Hindawi
Journal of Sensors
Volume 2022, Article ID 8277831, 10 pages
https://doi.org/10.1155/2022/8277831

Research Article

A Collaborative Filtering Method for Operation Maintenance Behavior in Power Monitoring Systems

Jinyu Wu, Wenwei Tao, Wenzhe Zhang, Zeming Jiang, and Gang Chen

China Southern Power Grid Co., Ltd., Huangpu District, Guangzhou, Guangdong 510623, China

Correspondence should be addressed to Jinyu Wu; wujinyu0301@163.com

Received 7 March 2022; Revised 12 April 2022; Accepted 18 April 2022; Published 20 May 2022

Academic Editor: Wen Zeng

As an important part of power infrastructure, a power monitoring system provides real-time data acquisition, state detection, and remote control of power equipment for the power grid and can deal with sudden anomalies in time. The operation and maintenance of the power monitoring system are very important to ensure the stable operation of power grid. The current mainstream remote operation and maintenance mode has internal threats such as misoperation of operation and maintenance personnel or malicious damage caused by attackers stealing operation and maintenance authority. Meanwhile, the existing operation and maintenance audit has the problems of high human resource cost and limited supervision of operation and maintenance personnel. To solve this problem, this paper proposes a collaborative filtering method for operation and maintenance behavior of power monitoring system called CFomb. Exploiting a keyword matching algorithm, CFomb determines the power resources accessed by operation and maintenance users from multiple operation instructions and extracts operation and maintenance behaviors. Referring to the collaborative filtering idea, the feature matrix decomposition scheme is introduced to train the access probability model based on the historical normal behavior of multiple operation and maintenance users, which provides a basis for real-time prediction of the access behavior probability of target operation and maintenance users. The OTSU binarization technique is used to determine the probability threshold of abnormal operation and maintenance behaviors, identify abnormal behaviors through threshold comparison, and send real-time alarms to operation and maintenance audit. The simulation experiment results show that the method in this paper can effectively identify the abnormal behavior of operation and maintenance users, reduce the overhead of manual audit, and help improve the power monitoring system's ability to respond to internal threats of operation and maintenance.

1. Introduction

As an important part of power infrastructure, the power monitoring system provides reliability support for the stable operation of power grid. In order to ensure the normal operation of the power operation system, it is necessary to carry out routine operation and maintenance for the system function, related equipment, hardware and software, and other internal resources, to deal with emergencies and abnormalities in time. The power monitoring system is distributed and deployed in different power stations, power distribution stations, and dispatching centers at all levels. Therefore, in case of employing on-site operation and maintenance, on the one hand, it will be difficult to deal with the emergencies anywhere in the system; on the other hand, it will be hard to control the operation and maintenance of power monitoring system. As a result, the operation and maintenance platform of power monitoring system needs to realize both remote and centralized operation and maintenance management.

The current typical operation and maintenance platform of the power monitoring system is generally built based on bastion host [1–3], whose technology can manage operation and maintenance accounts and assets uniformly and set the buffer to allow the assets of the power monitoring system to realize remote operation and maintenance without direct exposure to the outside. The administrator of bastion can configure the access strategy of operation and maintenance users. When logging in to the power monitoring system with fortress machine technology to implement operation and maintenance, the operation and maintenance personnel can record

the operation and maintenance process of operation and maintenance personnel in real time, and the auditors can audit the operation and maintenance process of operation and maintenance personnel according to the audit rules, so as to achieve the supervision of operation and maintenance personnel. In addition, the bastion can isolate the internal resources of the power monitoring system from external exposure, centralize the identity of operation and maintenance users, and achieve centralized access control of operation and maintenance work. However, the remote operation and maintenance mode of the bastion host power monitoring system neglects the protection against internal threats, resulting in the high human resource cost of security audit.

To solve this problem, this paper proposes a collaborative filtering method (CFomb) for the operation and maintenance behaviors of power monitoring system. Exploiting keyword matching algorithm, CFomb determines the power resources accessed by operation and maintenance users from multiple operation instructions and extracts operation and maintenance behaviors; by reference to the collaborative filtering idea of the recommended system, the feature matrix decomposition scheme is introduced to train the access probability model based on the historical normal behavior of multiple operation and maintenance users, which provides a basis for real-time prediction about the access behavior probability of operation and maintenance users; OTSU binarization technique is used to determine the probability threshold of abnormal operation and maintenance behaviors, identify abnormal behaviors through threshold comparison, and send real-time alarms to operation and maintenance audit. Finally, the behavior that multiple operation and maintenance users of the power monitoring system access multiple power resources was simulated, and an experiment was carried out. The experimental results show that the methods proposed in this paper can help build a behavior probability prediction model for users and resources based on user behavior patterns and determine whether random behavior of users is abnormal in accordance with the thresholds generated automatically.

The overall structure of the paper is shown in Figure 1.

2. Related Work

Various information security issues are introduced in the informatization development of the power industry. Among others, the internal threats of information system, as a hot issue in the research on current general information system security, have drawn increasing attention [4–7]. In terms of internal threat recognition methods, the traditional way was to audit the historical access logs of operation and maintenance users and detect the internal attacks that occurred by afterward examination. For example, Liu et al. [8] proposed the Log2vec method to detect the abnormal behavior of system. Based on the log information, this method extracts multiple factors, such as the sequential relationship between user behavior sequences within a day, the relationship between behavior sequences on different dates, and the behavior topology relationship of resource access by users, to carry out behavior modeling; the logging behavior is transformed into a vector using graph neural networks; the behavior vector is separated from abnormal behavior with the clustering algorithm, and then the insiders responsible for the abnormal behavior are traced. Gu and Guo [9] proposed an internal threat detection method based on role abnormal behavior mining, which mines the role abnormal behaviors using the sequence pattern and carries out pattern matching with KMP algorithm to recognize abnormal users. However, these methods require studying the historical data offline, failing to analyze new data in real time and to locate the malicious behaviors in real time. Therefore, it is difficult to avoid the losses caused by internal threats.

Rashid et al. [10] simulated the weekly normal behaviors of each user using the hidden Markov model and then applied them to the detection of the significant deviation between abnormal behaviors and normal behaviors. Happa [11] used the EM algorithm to train a GMM for the behaviors of each user in the first month, to simulate the normal behaviors of the user. The trained GMM is applied to computing the likelihood of input observations to indicate the possibility of the input. If the likelihood is smaller than the threshold, the observation will be detected as abnormal.

However, the above methods merely consider modeling the normal behaviors of users using the sequential features of a single user's behaviors, failing to consider the correlation between user behaviors. As recommendation systems analyze and model user behavior data, they predict and recommend products that users do not use but are likely to be interested in. Collaborative filtering algorithm is a key algorithm in recommendation systems. Collaborative filtering can use user behaviors similar to those of the target user to infer the target user's preference for a specific product and then make recommendations accordingly based on this preference. Therefore, by reference to the collaborative filtering idea, in the abnormal behavior detection of internal threats on the operation and maintenance platform of the power monitoring system, this paper not only considers the historical longitudinal features of operation and maintenance users but also integrates the transverse impacts among similar operation and maintenance users.

Compared with the existing work, the main innovations of this paper are as follows.

(1) The introduction of the feature matrix decomposition method to train the access probability model, which is simple to compute, easy to obtain training data, and does not require complex processing, can be applied to more power monitoring system O&M scenarios

(2) Combining collaborative filtering idea and OTSU binarization method to achieve probability prediction of real-time access behavior of O&M users and adaptive selection of probability threshold of abnormal O&M behavior, supporting more efficient and safe development of power monitoring system O&M

(3) Simulation experiments are carried out based on the OTSU access behavior dataset for power monitoring system multiple O&M users to access multiple power

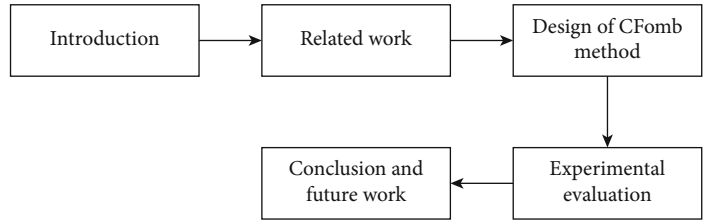

Figure 1: The overall structure of the paper.

resources behavior, and the results show that the proposed method can effectively identify the abnormal behavior of power monitoring system O&M users and can reduce the cost of human O&M audit

3. Design of the CFomb Method

3.1. Overall Design of the CFomb Method. The overall architecture of the CFomb method is shown in Figure 2. The method mainly consists of three modules, i.e., operation and maintenance behavior extraction, behavior analysis, and behavior alarm. The operation and maintenance behavior extraction module and the behavior alarm module are deployed in the bastion host on the operation and maintenance platform of the power monitoring system, while the computation nodes of the behavior analysis module can be deployed separately. In the CFomb method, the operation and maintenance behavior extraction module provides real-time behavior input for the behavior analysis module, which recognizes abnormal behavior and outputs abnormal behavior alarm information to the behavior alarm module, while the behavior alarm module displays the alarm information to the auditor and feeds back to the behavior analysis module.

3.2. Extraction of Operation and Maintenance Behaviors. As the processing object of the CFomb method, the description of an operation and maintenance behavior requires defining the specific operation and maintenance users and objects.

The operation and maintenance user can be a user with the operation and maintenance permission for the server of the power operating system. The user may carry out operation and maintenance based on the server of the power monitoring system within the preset time in accordance with the requirements of operation and maintenance [12]. When an operation and maintenance user carries out the operation and maintenance, the user needs to determine the target power resources to be accessed firstly and then input the corresponding operating instruction into the operation and maintenance platform of power monitoring system based on the determined target power resources. The operation and maintenance platform will obtain the corresponding operating instruction and determine the target power resources in accordance with the operating instruction, so that the target operation and maintenance user can smoothly carry out the corresponding operation and maintenance.

The operation and maintenance users and the power resources shall have unique identification information on the operation and maintenance platform of the power monitoring system. In CFomb method, the operation and maintenance behavior extraction module inserts the instruction extraction points into the bastion host to extract the real-time operation and maintenance instructions of operation and maintenance users. Through comparing and analyzing the instructions and keyword database of operation and maintenance objects, the module determines the operation and maintenance users and the resource objects accessed and outputs the operation and maintenance behaviors. For each instruction input by users, the algorithm indicated in Algorithm 1 is called to extract the user behavior in the instruction. As the input of the behavior analysis module, the behavior information output is used to recognize the current abnormal behavior of a user.

3.3. Analysis of Operation and Maintenance Behaviors. As the core of CFomb method, the behavior analysis module needs to collect the historical access records of users to construct the behavior matrix and obtains the user model and resource model via behavior matrix decomposition. When the behavior analysis module obtains the real-time behavior information input of the behavior extraction module, it can extract the corresponding features from the user model and resource model, predict the probability of occurrence of such behavior, and determine whether the behavior is normal or abnormal based on the probability threshold classification. In addition, the behavior buffer will continuously record user behavior and take it as the training set adjustment information of follow-up iterative training to update the learning model.

3.3.1. Construct the Training Set. Stemming from the statistics on historical behaviors of operation and maintenance users, the training set employs the frequency of resource access by users to describe the probability of resource access by users. Since the scope of historical access behavior and resource access by users is limited, the training set merely contains the information of resource access frequency of a few users.

The training set is represented in the form of frequency matrix $P_{M\times N}$ of M rows and N columns, among which the number of matrix row M represents the total quantity of users on the operation and maintenance platform, while the number of matrix column N represents the total quantity of resources on the operation and maintenance platform, and the behavior space (i.e., the scale of frequency matrix) is $M \times N$. Since the training set is a sparse matrix, it can be stored in the form of triple $(i, j, p_{i,j})$, among which $p_{i,j}$

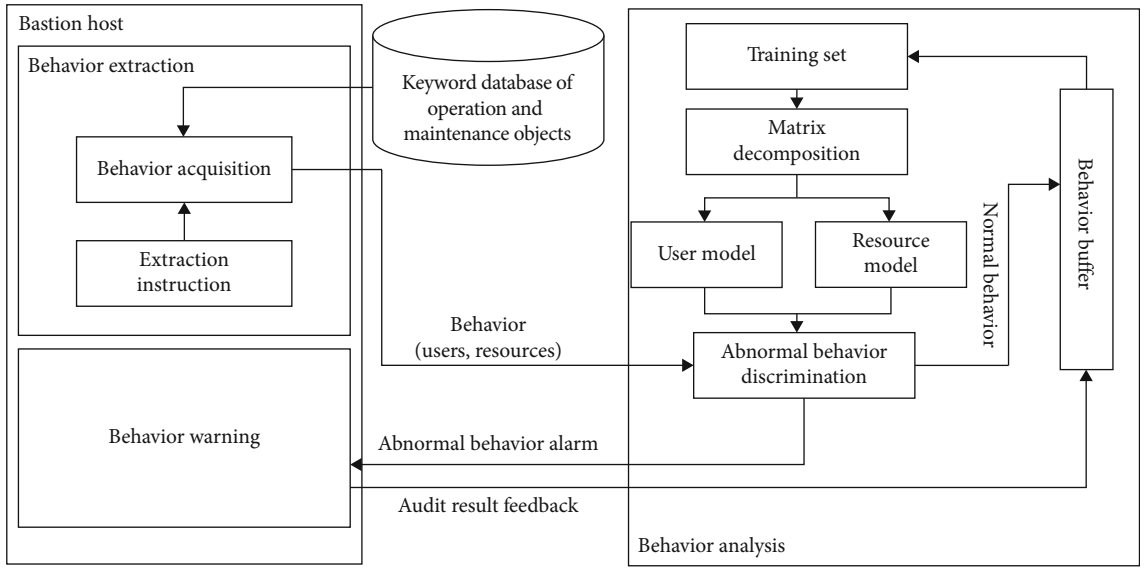

Figure 2: Architecture of CFomb.

Input: a real-time operating instruction character string I of a user, the current user id user_id, and resource keyword database source_db
Output: behavior information{user_id, source_id} containing user id and resource id
1 Divide instruction I into multiple words
2*for word in each words.* /∗Traverse the words in the instruction∗/
3 use the *word* to query the keyword database to obtain the resource id *source_id*
4 if (successful query)
5 output the behavior information {*user_id, source_id*}
6 *end if*
7*end for*

Algorithm 1: Workflow of operation and maintenance behavior extraction algorithm.

represents the element $(1 \le i \le M, 1 \le j \le N)$ in the ith row and the jth column.

Based on the statistics on the historical behaviors of operation and maintenance users, the construction of the training set is updated continuously with the new access behavior of operation and maintenance users. If the record on user behaviors is unavailable in the initial construction stage of operation and maintenance platform, it is necessary to collect the user behaviors for a period of time before enabling the CFomb method for the follow-up work.

3.3.2. Matrix Decomposition. With the aim of predicting the probability of random behavior in behavior space, the CFomb method introduces the collaborative filtering algorithm based on matrix decomposition and carries out matrix decomposition of the training set to construct the feature model of users and resources [13]. The user model represents the matrix $X_{M\times K}$ of M rows and K columns, the resource model represents the matrix $Y_{N\times K}$ of N rows and K columns, among which K represents the implied feature dimension, whose value is much smaller than any one of M and N.

Matrices $X_{M\times K}$ and $Y_{N\times K}$ represent the feature distribution of M users and N resources in K dimensional feature space. The ith column in matrix X and the jth column in matrix Y represent the K dimensional eigenvector of user i and resource j, respectively. The similarity of these $M+N$ eigenvectors can be obtained through inner product computation, which includes similarity between users and resources, similarity among users, and similarity among resources. Among these, the similarity between users and resources represents the prediction of the probability of resource access by users.

Therefore, the user model and resource model obtained by the matrix decomposition algorithm shall guarantee that the product of matrix X and matrix Y approaches, as much as possible, the frequency matrix P indicated in the training set, i.e., frequency matrix P can be represented by X and Y in the form of the following formula.

$$P_{M\times N} \approx X_{M\times K}\left(Y_{N\times K}\right)^{T}. \tag{1}$$

For the user i and the resource j designated randomly in the behavior space, the ith column and the jth column can be taken, respectively, in the user model matrix X and the resource model matrix Y to obtain K dimensional column vectors X_i and Y_j. Therefore, the predictive value of the

Input: Behavior frequency matrix $P_{M\times N}$ of resource access by users
Output: User matrix $X_{M\times X}$ and resource matrix $Y_{N\times k}$
1 Construct the loss function *Loss*
2 Randomly initialize user model X and resource model Y
3*while* (*Loss* does not converge)
4 *for i from 1 to M*
5 *for j from 1 to N*
6 *if* ($p_{i,j}$is recorded in the training set)
7 $e = X_iY_j - p_{i,j}$ / ∗Predictive value of model- Corresponding value of training set∗/
8 *end if*
9$x_{i,j} = x_{i,j}$-2 ∗ *alpha*(e ∗ $y_{i,j}$+*lambda* ∗ $x_{i,j}$)/ ∗ *alpha represents the learning rate* ∗ /
10$y_{i,j} = y_{i,j}$-2 ∗ *alpha*(e ∗ $x_{i,j}$+*lambda* ∗ $y_{i,j}$) / ∗ *delta represents the regular terms* ∗ /
11*end for*
12*end for*
13 Output X_i and $Y_j$as user matrix $X_{M\times k}$ and resource matrix $Y_{N\times k}$

Algorithm 2: Workflow of computing user and resource models by alternating least squares method.

Input: Set S of all probability prediction values of behavior space
Output: **A**bnormal behavior discrimination threshold T
1 Obtain the behavior probability set S
2*maxScore =0*
3*for t = min(S) to max(S) by 0.0001*/∗t represents the candidate threshold, traverseall the probabilities between *min(S)* and *max(S)*, with an increased step size of 0.001 (0.1%)∗/
4S_1={s|s∈S,s≤t} S_1S_2={s|s∈S,s>t}
5 Obtain the mean values of m_1m$_1$and m_2m$_2$ in S_1 and S_2
6 Obtain ratios p_1 andp_2 ofS_1 andS_2 in S
7 Obtain the between-cluster variance s^2 of S_1 andS_2 with Formula (6)
8*if (maxScore*< s^2)
9 $T = t$
10 *end if*
11*end for*
12 Output the final threshold T

Algorithm 3: The workflow of calculating abnormal behavior discrimination threshold T based on OTSU.

element in the ith row and the jth column of matrix P can be obtained by Formula (2). In the formula, $\widehat{p}_{i,j}$ represents the estimated value of the element $p_{i,j}$ in the ith row and the j th column of the frequency matrix P by user model X and resource model Y.

$$p_{i,j} \approx \widehat{p}_{i,j} = X_i Y_j^T. \tag{2}$$

For the decomposition of matrices, the CFomb method employs ALS (alternating least squares).

First, the loss function is constructed as formula (3), among which P_0 represents the set of numbers of rows and columns corresponding to the recorded users and resources. λ represents the coefficient of regular terms for preventing overfitting.

$$\text{Loss} = \sum_{(i,j)\in P_0} \left(p_{i,j} - X_i Y_j^T\right) + \lambda \sum_{i=1}^{M} \|X_i\|^2 + \lambda \sum_{j=1}^{N} \left\|Y_j\right\|^2. \tag{3}$$

Formulas (4) and (5) represent the gradient descent iteration formulas for this loss function.

$$X_i = X_i - 2\alpha \left[\sum_{j=1}^{N} \left(X_i Y_j^T - p_{i,j}\right) Y_j + \lambda X_i\right], \tag{4}$$

$$Y_i = Y_i - 2\alpha \left[\sum_{i=1}^{M} \left(X_i Y_j^T - p_{i,j}\right) X_i + \lambda Y_j\right]. \tag{5}$$

The algorithm that computes user model X and resource model Y by ALS is shown in Algorithm 2.

3.3.3. Abnormal Behavior Recognition. The abnormal behavior recognition by the CFomb method is divided into two steps: the first step is to quickly predict the probability of real-time behavior, and the second step is to discriminate whether the current behavior is abnormal based on the abnormal threshold.

The user model X and resource model Y can be obtained by the matrix decomposition method provided in 3.3.2. For

Input: User model$X$, resource model Y, abnormal behavior discrimination threshold T, and the extracted user behavior informationi and j (User i accesses Resource j)
Output: 1 /0; 1 indicates the normal behavior, and 0 the abnormal behavior
1 Extract vector X_i in the *ith* column of user model and vector$Y_j$in the*jth* column of resource model
2*p* = $X_i(Y_j)T$ /∗ p represents the prediction probability, which is obtained by computing the inner product of vectors $X_i$andY_j ∗ /
3*if*(p>T)
4 Output 1
5*else*
6 Output 0
7*end if*

Algorithm 4: The workflow of abnormal behavior recognition algorithm.

Input: The extracted user behavior information*i*and j (user i accesses resource j),the audit *tag*,and the BUFFER_SIZE
Output: update the training set
1 Obtain the behavior information *i, j*
2 Set the default integration ratio as *alpha*=1
3*if (tag=1)* / ∗ *audit feedback enters the buffer* ∗ /
4 *alpha* = BUFFER_SIZE
5*end if*
6*action*[*i*][*j*] += *alpha* /∗statistics on behaviors in the buffer∗/
7*user*[i] += *alpha*/∗update the amount of user behavior∗/
8*if* (*user*[*i*]>BUFFER_SIZE)
9 Employ Formula (7) to update the information corresponding to user *i* in the training set
10 *user*(*i*)=0
11 *action*[*i*][*j*]=0
12*end if*

Algorithm 5: The workflow of O&M behavior buffer.

the behavior that random user i accesses random resource j, the eigenvector of user i can be obtained by extracting the ith column of X, and that of resource j can be obtained by extracting the jth column of Y. The inner product of these two eigenvectors represents the probability prediction of such behavior.

The recognition of abnormal behavior can be completed by determining the abnormal behavior discrimination threshold after the behavior prediction probability is obtained. The CFomb method has introduced the OTSU algorithm to adaptively determine the abnormal behavior discrimination threshold.

First, the statistics on the prediction probability of all behaviors can be made in behavior space via user model and resource model. The prediction probability of abnormal behavior will be concentrated in the lower probability interval, while that of normal behavior will be concentrated in the higher probability interval. If an existing probability threshold T divides the behavior into two maximum sets of between-cluster variance σ^2 in the behavior space, the threshold T can be taken as the probability threshold for abnormal behavior discrimination.

Formula (6) provides the computation of between-cluster variance in operation and maintenance behavior in the CFomb method.

$$\sigma^2 = p_1 \cdot p_2 \cdot (m_1 - m_2)^2. \quad (6)$$

In the formula, p_1 represents the proportion of behaviors with a prediction probability not more than the threshold T in the behavior space, p_2 represents the proportion of behaviors with a prediction probability above the threshold T in the behavior space, m_1 represents the mean probability of behaviors with a prediction probability not more than threshold T, and m_2 represents the mean probability of behaviors with a prediction probability above the threshold T.

The specific algorithm steps for computing threshold T based on OTSU algorithm are shown in Algorithm 3.

Furthermore, the real-time operation and maintenance behavior can be realized in accordance with operation and maintenance user model X, power resource model Y, and abnormal behavior discrimination threshold T, with the specific algorithm shown in Algorithm 4.

3.3.4. Construction of Behavior Buffer. The CFomb method will continuously collect the behavior information of users during operation. In order to implement the dynamic adjustments to the user model and resource model, the CFomb method designs a behavior buffer in the behavior analysis module for receiving the behavior information of operation and maintenance users and feeding back new operation and maintenance user behaviors to the training model, thereby adjusting the user model and resource model.

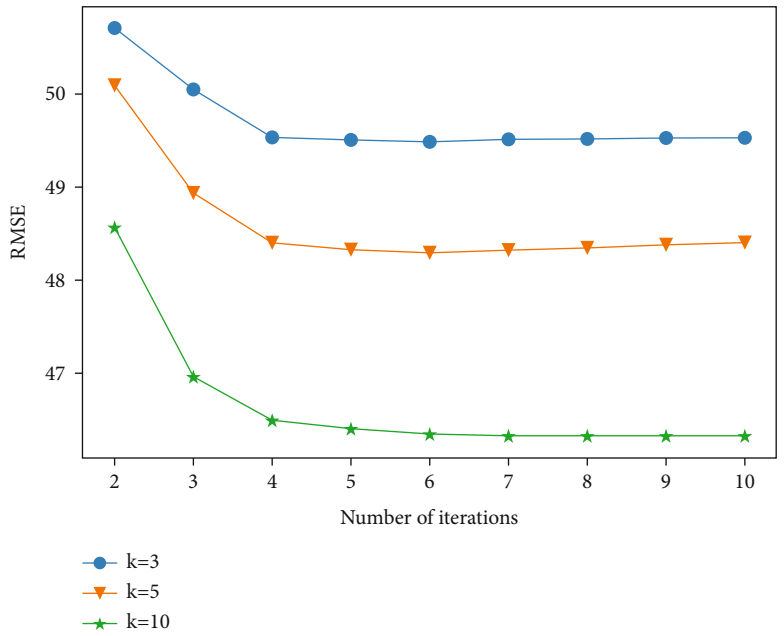

Figure 3: RMSE curve of the model with different values of k.

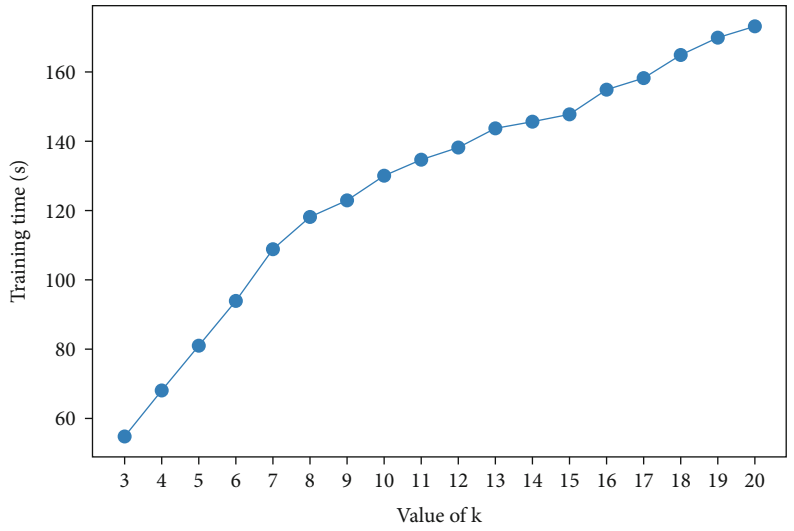

Figure 4: Training time curve under different values of k.

The behavior information to be written in the behavior buffer will be stored in the form of triple $(i, j, \text{action}_{i,j})$, indicating that the number of times stored in the buffer for accessing resource j by an operation and maintenance user i is $\text{action}_{i,j}$. Moreover, the behavior buffer maintains a counting sequence times_i, in which $1 \le i \le M$, indicating that times_i behavior information of the ith operation and maintenance user is recorded in the buffer. Formula (7) represents the computational formula in which the access behavior of an operation and maintenance user i to resource j in the behavior buffer is integrated into the training set.

$$p_{i,j} = (1 - w)p_{i,j} + w \cdot \frac{\text{buffer}_{i,j}}{\text{times}_i}, \tag{7}$$

where w $(0 < w < 1)$ represents the writing weight of user buffer. The higher the writing weight, the higher the change rate of training set. The specific values will not be specified in this paper. Algorithm 5 provides the workflow of operation and maintenance behavior buffer.

4. Experimental Evaluation

4.1. Data Preparation. In order to verify the recognition efficiency of the CFomb method on the users' abnormal behaviors of access to the resources on the operation and maintenance platform of power monitoring system, the behavior of resource access by operation and maintenance users is simulated under the background of operation and maintenance [14–15], and the dataset of resource access by

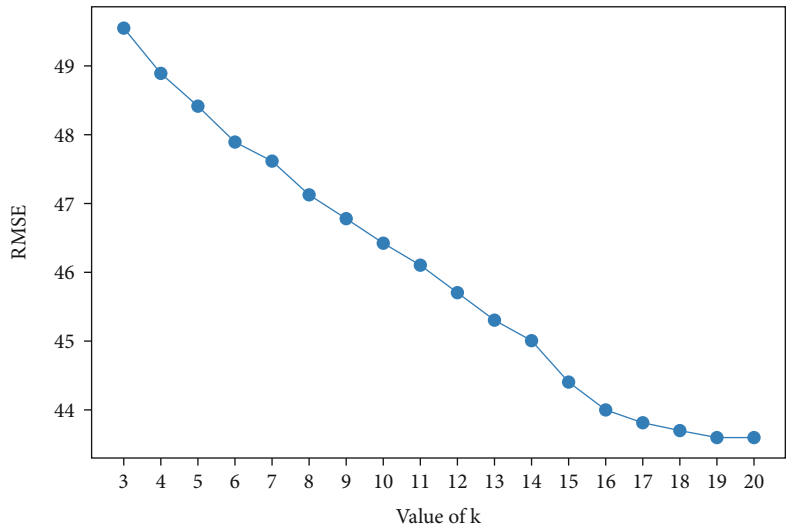

Figure 5: RMSE under different values of k.

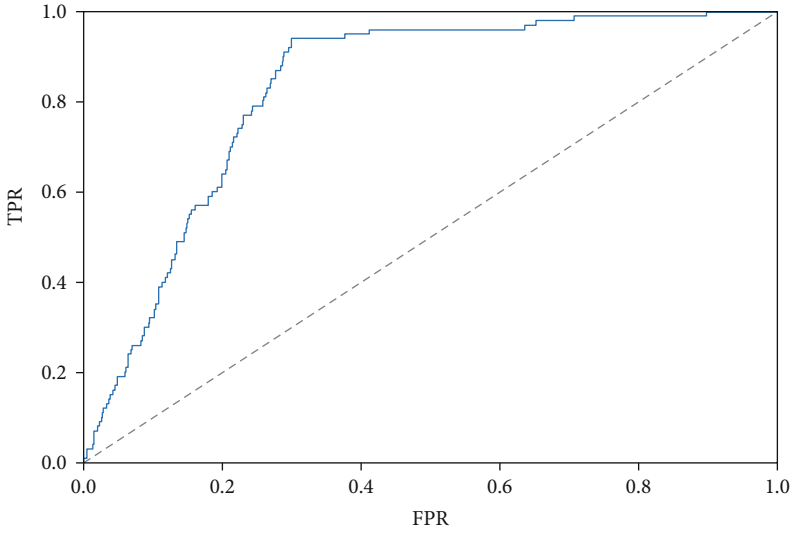

Figure 6: ROC curve of the test case.

Table 1: Abnormal behavior identification confusion matrix.

	Recognition		Total
	Abnormal	Normal	
Actual			
Abnormal	94 (TP)	6 (FN)	100
Normal	1,513 (FP)	3,487 (TN)	5,000
Total	1,607	3,493	5,100

operation and maintenance users is generated, based on which the recognition capability of the CFomb method on abnormal behaviors is verified. The dataset was obtained from the internal network data of China Southern Power Grid Corporation, which contains a total of 107,670 records on 1,170 users' access to 7,455 resources. Moreover, 1,759 access records are generated by simulating the malicious user access behavior to verify the recognition effect.

4.2. Influence of Feature Dimension on Model Training. In the implementation of the algorithm of user behavior matrix decomposition, it is necessary to first determine the values of feature dimension k in user model and resource model. The value of k represents the dimension describing the features of users and resources when predicting the probability of resource access by users, which affects the prediction accuracy of the training model, model size, and training time. In order to verify the influence of k on the prediction accuracy of training model, the predictive RMSE of model for 2 to 10 rounds of training is recorded, respectively, when the values of k are 3, 5, and 10, and the RMSE decline curve was drawn. The experimental results are shown in Figure 3. It is obvious that the larger the value of k in the training model is, the smaller the model error is; and all the training models converge to the steady state after the fifth round.

In order to further determine the influence of the values of k on model training, the number of training times is set as 5, and the training time and RMSE with the value of k ranging from 3 to 20 are recorded, as shown in Figures 4 and 5, respectively. It can be seen that when the value of k ranges from 3 to 20, the increase in training time and the decline in RMSE are linear and variable.

Since it is necessary to continuously update the training set to implement iterative training in the context of operation and maintenance and quickly output model on the premise of ensuring the accuracy of the training model. The value of k in this experiment is set as 10.

4.3. Threshold Sensitivity Analysis Based on the Test Set. The user model and resource model are obtained via the training on the behavior dataset in 4.2. Through the models, the probability prediction on all behaviors in behavior space can be implemented. In order to realize the abnormal behavior recognition, the abnormal behavior discrimination threshold shall be determined. In order to verify whether the probability prediction value of user behaviors obtained from the training model can distinguish the normal behavior from the malicious behavior, 100 malicious access behaviors and 5,000 normal behaviors are extracted to form a test set to test the training model.

In the experiment, the probabilities of all behaviors in the test set in the user model and the resource model are predicted, and the sensitivity of the data in dataset on the prediction probability threshold is verified using ROC (receiver operating characteristic) curve in combination with the tags indicating whether the behaviors are abnormal. The ROC curve can depict the relationship between false positive rate (FPR) and true positive rate (TPR) when test data classify the abnormal behaviors at different thresholds. TPR is the proportion of correctly identified positive data to the total positive data, i.e., the recall rate, while FPR indicates the percentage of negative data predicted to be positive when the actual value is negative. The specific formulas (8) and (9) of FPR and TPR are as follows:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{8}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}. \tag{9}$$

In the formula, TP, FP, TN, and FN represent true positive, false positive, true negative, and false negative, respectively. Since it is generally assumed that behaviors equal to or lower than the probability threshold are abnormal, the behaviors lower than the threshold are determined as positive in this experiment [16, 17].

The ROC curve of test set data in this experiment is shown in Figure 6. The threshold sensitivity of data in test set can be represented by AUC (area under the curve) [18, 19], and the AUC of the test data is 82.80%. The ROC curve indicates that the probability threshold has a certain capability to classify normal and abnormal behaviors in the data of test set [20–22]. However, a higher accuracy rate of abnormal behavior recognition will be accompanied by FPR increase to some extent [23–25].

4.4. Recognition Effect of Abnormal Operation and Maintenance Behavior. In practical application, it is necessary to adaptively generate the abnormal behavior discrimination threshold in accordance with the probability distribution in behavior space; therefore, the CFomb applies the OTSU algorithm to the computation of the abnormal behavior discrimination threshold. In the experiment, the probabilities of all behaviors in behavior space are calculated, and the threshold of 0.0927 is obtained with the OSTU algorithm.

Afterward, the effect of abnormal behavior recognition proposed in this paper is further tested at the threshold of 0.0927. There are a total of 5,100 behavior records in the experiment. The confusion table (as shown in Table 1) is obtained based on the statistics on recognition results, with the corresponding TPR and FPR of 94% and 30.26%, respectively. The data indicates that 94% of malicious behaviors are recognized at the threshold of 0.0927, with 30.26% of normal behaviors determined as abnormal behaviors. Moreover, of all the behaviors, 31.51% are determined as normal behaviors, indicating that CFomb can exclude 68.49% of behavior records in manual audits, which significantly reduces the workload of security audit.

5. Conclusion and Future Work

Based on the matrix decomposition collaborative filtering method, this paper proposes the CFomb-based abnormal behavior collaborative filtering method under the background of internal attacks against the operation and maintenance platform of the power monitoring system. This method obtains a user model and resource model via training in global data access by users, combined with collaborative filtering idea and OTSU binarization method to realize the probability prediction of real-time access behavior of O&M users and adaptive selection of probability threshold of abnormal O&M behavior to support more efficient and safe development of power monitoring system O&M. The experiment indicates that the recognition effect is obvious, which can effectively prevent internal threats to the operation and maintenance platform of the power monitoring system and significantly reduce the workload of manual audits. The method proposed in this paper overcomes the inefficiency of traditional methods and provides new ideas for the operation and maintenance mode of power monitoring systems. Due to the limitation of the data in the training set, the probability prediction model has certain errors. Based on this paper, we will lower the misjudgment rate in the future in combination with other behavior features of operation and maintenance users as well as the feedback iteration of operation and maintenance personnel on the training model, to further study more effective methods for recognizing abnormal behaviors.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no competing interests.

References

[1] M. S. LiQ, S. Zhang, M. Wu, J. Zhang, and M. Taleby, "Safety risk monitoring of cyber-physical power systems based on ensemble learning algorithm," *Access*, vol. 7, pp. 24788–24805, 2019.

[2] H. Shi, "Design and application of operation and maintenance security management and control system based on bastion machine technology," *China Management Information*, vol. 19, no. 24, pp. 44-45, 2016.

[3] W. Zheng, H. Chen, and P. Wu, "Demand and application scenario design of online operation and maintenance audit platform for power monitoring system," *Network Security Technology and Application*, vol. 2020, no. 11, pp. 134-135, 2020.

[4] H. Wu, "Analysis of preventive measures for network information security of electric power enterprises information system," *Engineering*, vol. 2018, no. 12, p. 73, 2018.

[5] G. Yang, J. Ma, A. Yu, and D. Meng, "Research on insider threat detection," *Journal of Information Security*, vol. 1, no. 3, pp. 21–36, 2016.

[6] S. Yuan and X. Wu, "Deep learning for insider threat detection: review, challenges and opportunities," *Computers & Security*, vol. 104, p. 102221, 2021.

[7] H. Peng, *Research on Insider Threat Detection Method Based on User Behavior*, Beijing Jiaotong University, Beijing, 2019.

[8] F. Liu, Y. Wen, D. Zhang, X. Jiang, and D. Meng, "Log 2vec: a heterogeneous graph embedding based approach for detecting cyber threats within enterprise," in *the 2019 ACM SIGSAC Conference. ACM*, 2019.

[9] G. Zhaojun and G. Jingxuan, "An insider threat detection method based on character abnormal behavior mining," *Computer Engineering and Design*, vol. 41, no. 10, pp. 2740–2746, 2020.

[10] T. Rashid, I. Agrafiotis, and J. R. C. Nurse, "A new take on detecting insider threats: exploring the use of hidden Markov models," in *Proceedings of the 8th ACM CCS International workshop on managing insider security threats*, pp. 47–56, 2016.

[11] J. Happa, "Insider-threat detection using Gaussian mixture models and sensitivity profiles," *Computers & Security*, vol. 77, pp. 838–859, 2018.

[12] Z. Khan, S. Zubair, K. Imran, R. Ahmad, S. A. Butt, and N. I. Chaudhary, "A new users rating-trend based collaborative denoising auto-encoder for top-N recommender systems," *IEEE Access*, vol. 7, pp. 141287–141310, 2019.

[13] W. Wang, J. Chen, J. Wang, J. Chen, J. Liu, and Z. Gong, "Trust-enhanced collaborative filtering for personalized point of interests recommendation," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 6124–6132, 2020.

[14] C. Zhang and C. Wang, "Probabilistic matrix factorization recommendation of self-attention mechanism convolutional neural networks with item auxiliary information," *IEEE Access*, vol. 8, pp. 208311–208321, 2020.

[15] Y. Zhang, Y. Wang, and S. Wang, "Improvement of collaborative filtering recommendation algorithm based on intuitionistic fuzzy reasoning under missing data," *IEEE Access*, vol. 8, pp. 51324–51332, 2020.

[16] J. Chen, J. Han, X. Meng, Y. Li, and H. Li, "Graph convolutional network combined with semantic feature guidance for deep clustering," *Tsinghua Science and Technology*, vol. 27, no. 5, pp. 855–868, 2022.

[17] M. Heydarian, T. Doyle, and R. Samavi, "MLCM: multi-label confusion matrix," *Access*, vol. 10, pp. 19083–19095, 2022.

[18] S. Zhang and M. Abdel-Aty, "Real-time pedestrian conflict prediction model at the signal cycle level using machine learning models," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 176–186, 2022.

[19] S. Lin, P. Rouse, Y. Wang, and F. Zhang, "A statistical model to detect DRG outliers," *IEEE Access*, vol. 10, pp. 28717–28724, 2022.

[20] Z. Huang and D. Chen, "A breast cancer diagnosis method based on VIM feature selection and hierarchical clustering random forest algorithm," *IEEE Access*, vol. 10, pp. 3284–3293, 2022.

[21] M. Wu, L. Tan, and N. Xiong, "A structure fidelity approach for big data collection in wireless sensor networks," *Sensors*, vol. 15, no. 1, pp. 248–273, 2015.

[22] S. Huang, A. Liu, S. Zhang, T. Wang, and N. N. Xiong, "BD-VTE: a novel baseline data based verifiable trust evaluation scheme for smart network systems," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 2087–2105, 2021.

[23] H. Li, J. Liu, K. Wu, Z. Yang, R. Liu, and N. Xiong, "Spatio-temporal vessel trajectory clustering based on data mapping and density," *IEEE Access*, vol. 6, pp. 58939–58954, 2018.

[24] K. Gao, F. Han, P. Dong, N. Xiong, and R. du, "Connected vehicle as a mobile sensor for real time queue length at signalized intersections," *Sensors*, vol. 19, no. 9, p. 2059, 2019.

[25] P. Yang, N. Xiong, and J. Ren, "Data security and privacy protection for cloud storage: a survey," *IEEE Access*, vol. 8, pp. 131723–131740, 2020.