the original transaction set to fill the corresponding array vector, and then generate the corresponding vertical array from the array vector. The transaction number of each item is supported by the vertical array statistics (that is, all the transaction numbers in the columns below each item), and the support count of each item needs to be calculated in combination with the number of occurrences of the corresponding transaction in the array vector, and then the support transaction is deleted. The number is less than the minimum support level of $s(w)$,

which in turn leads to a frequent 1-item set. Count the number of frequent 1-item centralized transactions. If it is greater than 2, continue to execute downwards. If less than 2, the current 1-frequent item set is the result of the final frequent item set. Except for the transaction code whose transaction length is less than 2 in the vertical array, the candidate 2-item set is obtained by the two-two connection of the items in the frequent 1-item set. The transaction set supporting the candidate 2-item set can be generated by supporting the connection in the vertical array. The support transaction of the two transactions of the item is obtained by intersection, and the frequent 2-item set can be obtained by deleting the candidate 2-item with the support less than the minimum support. And so on, until all user behavior correlations are analyzed.

In order to enable the smart home system to provide services to users accurately and efficiently, a decision tree is established to analyze the user behavior patterns after analyzing the relevance. The data analyzed by the Apriori algorithm generates a data set containing many attributes according to the degree of association. To create a data set containing many attributes into a tree, you need to split the attributes of the data set. This paper chooses the splitting attribute with the maximum gain rate indicator.

The information gain rate is the ratio of the information gain to the amount of information divided. Assuming that the sample data set for training is H and the data set contains H_s tuples, $s = \{1, 2, \cdots, n\}$, the expectation is:

$$i(H) = -\sum_{s=1}^{n} \frac{H_s}{H} \log_2 \frac{H_s}{H} \qquad (8)$$

Assume that attribute A has m values of $a_1, a_2, \cdots, a_m$ and divides sample data set H into $\{H_1, H_2, \cdots, H_m\}$. After dividing data set H according to A attribute, it continues to calculate the information entropy required for the next split:

$$i_A(H) = \sum_{j=1}^{n} \frac{|H_j|}{|H|} \times i(H_j) \qquad (9)$$

The information gain is the difference between the original information demand and the new demand. The information gain rate normalizes the information gain using the "column information" value. Get the following

$$l_A(H) = -\sum_{j=1}^{n} \frac{|H_j|}{|H|} \times \log_2 \frac{|H_s|}{|H|} \qquad (10)$$

This value represents information generated by dividing the training data set D into m partitions corresponding to the m outputs of the attribute A test. A preliminary user behavior decision tree model can be obtained as shown (Fig. 4):
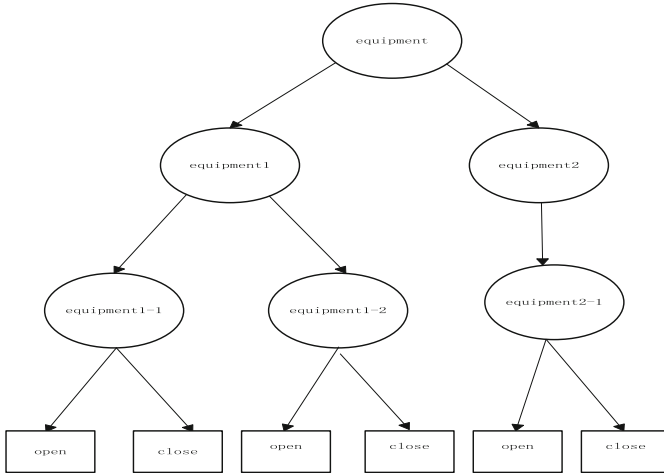
Fig. 4. User behavior decision tree

The smart home system provides users with corresponding services according to the user behavior decision tree, thereby improving the service efficiency of the smart home and reducing energy consumption. So far, the design of cloud computing-based intelligent home user behavior pattern mining analysis method has been completed.

3   Test Experiment

This paper designs a smart home user behavior pattern mining analysis method based on cloud computing. In order to verify the performance of the method, the comparison experiment between the design and the traditional method is carried out, and the relevant verification is completed through experiments.

3.1   Experimental Data Processing

The data used in this experiment is the power consumption data of 2 households for 2 weeks, and is collected by 3 electric meters. The first electric meter collection kitchen mainly includes a microwave oven, an oven, etc.; the second electric meter collects the laundry room, mainly including a washing machine and a drum type dry cleaning machine; and the third electric meter collects a water heater and an air conditioner. Because the data is inevitably erroneous, the missing data points are filled with the average of the last two numbers. The experimental data is processed on a computer configured as shown in the following table (Table 1).

Table 1. Data processing computer configuration.

Project	Configuration	Function
CPU	Intel Core i7-3596 3.25 GHz	Operating platform system
Hard disk	500 G	Storage of experimental data
Graphics chips	AMD RX57960	Drive display
Memory capacity	16 GB	–
Ethernet card	Intel PWLA8391GT	Connect to the internet
Operating system	Windows 8.1	Realize experimental operation

The processed experimental data was used as the experimental object of the experiment, and the relevant experimental verification was completed according to the experimental content.

3.2 Experimental Content

The experiment adopts the form of comparative experiment. The experiment group is the cloud computing-based smart home user behavior pattern mining analysis method designed in this paper. The experimental reference group is the traditional user behavior pattern mining analysis method. The experimental comparison index is the daily power consumption of the smart home system using the experimental group and the reference group. The robustness of the experimental group method and the reference group method is verified by comparing the experimental indexes. In the experimental environment shown in the figure below, the experimental verification is completed (Fig. 5).
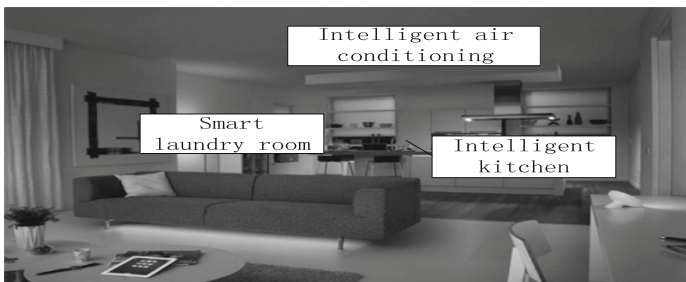
Fig. 5. Lab environment

3.3 Experimental Result

The experimental results are shown in the following table. The data in the table are analyzed and the corresponding experimental conclusions are obtained (Table 2).

Table 2. Power consumption (kW/h)

Experimental days	Experimental group	Control group
1	2.74	5.32
2	2.69	5.06
3	2.55	5.28
4	2.43	5.21
5	2.22	5.54
6	2.28	5.66
7	2.06	5.44
8	2.16	5.05
9	2.16	5.05
10	2.26	5.76
11	2.09	5.49
12	2.15	5.69
13	2.12	5.26
14	2.04	5.60

Analysis of the above table, in the two-week experiment process, the overall power consumption of the smart home system using the experimental group method has a downward trend, and the overall power consumption of the smart home system using the reference group method is constantly fluctuating. Analysis of the data in the table, the power consumption of the smart home system using the experimental group method is significantly lower than the power consumption of the application reference group method. The average power consumption of the smart home system using the experimental group method is 2.28 kW/h, and the average power consumption of the smart home system using the reference group method is 5.34 kW/h, and the reference group is about 2.35 times of the experimental group. In summary, the cloud computing-based smart home user behavior pattern mining analysis method designed in this paper has better robustness and can reduce the energy loss of smart home systems.

4   Conclusion

This paper designs a cloud computing-based smart home user behavior pattern mining analysis method, and compares it with the traditional method to verify that it has better robustness and good application space.

References

1. Xiao, D., Wang, Q., Cai, M., et al.: Research on implicit interference detection based on knowledge graph in smart home automation. Chin. J. Comput. **42**(06), 1190–1204 (2019)
2. Gao, Y., Bao, F.: Research on mobile user behavior mining integrated with location scenarios. Math. Pract. Theory **48**(16), 72–84 (2018)
3. Wang, Y.: Research on smart home security access control technology based on internet of things and cloud computing. Video Eng. **42**(08), 147–150 (2018)
4. Zhai, P., Fang, X., Liu, X., et al.: Decomposed mining method for process model based on behavior feature net. Comput. Integr. Manuf. Syst. **24**(07), 1690–1697 (2018)
5. Luo, H., Yang, Y., Wang, J., et al.: A cloud computing based mobile user behavior analysis system. Control Eng. China **25**(02), 218–223 (2018)
6. Zhang, Z., Zhang, S., Zeng, J.: Method of attacking temporal pattern privacy of users' behavior in social media. Comput. Eng. Appl. **53**(17), 14–19 (2017). +142
7. Chen, X., Xiao, B.: Emerging sequences pattern mining based on location information. Comput. Sci. **44**(07), 175–179 (2017)
8. Ge, X., Ren, C., Duan, J.: Analysis and research on user behavior patterns in the perspective of socialized electronic commerce. J. Anhui Electr. Eng. Prof. Tech. Coll. **22**(02), 91–95 (2017)
9. Li, S., Liu, J., Shao, F.: Analysis of user's browsing interest data mining combining Web log with user's browsing behavior. Mod. Electron. Tech. **40**(05), 22–25 (2017)
10. Shi, D., Li, H., Yang, R., et al.: Mining user frequent behavior patterns in daily life. J. Natl. Univ. Defense Technol. **39**(01), 74–80 (2017). Author, F.: Article title. Journal 2(5), 99–110 (2016)