Design of Human Resources Multi-dimensional Evaluation System Based on Big Data Mining

Haonan Chu[1] and Junlin Li[2](✉)

[1] School of Labor Relations and Human Resource, China University of Labor Relations, Beijing 100048, China
[2] Beijing Union University, Beijing 100101, China
bfgb545@163.com

Abstract. In order to improve the load quota and throughput performance of human resources multi-dimensional evaluation system, this study proposes a human resources multi-dimensional evaluation system based on big data mining. The system includes four parts: data acquisition module, data preprocessing module, evaluation module and feedback incentive module; According to the demand of system mining human resource data, the decision tree algorithm is used to mine human resource data, and then according to the evaluation purpose, characteristics and principles, the evaluation index construction steps are designed to build a multi-dimensional evaluation index system. On the basis of determining the factor set of evaluation indicators, calculate the weight of evaluation indicators, so as to establish a multi-dimensional evaluation model of human resources and realize the multi-dimensional evaluation of human resources. The experimental results show that the load and throughput test indexes of the system are within the design standard, which proves that the system has achieved the expected goal.

Keywords: Big data mining · Human resources · Multi-dimensional evaluation · System design

1 Introduction

In the final analysis, the competition between enterprises is the competition between the pros and cons of corporate human resources. With the development of science and technology, the economic rate of return of human resources is significantly higher than that of other resources. Statistics in the United States in recent years show that the income gap caused by human resources and physical capital investment is still widening. The ratio of the two has reached 4:1, which fully confirms that the role of human resources in economic development is much higher than other resources, and Plays a vital role in the development of enterprises [1].

Organizational productivity depends on the use and control of three factors, these three factors are capital, methods and human resources. Effectively organizing the system requires information about what is happening, as well as a mechanism for correcting

S.-H. Wang and Y.-D. Zhang (Eds.): ICMTEL 2022, LNICST 446, pp. 459–469, 2022.
https://doi.org/10.1007/978-3-031-18123-8_35

or adjusting inputs [2]. The productivity gains obtained by capital can generally be measured through sophisticated accounting systems. The benefits obtained by the method can be evaluated by a control system similar to this one. However, the contribution of human resources to productivity is difficult to measure. This dynamic performance can only be evaluated by the work output or work behavior of employees in a period of time. Performance appraisal is to use productivity effects and efficiency criteria to assign values to employee behavior or work output [3].

At present, the performance evaluation methods commonly used by most enterprises in China are still based on the traditional performance evaluation methods based on empirical judgment. Therefore, there are some problems, such as large subjective randomness, poor scientificity, single evaluation means, lack of objective evaluation standards, closed, opaque and closed evaluation, resulting in large deviation in the evaluation results. It frustrates the work enthusiasm of managers and employees, resulting in low organizational efficiency and difficult to achieve enterprise goals [4]. Therefore, establishing and improving a scientific and reasonable performance evaluation system is an urgent task for enterprise human resources development and management [5].

Therefore, in order to improve the CPU and memory occupancy rate, system load capacity and throughput, this study designed a multidimensional evaluation system of human resources based on big data mining, in order to provide better technical support for human resources work.

2 System Design

2.1 System Overall Architecture Design

The design of the human resources multi-dimensional evaluation system is divided into four parts: data acquisition module, data preprocessing module, evaluation and assessment module and feedback incentive module, as shown in Fig. 1.

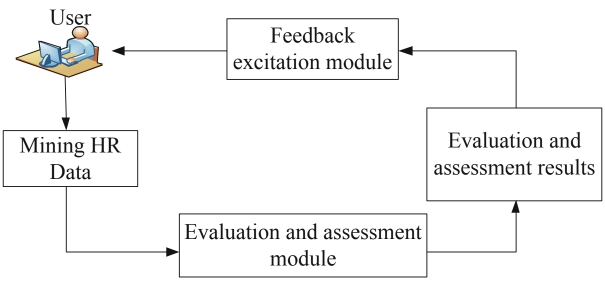


Fig. 1. Overall framework of the system

As can be seen from Fig. 1, the system in this paper takes employees as the main body of the system, and the final evaluation results will be fed back to employees, thus forming a closed loop for the whole system.

Firstly, data mining technology is used to process and analyze the relevant data of employees' work, and key characteristic indexes are extracted; Then, after preprocessing the extracted key indicators, they are used as the input of artificial intelligence algorithm according to the weight, and analyzed by artificial intelligence algorithm after training; Finally, grade evaluation is given to employees' work; Due to the transparency of the algorithm, the evaluation process of the system can be output and used as a feedback incentive to the evaluated employees.

2.2 Big Data Mining Human Resource Data

In data mining classification technology, there are mainly decision tree method, Bayes method, neural network method and rough set method [6]. A decision tree is a tree structure similar to a flowchart, in which each internal node of the tree represents a test of an attribute (value), and its branches represent each result of the test; and each leaf node of the tree It represents a category. The highest node of the tree is the root node. Its core algorithm is ID3 algorithm. For this reason, the decision tree method of ID3 algorithm is adopted to mine human resource data.

Assuming that the data set S contains a collection of s data samples, the category attribute can take m different values, corresponding to m different categories C_i, where $i \in \{1, 2, \cdots, m\}$. Assuming that s_i is the number of samples in category C_i, then the amount of information I required to classify a given data object is:

$$I(s_1, s_2, \cdots, s_m) = -\sum_{i=1}^{m} p_i \log(p_i) \tag{1}$$

In formula (1), $p_i = \frac{s_i}{s}$ represents the probability of belonging to category C_i for any data object.

Assuming that an attribute A takes v different values $\{a_1, a_2, \cdots, a_v\}$, the attribute A can be used to divide the set S into v subsets $\{S_1, S_2, \cdots, S_v\}$, where S_j contains the data samples of the S sets of attributes A and a_j values. If the attribute A is selected as the test attribute (used to divide the current sample set), let s_{ij} be the number of samples of the attribute C_i category in the subset S_j. . Then the information (entropy) h required to divide the current sample set using the attribute A can be calculated as follows:

$$h(A) = \sum_{j=1}^{v} \frac{s_{1j} + s_{2j} + \cdots + s_{mj}}{s} I\left(s_1, s_{2j}, \cdots, s_m\right) \tag{2}$$

In formula (2), item $\frac{s_{1j}+s_{2j}+\cdots+s_{mj}}{s}$ is regarded as the weight of the j subset, which is the sum of the sample data of attribute A and a_j in all subsets divided by the total number of samples in the S set [7]. The smaller the $h(A)$ calculation result, the more "pure" (good) the result of the subset division. For a given sub-set S_j, its information is:

$$I\left(s_1, s_{2j}, \cdots, s_m\right) = -\sum_{i=1}^{m} p_{ij} \log\left(p_{ij}\right) \tag{3}$$

In formula (3), $p_{ij} = \frac{s_{ij}}{|S_j|}$ represents the probability that any data sample in subset S_j belongs to category C_i. The information gain obtained by using attribute A to divide the corresponding sample set of the current branch node is:

$$Z(A) = I\left(s_1, s_{2j}, \cdots, s_m\right) - h(A) \tag{4}$$

In formula (4), $Z(A)$ represents the reduction of (information) entropy obtained by dividing the set according to the value of attribute A.

The decision tree induction algorithm calculates the information gain of each attribute, selects the attribute with the largest information gain as the test attribute of a given set, and generates the corresponding branch node. The generated nodes are marked as corresponding attributes, and corresponding (decision tree) branches are generated according to different values of this attribute, and each branch represents a divided sample subset [8].

The sample subset of the data set divided by the above process is the human resource data mined by the decision tree algorithm in the mass data. Next, based on these data, a multidimensional evaluation index system of human resources is established.

2.3 Construct a Multi-dimensional Evaluation Index System

The main purpose of this system is to achieve the following 8 purposes: to provide a basis for the promotion/demotion/transfer and resignation of employees, the organization's feedback on employee performance appraisal, to evaluate the contribution of employees and teams to the organization, and to provide a basis for employees' compensation decisions., Evaluate the decision-making of recruitment and selection and work distribution, understand the training and education needs of employees and teams, evaluate the effects of training and employee career planning, and provide information for work planning/budget evaluation/human resource planning.

Therefore, the system of this study needs to have five characteristics of conformity to reality, sensitivity, reliability, acceptability and practicality.

Based on this, the research index u is mainly composed of three parts: evaluation element a, evaluation mark b and status scale c, namely:

$$u = a + b + c \tag{5}$$

In formula (5), a represents the basic unit of the evaluation object; b is mainly used to reveal the key identifiable characteristics of the evaluation elements. It has a variety of forms. From the descriptive connotation, there are objective forms, subjective forms, semi-objective and semi-subjective forms. In terms of expression, there are short sentence, question, and direction indicators; c represents the degree of difference in evaluation elements or signs, and the state order and scale. It can be divided into quantifier, quantitative, hierarchical, symbolic, and graphical, Definition formula, comprehensive formula, etc. Therefore, multi-dimensional evaluation indicators have the functions of materialized connection, unified guidance, prevention of subjective one-sidedness, and deepening of understanding. To this end, according to the principle of homogeneity, reliability, universality, independence, completeness, and structure, a multi-dimensional evaluation index system is constructed according to the following steps:

1. Design of index content. It mainly includes the formulation of evaluation elements, the selection of evaluation signs and the division of evaluation scales. The design of each content has different ways and methods for candidates [9].
2. Classification, merger and design. Divide indicators by category.
3. Quantification of indicators. Quantification mainly includes three tasks: weighting, scoring and scoring. The specific contents are as follows:

   1) Weighting. Weighting refers to comparing the importance of different indicators in the "system". Weighting is actually a process in which the evaluation index system is vertically equalized and can be added. The weighting methods are: subjective experience method, A, B, C classification weighting method, expert survey method, comparative weighting method, Del non-weighting method, analytic hierarchy process weighting, multiple regression analysis weighting method, principal factor analysis weighting method, standard deviation Weighting method, etc.
   2) Assign points. That is, according to certain rules, a certain score is given to the "standard status" of each indicator and the degree of difference. There are many ways to assign points, including standard assignments, grade assignments, regular assignments, random assignments, precise assignments, fuzzy assignments, absolute assignments, relative assignments, secondary assignments, and statistical assignments., Decentralized scoring, these scoring methods have their own advantages and disadvantages, and the scoring method should be determined according to the specific evaluation situation.
   3) Scoring. Score cent is to point to when evaluating or after evaluating to evaluate result quantification and express, the form basically has statistical method, calculation method, judgment method, choice method.

4. Indicators are applicable.
5. Index test. Judge whether the index is qualified according to the inspection results. If not, return to step 1 to modify the index; If the indicators are qualified, the constructed indicators are output.

According to the above five steps, the indicators constructed in this study are shown in Fig. 2.

According to the multi-dimensional evaluation index system shown in Fig. 2, human resources are evaluated in six aspects: environment, personality, performance, ability, knowledge and morality.

2.4 Multi-dimensional Assessment of Human Resources

According to the index system shown in Fig. 2, the fuzzy comprehensive evaluation method is adopted to evaluate human resources in multiple dimensions. Based on this, the evaluation factor set is established as follows:

$$U = \{u_1, u_2, \cdots, u_n\} \tag{6}$$

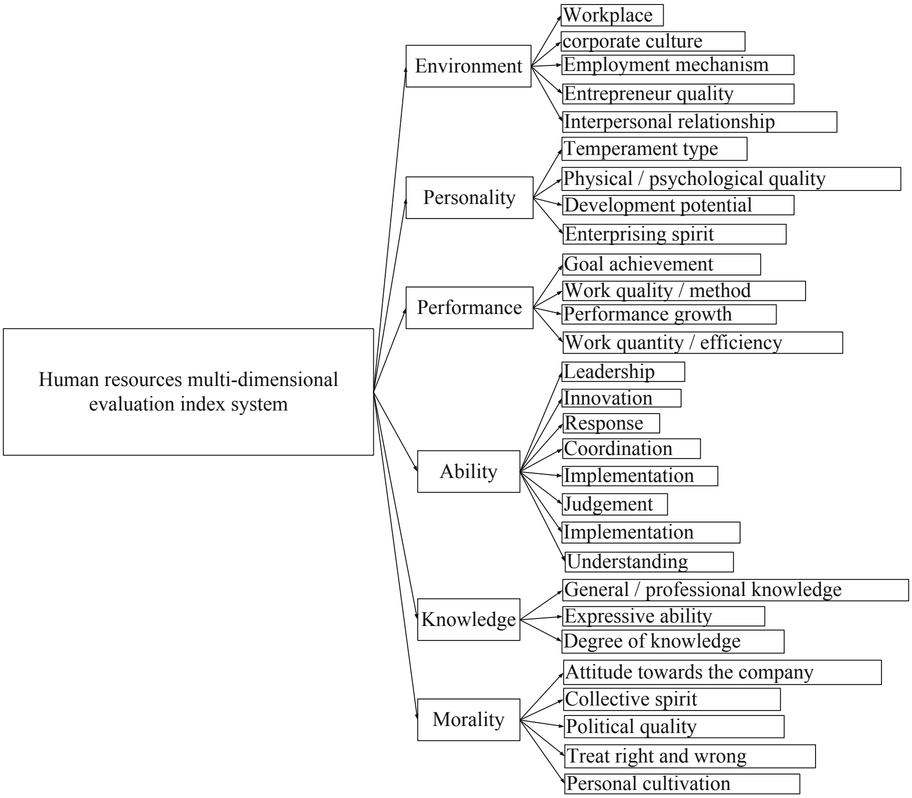


Fig. 2. Multi-dimensional evaluation index system

In formula (6), U represents the evaluation factor set; $u_i(i = 1, 2, \cdots, n)$ represents the i evaluation index, and n represents the number of indexes.

According to the evaluation factor set shown in formula (6), an initial matrix of initial data indicators of n evaluation indicators of m samples is established:

$$X = \{x_{ij}\}_{m*n} \tag{7}$$

In Eq. (7), X represents the initial matrix of dimension $m * n$; x_{ij} represents the element in row i and column j of the matrix. Since there are great differences in the dimension, order of magnitude and degree curve of each index, the initialization data shall be standardized:

$$x'_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}} \tag{8}$$

In formula (8), x'_{ij} represents the data after x_{ij} normalization. According to formula (8), the standardized matrix X' of the data can be obtained:

$$X' = \left\{x'_{ij}\right\}_{m*n} \tag{9}$$

The j information entropy value H_j obtained from the unit entropy function H is:

$$\begin{aligned} H &= -k \sum x_i' \lambda n x_j' \\ H_j &= -k \sum x_i' \lambda n x_{ij}' \end{aligned} \tag{10}$$

In formula (10), k represents Boltzmann's constant; λ represents the number of index states [10]. For a system with completely disordered information, the order is zero, and its entropy value is the largest, $H = 1$. When each sample of m is in a completely disordered distribution state, $x_{ij}' = \frac{1}{m}$, then:

$$H = -k \sum_{i=1}^{m} \frac{1}{m} \lambda n \frac{1}{m} = k \sum_{i=1}^{m} \frac{1}{m} \lambda nm = k \lambda nm = 1 \tag{11}$$

So there is: $k = (\lambda nm)^{-1} \quad 0 \le H \le 1$.

Since the information entropy H_j can be used to measure the utility value of the information of the j index (the number of indicators), when it is completely disordered, $H_j = 1$, at this time, the information of H_j (that is, the number of the j index) is important for the comprehensive evaluation The utility value is zero. Therefore, the information utility value of a certain index depends on the information entropy value H_j of the index, and the difference $E_j = 1 - H_j$ of 1.

It can be seen that using the entropy method to estimate the weight of each indicator is essentially calculated by using the value coefficient of the indicator information. The higher the value coefficient, the greater the importance for evaluation. Therefore, the weight W_j of the j index is:

$$W_j = \frac{E_j}{\sum_{i=1}^{m} E_j} \tag{12}$$

Assuming that the comment set is V and there are n evaluation samples, then there are:

$$V = \{v_1, v_2, \cdots, v_n\} \tag{13}$$

According to the comment set shown in formula (13), the membership degree of the comment set is set to $\varsigma(u_n) = e^{-\frac{2n-1}{10}}$, where e represents the natural constant. If there are m evaluation factors, the fuzzy relationship between the evaluation factor set U and the comment set V can be expressed by the evaluation matrix R:

$$R_i = \{r_{ij}\}_{n*m} \tag{14}$$

In formula (14), $n * m$ represents the dimension of the matrix; r_{ij} represents the element in row i and column j of the matrix. There is the following relationship with the evaluation factor set U and the comment set V:

$$r_{ij} = ur(u_i, v_i)\left(0 \le r_{ij} \le 1\right) \tag{15}$$

When the weight term W and the fuzzy relationship matrix R are known, the compound operation of the fuzzy matrix can be used to establish the fuzzy comprehensive model P of the evaluation index system:

$$P = W_{il} * R_{il} \tag{16}$$

In formula (16), W_{il} represents the evaluation weight vector of the i index and the group l; R_{il} represents the evaluation matrix of the i index and the l group.

The calculation results of Eqs. (12) and (14) are brought into Eq. (16), and the multi-dimensional evaluation results of human resources can be obtained. Realize multi-dimensional evaluation of human resources.

3 System Test and Result Analysis

In order to verify the practical application performance of the above designed human resources multi-dimensional evaluation system based on big data mining, the following experiments are designed.

In order to verify the performance of the client, network communication and server at the same time, the LoadRunner stress test tool is used to generate a large number of users in the system, make the system in an overload state, and detect the usage of CPU and memory of the system server, as well as the maximum, minimum and average response time of the system page.

Considering the performance requirements of the design system in practical application, the following experimental environment is designed: Set the maximum number of resource access users to 2000, the maximum response time for resource requests to 1s, and the CPU and memory usage to less than 30%. If the response time exceeds 1s and the CPU and memory usage exceeds 30%, the system does not meet the design requirements and needs to be rectified.

The test steps are as follows:

1. Use the LoadRunner stress test tool to create a script on the computer and enter the website to generate the system stress test parametric setting script.
2. Connect the generated script with the system, create different resources in the system and record the script;
3. Set the number of virtual users loading resources in different resources to test the system pressure.

3.1 This Section Describes How to Test the System CPU and Memory Usage

First, test the load rate of the system, and the result is shown in Fig. 3.

It can be seen from Fig. 3 that when the virtual users in the system reach 10,000, the CPU and memory occupied by its running resources are 25% and 27%, respectively, which are 5% and 3% smaller than the maximum occupancy rate, which is already very high. Close to the limit of CPU and memory usage.

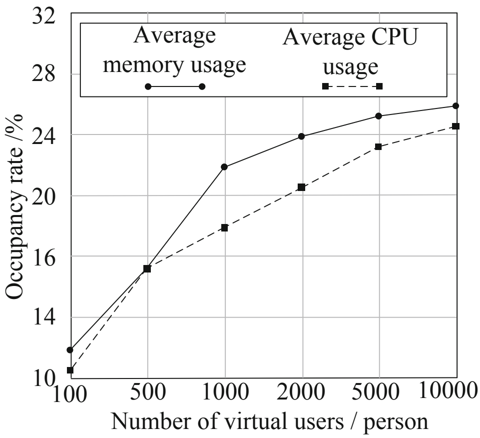


Fig. 3. CPU and memory usage

3.2 System Load Rate Test

The system runs the Vuser test script designed in this experiment 30 times to perform system pressure test. When the number of script runs reaches 30, after 5 min, the script runs are reduced by 5 times every 30s to perform system decompression test. The system load test result is shown as in Fig. 4.

It can be seen from Fig. 4 that under the load pressure shown in Figure (a), the system load test time is 11 min and 5 s, the average number of clicks is 160.54 s/time, and the average transaction response time is 4.467s, both in Within the scope of the system test standard designed in this experiment.

3.3 System Throughput Test

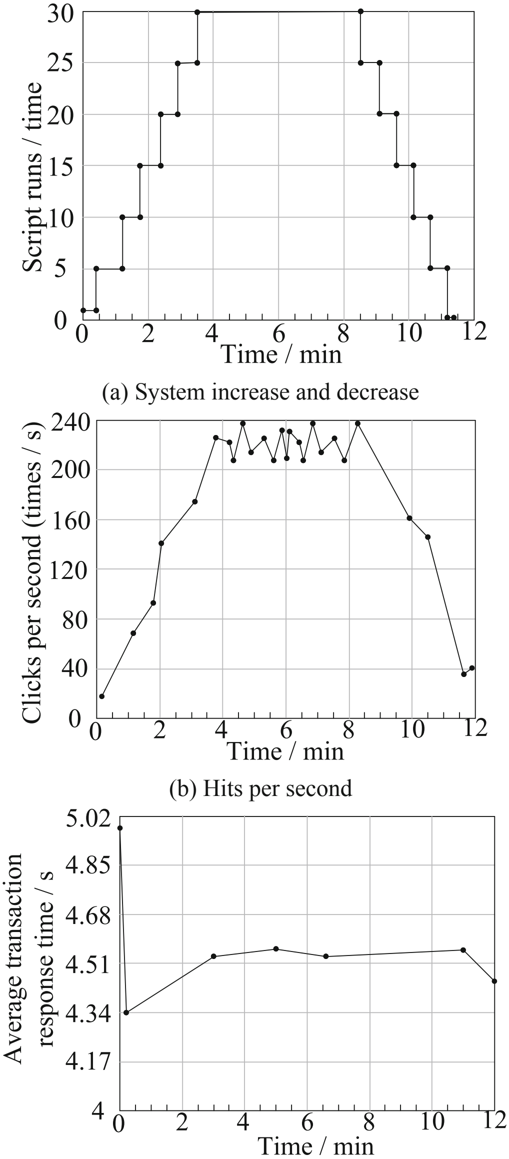
Based on the system load test results, test the system throughput, and the test results are shown in Fig. 5.

As can be seen from Fig. 5, the system throughput will change with the change of the number of concurrent users in the system, showing a direct proportional relationship with the system load. Thus, the server operation of this system tends to be stable gradually.

Based on the above three experimental results, it can be seen that the function and performance of the evaluation system designed in this paper have reached the expected goal, and it is suitable for practical application.

4 Conclusion

In this paper, combined with the theory of performance evaluation, according to human resources evaluation principles, objectives, characteristics, the establishment of performance evaluation index system. Then the fuzzy comprehensive evaluation model of performance evaluation is constructed by using fuzzy mathematics theory. From the



(a) System increase and decrease

(b) Hits per second

(c) Average transaction response time

Fig. 4. System load test result graph

overall experimental results, the system in this paper has a good application effect in load and throughput.

Because the methods and theories based on in the research process have been quite mature and perfect, so the research on these theories and methods is grounded and

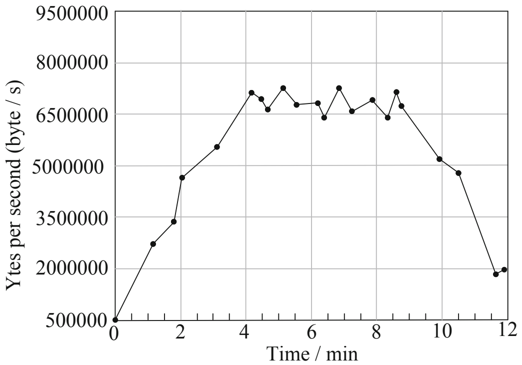


Fig. 5. System throughput test

scientifically feasible. At the same time, the method as an expert system can also provide a certain reference value for other enterprises system evaluation.

References

1. Xie, R.: Design of Java EE based human resource management system with friendly man? machine interaction. Mod. Electron. Techn. **44**(8), 114–118 (2021)
2. Su, B., Liu, S., Lu, Y., et al.: Evaluation of human resource allocation of primary healthcare in China: based on agglomera-tion degree. Chin. J. Health Policy **14**(4), 49–54 (2021)
3. Song, S., Wei, L., Liu, Y., et al.: Application of Data Envelopment Analysis in evaluation human resources efficiency of nursing in department of cardiology. Chin. Nurs. Manage. **21**(5), 745–749 (2021)
4. Feng, T., Zhao, M., Bu, Q.: Practice of the comprehensive evaluation system through combining the workload of RBRVS with specific performance. Chin. Hosp. Manage. **41**(8), 51–53 (2021)
5. Yi, W., Cheng, Z.: Research on the performance evaluation mechanism of new think tank: from the dual perspective of stimulation and constraint. J. Xiangtan Univ. (Phil. Soc. Sci.) **44**(3), 56–60 (2020)
6. Congman, W., Yunye, Z.: Research on the evaluation of human resource supply quality of domestic service in Beijing, Tianjin and Hebei Based on AHP-TOPSIS Model. Contemp. Econ. Manage. **42**(1), 78–83 (2020)
7. Zhu, Y., Yu, W.: Construction of evaluation index system of china's public sports resource allocation level in new era. J. Wuhan Inst. Phys. Educ. **54**(3), 5–12 (2020)
8. Weng, Y., Cheng, L., Bai, Y., et al.: Construction of evaluation index system of nursing human resources efficiency. Chin. Nurs. Res. **33**(11), 1855–1859 (2019)
9. Jia, H., Lijuan, X., Huan, J.: Development of indexes for evaluating nursing human resource efficiency in geriatric department based on data envelopment analysis. J. Nurs. Sci. **34**(14), 53–55 (2019)
10. Chen, L., Wang, C.: Research and simulation of failure-aware algorithms for trusted cloud resource scheduling. Comput. Simul. **37**(10), 334–337+356 (2020)
11. Zhu, B., Chen, G., Li, P., et al.: Design of human resource management system based on B/S mode and MySQL. Mod. Electron. Tech. **44**(14), 65–69 (2021)
12. Xu, J., Liu, C.: Design of evaluation system based on artificial intelligence and deep data analysis technology. Electron. Des. Eng. **29**(6), 179–183 (2021)