Comprehensive Control Method of Network Teaching Data Scheduling Based on Fuzzy Mathematics

Mei-li Dai(✉) and Li An

Rizhao Polytechnic, Rizhao 276800, China

Abstract. Aiming at the problem of dividing and scheduling network teaching resources in the network environment, a more complete fuzzy mathematics theory and an example are used to introduce a method of dividing and scheduling teaching resources in detail. The main methods used during this period are: the introduction of data standardization methods to eliminate the influence of dimensions; a method of comprehensively considering various indicators and then processing them separately, in order to avoid the "barrel effect"; For some articles only relying on the error of using fuzzy similarity matrix to divide teaching resources has been corrected, and the final division should use fuzzy equivalent matrix.

Keywords: Teaching resources · Fuzzy clustering analysis · Barrel effect · Data standardization · Transitive closure · Fuzzy equivalent matrix

1 Introduction

Teaching data scheduling refers to making full use of the network teaching information resources in different time and space, using computer technology to automatically analyze, synthesize, distribute and use the network teaching information obtained according to the time sequence under certain criteria, so as to obtain the consistent description and interpretation of the tested object, so as to complete the required decision-making and estimation tasks, and make the teaching system obtain its components More superior performance [1]. The existing network teaching data scheduling methods include the least square method based on nonlinear optimization class and the method based on Kalman filter. However, the measurement of U1 can not be based on the measurement of various factors such as Kalman filter equation, which is based on the observation of various factors. However, the least square method based on nonlinear optimization class has great difficulties in the implementation of the algorithm due to the uncertainty of the available measurement elements and the need for iterative solution, which can not meet the requirements of real-time scheduling calculation [2, 3]. How to take a simple and effective method to schedule network teaching information is the key to improve the accuracy and efficiency of network data processing.

As an eye-catching next-generation network system, the grid has its own distinctive features, and thus has created its powerful functions. Simply put, the grid is the use of

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2021
Published by Springer Nature Switzerland AG 2021. All Rights Reserved
W. Fu et al. (Eds.): eLEOT 2021, LNICST 389, pp. 205–217, 2021.
https://doi.org/10.1007/978-3-030-84383-0_18

all the massive and available connections on the Internet (i.e. Idle) resources (mainly including computing resources and data storage resources) are integrated into a super virtual computer. The user does not need to understand or even understand its background operation (that is, transparent to the user), just through a relatively simple user interface, This super virtual computer can be used to quickly and efficiently complete tasks such as massive data operations, massive data storage management, and massive information search [4].

Aiming at the above problems, this paper proposes a network teaching data scheduling method which combines fuzzy mathematics and non negative eigenvector theory. First of all, the method of data standardization is used to eliminate the influence of data dimension. Secondly, considering the influence of each index, fuzzy mathematics method is used to schedule and control the network teaching data. Finally, the effectiveness of the method is verified by practical application, and a better scheduling effect is obtained.

2 Network Teaching Resources

In a broad sense, network teaching resources are all kinds of teaching resources that are connected to the Internet, and all kinds of educational software resources installed on it. Users can get corresponding network teaching resources and resources according to their different needs and user rights service.

Through the above analysis, it can be seen that the scope of online teaching resources is very large, and the specific performance of each resource is not the same. In order to achieve effective management and efficient use to meet the needs of different students and users, a very critical step is to Division and scheduling of teaching resources. In view of the characteristics of the above teaching network resources, it can be found that applying the fuzzy clustering analysis method in fuzzy mathematics theory can solve the above problems well.

3 Fuzzy Cluster Analysis

Fuzzy equivalent matrix is a classical fuzzy clustering method, which combines the standard residual error with the measurement difference of adjacent time to calculate the fuzzy eigenvalue fuzzy clustering, and find the best calculation threshold. According to the data results of individual data, the clustering of target data is completed. Fuzzy equivalent matrix clustering can effectively avoid residual pollution and residual inundation, and can flexibly select clustering results to improve the effectiveness of clustering.

In ordinary cluster analysis, given objects and their characteristics can be classified according to the inherent similarity of the data. The mathematical method of classifying the researched objects according to certain standards is called cluster analysis. This is a classification method of multivariate statistics “things gather together”. However, in the network teaching data, there are many types of data that are not clearly divided, the boundaries are fuzzy, and the relationship between them is more fuzzy. Therefore, fuzzy mathematics methods should be used to classify network teaching data. The cluster analysis using fuzzy mathematics method is called fuzzy cluster analysis [5, 6].

The object of this paper - network teaching resources, which can not be clearly divided between each other, the boundary is fuzzy, so it is feasible to use fuzzy mathematics to analyze it.

The following steps of fuzzy analysis are introduced

The first step is to standardize the data. ① Write out the data matrix; ② Standardize the data. Network teaching resources have many characteristics, each of which has a different dimension. In order to compare the data of network teaching resources with different dimensions, it is usually necessary to properly transform the original data to obtain a new sample. Standardization matrix. In this paper, the translation and standard deviation transformation method is used, and the new standard sample network teaching data obtained in the first step is used to eliminate the influence of dimensions, and a fuzzy matrix representing the fuzzy relationship between each network teaching data is established.

The second step is calibration (building a fuzzy similarity matrix). In order to classify network teaching data quantitatively, some quantitative indicators for classification must be determined, and some quantitative indicators that can indicate the degree of similarity between samples (or variables) are introduced, which are called cluster statistics.

The determined cluster statistics $r_{ij} = R(x_i, x_j)$ and $\underline{R}(x, y)$ are called membership function or membership degree, and their values represent the correlation degree of binary relationship (x, y) with fuzzy relation R, where R represents fuzzy similar relation, and the calculation methods mainly include similarity coefficient method and distance method. In this paper, Hamming distance method is used. A fuzzy matrix R with r_{ij} as its element can be obtained.

The third step is the clustering of online teaching resources. There are many clustering methods. Here we mainly introduce the clustering method based on fuzzy equivalence matrix—transitive closure method. When classifying network teaching resources according to fuzzy relations, the fuzzy relations must be fuzzy equivalence relations, which are reflexive, symmetrical and transitive [7].

According to the calibration, the fuzzy matrix is only a fuzzy similar matrix, and the fuzzy similar matrix only has reflexivity and symmetry, but it does not necessarily have transitivity, that is, R is not necessarily a fuzzy equivalent matrix (fuzzy equivalent matrix has reflexivity, symmetry and transitivity). In order to classify network teaching resources, R needs to be transformed into fuzzy equivalent matrix R''.

By finding the transitive closure $t(R)$, the fuzzy similarity matrix can be transformed into a fuzzy equivalent matrix, which is transitive, while retaining reflexivity and symmetry. Here is a practical simple method-the square method, seeking transitive closure $t(R)$.

Starting from the fuzzy similar matrix, the square is calculated in turn

$$R \to R^2 \to R^4 \to \cdots \to R^{2^4} \to \cdots, (i = 1, 2, 3, \cdots) \quad (1)$$

When $R^k \cdot R^k = R^{2k}$ appears for the first time (indicating that R^k is transitive!), R^k is the required transitive closure $t(R)$, which is the required fuzzy equivalent matrix R^*, namely $t(R) = R^*$. Then, take a fixed value λ of [0, 1] and specify:

For any two elements x_i and y_i in the universe, if:

$$r_{ij} \geq \lambda \tag{2}$$

Then x_i and y_i belong to the same class; otherwise, they do not belong to the same class [8].

4 Network Teaching Data Fusion Method Based on Fuzzy Closeness

Using the real-time network teaching data processing process of positioning smoothing and filtering, the three-dimensional processing results of group n are obtained:

$$e_i = (x_i, y_i, z_i), i = 1 \ldots n \tag{3}$$

According to the decision level fusion structure model, the distance between any two positioning results is calculated. The relative distance between the i positioning result and the j positioning result is as follows:

$$d_{ij} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2 + \left(z_i - z_j\right)^2} \tag{4}$$

The authenticity of the i th positioning result can be determined by the relative distance d_{ij} between the i-th positioning result and the positioning result of the j-th measuring device: the smaller the d_{ij}, the higher the authenticity of the i-th positioning result. Conversely, the larger the d_{ij}, the lower the authenticity of the i-th positioning result, and the fuzzy closeness function $a_{ij}(k)$ between the i-th positioning result and the j-th positioning result at different spatial positions at time k can be defined.

Network teaching data fusion is shown in Fig. 1.

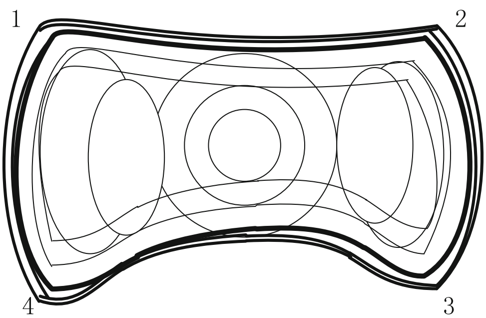

Fig. 1. Network teaching data fusion process

The definition of fuzzy closeness function should satisfy:

(1) When the relative distance between the two positioning results is larger, their fuzzy closeness is smaller, and the relative distance between the two positioning results is smaller, the mutual closeness between the data is greater, that is, $a_{ij}(k)$ should be proportional to the relative distance Inverse relationship

(2) If the relative distance of the data to itself is zero, the fuzzy closeness of the data to itself is 1;
(3) $a_{ij}(k) \in (0, 1)$, enabling data processing to take advantage of the membership function in fuzzy set theory, avoiding the absoluteization of fuzzy closeness between data. Therefore:

$$a_{ij} = \frac{1}{1 + d_{ij}} = \frac{1}{1 + \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}} \tag{5}$$

It can be seen that the definition form of formula (5) satisfies the properties of fuzzy closeness function, and this definition form of satisfying fuzzy closeness function $a_{ij}(k)$ is more in line with the reality of practical problems, and is convenient for specific implementation, so that the result of network teaching resource integration is more accurate and stable [9, 10].

Through the above method, we can get the fuzzy closeness degree of arbitrary positioning results at time k, and then obtain a fuzzy closeness matrix $A(k)$:

$$A(k) = \begin{bmatrix} a_{11}(k)\ a_{12}(k) \ldots a_{1n}(k) \\ a_{21}(k)\ a_{22}(k) \ldots a_{2n}(k) \\ \vdots \\ a_{n1}(k)\ a_{n2}(k) \ldots a_{nn}(k) \end{bmatrix} \tag{6}$$

According to the matrix $A(k)$, the comprehensive fuzzy closeness $u_{i(k)}$ of a certain positioning result $(x_i y_i z_i)$ and other positioning results is obtained, and a set of non-negative numbers is required from the theory of probability source merger:

$$b_1(k), b_2(k), \ldots b_n(k) \tag{7}$$

Make:

$$\begin{gathered} u_i(k) = b_1(k)a_{i1}(k) + b_2(k)a_{i2}(k) + \ldots + b_n(k)a_{in}(k) \\ i = 1, 2 \ldots, n \end{gathered} \tag{8}$$

Make:

$$\begin{gathered} U(k) = [u_1(k), u_2(k), \ldots, u_n(k)]^r \\ B(k) = [b_1(k), b_2(k), \ldots b_n(k)]^T \end{gathered} \tag{9}$$

Then formula (9) can be written as the following matrix form:

$$U(k) = A(k)B(k) \tag{10}$$

Because of $a_{ij} \geq 0$, $A(k)$ is a non-negative symmetric matrix. According to the properties of the non-negative symmetric matrix, it can be seen that $A(k)$ has a maximum fuzzy eigenvalue $\lambda(\lambda > 0)$, and only the eigenvector corresponding to the eigenvalue is positive, and $A(k) = \lambda B(k)$ is used to obtain the eigenvector $B(k)$, which is substituted into Eq. (10) as the fuzzy positioning result Proximity [11–13]. Expanding matrix (10), the i fuzzy closeness is:

$$u_i(k) = b_2(k)a_{i2}(k) + \ldots b_n(k)a_{in}(k) \tag{11}$$

Where $b_i(k)$ is the i th element of the eigenvector $B(k)$.

In the actual network teaching data fusion process, different positioning results need to be given different weights. The teaching system with better stability and higher reliability has greater weight. According to the previous discussion, the reliability and stability of the equipment with large fuzzy proximity degree are higher, and the weight should be greater, otherwise, the weight should be smaller. Therefore, fuzzy closeness can be used to represent the weight of the sensor. The clustering data is shown in Fig. 2:

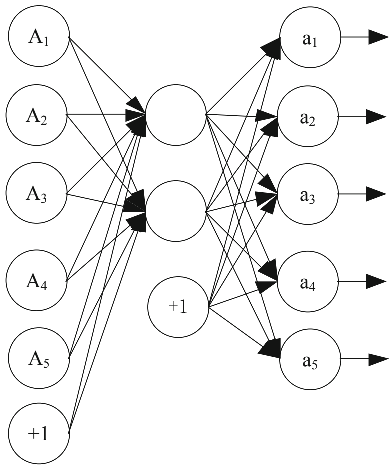

Fig. 2. Clustering data

Because in the rectangular coordinate system, the abnormal data sample and the normal data sample are inconsistent, so this paper completes the network teaching data processing work by establishing the rectangular coordinate system. Randomly select n data as the basic sample, draw the ROC curve through MATLAB software, judge the accurate value of the data, and then repeat this operation. Other samples are sampled to calculate the offset between the standard data and the measured data, the overlapped network teaching data are stored, and the different network teaching data are re integrated until the ROC curve is a coincidence line. The established rectangular coordinate system is shown in Fig. 3.

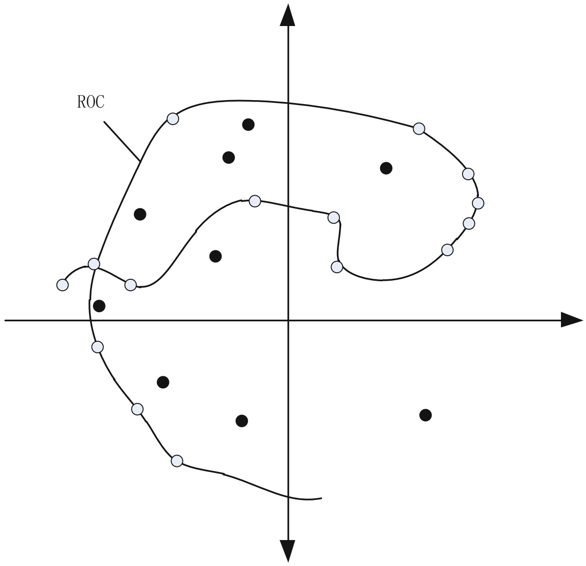

Fig. 3. Data processing rectangular coordinate system

The vertical axis direction of ROC curve represents the detection probability, and the horizontal direction represents the false alarm rate. According to the change of the curve, the higher the threshold value, the higher the accuracy of the system, and the stronger the self identification ability. If the ROC is not a smooth curve, it is necessary to divide the ROC curve into several sections to form a number of small trapezoids, and calculate the accurate value of each trapezoid area data. Through the addition of several trapezoid areas, the number of data is subtracted from the number of normal data, and the false alarm rate of abnormal data is obtained. The best working point is found in the ROC curve, and the positive likelihood ratio and Youden index are used to distinguish the error detection rate and false alarm rate.

After the discrete attribute data is obtained, the original network teaching data is mapped to the three-dimensional space by constructing a new linear function, and then divided in the space. The matrix is used to find the heterogeneous data, and the optimal solution is obtained. The cloud detection method mainly depends on the simulation database. If there is no normal data in the simulation database, the attack data cannot be detected. If the real-time coding of abnormal data and the normal data exceed a certain threshold, the normal data will also be attacked. The whole data is analyzed and processed to show the non-linear relationship between normal data and abnormal data. The data is compared by random sampling, and HMM model is established. Using the detection function of HMM model, different data can be identified through cloud test, so as to realize information distribution processing and increase the adaptability of processing process.

Let the weight of the measured value of i be $w(k)$. According to the principle of information sharing, the sum of the information estimated by the optimal fusion can be equivalently decomposed into the sum of the information of several measurement data, that is, one information can be shared by several subsystems.

$$\sum_{i=1}^{n} w_i = 1(0 \leq w_i \leq 1) \tag{12}$$

Normalize the fuzzy closeness of n measuring equipment to get their respective relative weights:

$$w_i(k) = \frac{u_i(k)}{\sum_{i=1}^{n} u_i(k)} \tag{13}$$

Therefore, at a certain time k, the fusion positioning result of the target is as follows:

$$e(k) = \sum_{i=1}^{n} w_i(k)e_i(k) \tag{14}$$

5 Using Fuzzy Cluster Analysis Method to Dispatch and Control Network Teaching Resources

Different from the previous method of considering a single index one by one, the example presented here uses two comprehensive indicators of machine computing performance and communication performance to divide machines at the beginning, which can avoid the "barrel effect" in advance and meet the real-time application requirements most concerned by ordinary users. The so-called "barrel effect" refers to the amount of water contained in a barrel, which ultimately depends on the shortest of the boards used to make the bucket. In this case, it refers to two "boards" of "computing performance" and "communication performance".

Step 1: Data standardization.

(1) Data matrix

Suppose domain $U = \{x_1, x_2, \cdots, x_{10}\}$ represents a network teaching system with 10 nodes. Each node uses two indicators of computing performance and communication performance to represent its characteristics, namely $x_i = (x_{i1}, x_{i2})$. The original data is shown in Table 1:

Table 1. Raw data

Node	Index	
	Computational performance	Communication performance
x_1	0.6	50
x_2	1.0	40
x_3	2.0	60
x_4	4.0	55
x_5	2.5	30
x_6	3.0	100
x_7	54.0	20
x_8	45.0	100
x_9	20.5	50
x_{10}	15.0	120

Therefore, the transposition matrix of the original data matrix is obtained, and the clustering results of different network teaching data are shown in Fig. 4 below:

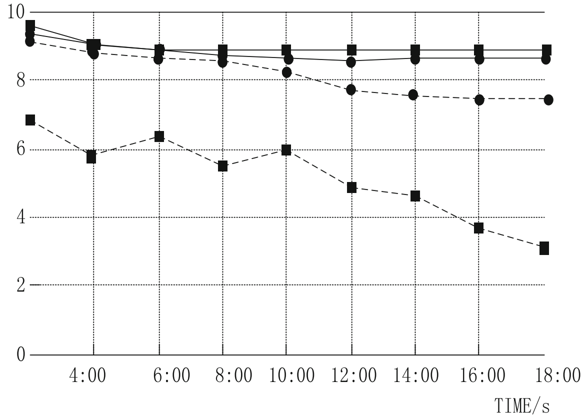

Fig. 4. Clustering results of different network teaching data

So far, the classification of network teaching resources has been basically completed. It can be seen that through $\lambda \in [0, 1]$ clustering from large to small, for different parameter λ, grid resources are clustered into different groups, and the nodes in each group can be regarded as nodes with similar comprehensive characteristics under the current parameter λ. Users can adjust the value of λ according to their needs. For larger application problems, by reducing the value of λ, the "logical groups with adjacent performance" can be merged into a large logical group. When λ is taken as 0, all nodes will form a logical group, and the scheduler will schedule the application decomposition tasks to all computing nodes as much as possible.

After successful grouping, according to the formula:

$$V = \frac{1}{M}\sum_{k=1}^{N} C_k \tag{15}$$

Calculate the cluster center value (or mean) V of the computing performance of each group and the cluster center value (or mean) V of the communication performance, where: M is the total number of nodes in the group; C_k is each node in the group Calculated performance value or communication performance value.

If the parameter λ is 0.911, the computational performance cluster center value of the first (i.e. the first class) $\{x_1, x_3, x_4, x_2, x_5\}$ of the group (class) is 0.911:

$$V_1 = \frac{1}{5}\sum_{i=1}^{5}(6.0 + 2.0 + 4.0 + 1.0 + 2.5) = 3.1 \tag{16}$$

Communication performance cluster center value:

$$V_2 = \frac{1}{5}\sum_{i=1}^{5}(50 + 60 + 55 + 40 + 30) = 47 \tag{17}$$

It is also worth noting that the maximum performance deviation ratio D can indicate the similar degree of performance of the nodes in this group

$$D = \frac{Max((Max_{k=0}^{n} C_h - V), (V - Min_{k=0} C_k))}{V} \tag{18}$$

Among them: M is the total number of nodes in the group. Obviously, the smaller the D, the closer the performance data of the nodes in the group.

So far, we can get a scheduling control algorithm that can network teaching resource data, and its steps are:

(1) Calculate the fuzzy equivalence matrix based on the comprehensive indicators of all computing nodes, that is, the closeness matrix;
(2) Determine the first-order λ matrix of the fuzzy equivalent matrix according to the value of λ and refer to the maximum performance deviation rate D;
(3) Obtain several logical groups of machines according to the λ order matrix;
(4) For problems of different scales and requirements, suitable or user-specified logical groups can be selected for scheduling according to the cluster center value.

6 Simulation Calculation

Assuming that the initial position of the network teaching data is (0, 0, 0), the scheduling movement is performed in the YOZ plane, sampling is performed at a period of 50 ms, and the data of 5 s is continuously simulated, and the fuzzy closeness formula is used to calculate the value of any two network teaching data. The fuzzy closeness degree is formed to form a fuzzy closeness degree matrix, the maximum eigenvalue of the matrix and its corresponding positive eigenvector are calculated, and then the comprehensive fuzzy closeness degree is obtained as the weight, and the weighted fusion calculation of network teaching data is performed.

Compare the actual scheduling results with the positioning results of the network teaching data scheduling method proposed in this paper. The graphs comparing the scheduling and positioning results of the two methods in the X direction, Y direction, and Z direction are respectively given, as shown in the figure below (Figs. 5, 6 and 7).

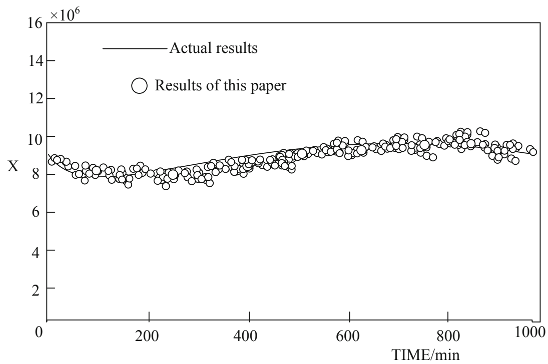

Fig. 5. T-X curve of dispatching positioning data

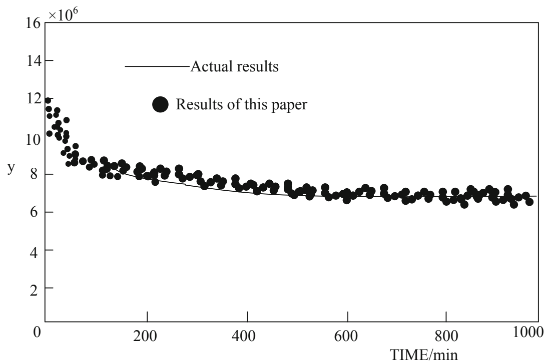

Fig. 6. T-Y curve of dispatch positioning data

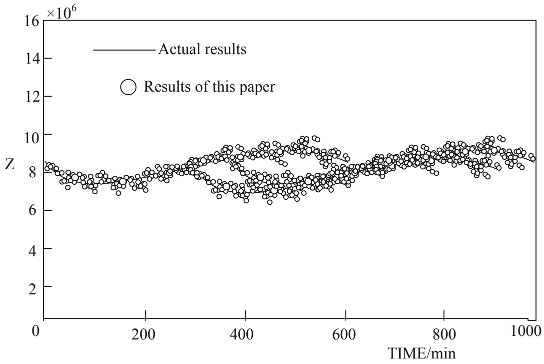

Fig. 7. T-Z curve of dispatching positioning data

It can be seen from the simulation results that, by comparing the results of intersection scheduling and fusion positioning in the X-axis, Y-axis, and Z-axis directions, the network teaching data scheduling and positioning results of the proposed method are very close to the actual results. It shows that the proposed method can effectively eliminate the adverse effect of the measured value on the fusion result and improve the processing accuracy.

7 Conclusion

According to the characteristics of network teaching resources, it is very suitable to use fuzzy clustering method in fuzzy mathematics to classify them. The early method of processing a single indicator and then comprehensively processing is more suitable for users who have special needs for grids. The method in this article is suitable for mass users who have general needs for network teaching resources and are more concerned about comprehensive performance.

References

1. Du, P.: Simulation of network multi-source data scheduling based on multi-goal mathematical planning. Comput. Simul. **37**(1), 343–346,447 (2020)
2. Hong, T., Gong, W.: The method of trust evaluation based on fuzzy mathematics. Tech. Autom. Appl. **39**(2), 41–44 (2020)
3. Yu, W., Jia, C., Di, S., et al.: Evaluation of seawater intrusion degree based on fuzzy mathematics and improved analytic hierarchy process. Mar. Sci. **42**(6), 21–31 (2018)
4. Liu, L., Zhang, Q., Gao, J., et al.: Data-based adaptive fuzzy comprehensive evaluation method well selection. Comput. Digit. Eng. **46**(10), 1980–1983 (2018)
5. Liu, G., Shi, X.: An improved fuzzy clustering fastest time pattern of cloud computing task scheduling scheme. Control Eng. China **25**(11), 2092–2096 (2018)
6. Qian, K., Qin, M., Song, Y., et al.: Multi-objective optimal scheduling based on fuzzy membership for electric bus charging station. Electr. Power Constr. **41**(2), 118–124 (2020)

7. Deng, C., Zhang, S., Yang, W., et al.: Real-time scheduling technology of photovoltaic-battery generation system based on intelligent fuzzy logic control. Power Syst. Technol. **42**(10), 3227–3233 (2018)
8. Xu, L., Xiong, M., Xiao, Y.: On-line prediction of application runtime using schedule historical data. Appl. Res. Comput. **37**(3), 763–767 (2020)
9. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks. Entropy **21**(9), 902 (2019)
10. Li, W., Qi, H., Xu, R., et al.: Data center network flow scheduling progress and trends. Chin. J. Comput. **43**(4), 600–617 (2020)
11. Dang, C., Liu, Y.: Research on network priority scheduling optimization method based on TRDP. Electron. Meas. Technol. **43**(2), 59–64 (2020)
12. Liu, S., Li, Z., Zhang, Y., et al.: Introduction of key problems in long-distance learning and training. Mob. Netw. Appl. **24**(1), 1–4 (2019)
13. Liu, S., Lu, M., Li, H., et al.: Prediction of gene expression patterns with generalized linear regression model. Front. Genet. **10**(8), 120 (2019)