Human Resource Scheduling Control Method Based on Deep Reinforcement Learning

Shaoping Zhang(✉) and Bo Sun

School of Labor Relations and Human Resource, China University of Labor Relations, Beijing 100048, China
kaka5411000@126.com

Abstract. With the increasing importance of human resource management in project process management, human resource scheduling and control methods also need to keep pace with the times. Continuing to use the traditional human resource scheduling control method will waste a lot of potential value of human resources. Therefore, a human resource scheduling control method based on deep reinforcement learning is proposed. By constructing the human resource scheduling control model, the minimum human cost expenditure under various constraints is obtained; Design the deep reinforcement learning algorithm, and design the scheduling algorithm for specific scheduling objectives for the human resources scheduling control center; Create a model-based human resource scheduling management and control evaluation system, and improve the relationship between the comprehensive evaluation value and the advantages and disadvantages of multi project human resource scheduling. Experiments show that this human resource scheduling control method can guide different problems to allocate goals. Considering the time factor, the optimal solution of human resource scheduling and control can be obtained.

Keywords: Deep reinforcement learning · Human resources · Dispatching control

1 Introduction

Resource management generates a resource list for a single resource or a group of resources. Resource allocation strategies for various tasks can be configured, but the complexity of human resources is not considered. At present, scholars at home and abroad have many research schemes about human resource scheduling. These studies regard the human resource scheduling problem as a combinatorial optimization problem, and can roughly divide the methods to solve the human resource scheduling into three categories: heuristic algorithm, local search and optimization algorithm [1]. In recent years, scholars at home and abroad have expanded based on the above three solutions and developed more theoretical and practical results. In China, there are solutions to solve the human resource model according to the genetic algorithm to obtain the optimal solution, and there are also solutions to using the input scheduling strategy. The relevant

S.-H. Wang and Y.-D. Zhang (Eds.): ICMTEL 2022, LNICST 446, pp. 368–382, 2022.
https://doi.org/10.1007/978-3-031-18123-8_28

human resource models can obtain a probability distribution of the project cost, and the corresponding resource adjustment solution can be obtained by using the stochastic optimization. Or the human resource scheduling method based on process agent is used to model human resources, and the task and consortium bidding model are described and defined [2, 3].

The scheduling plan is generated according to the user's preference for the target. In foreign countries, one kind of modeling is based on dividing the software process into four stages: coding, rework, testing and project closure to construct the resource model and design the algorithm to solve the scheduling [4]. Another model scheduling strategy is to set personnel as fuzzy variables based on their professional skill level and uncertain activity cycle, and solve the scheduling for such a resource model. The time constraint grid model of task relationship is established, the model is modeled with mathematical formula, the relationship between tasks is judged according to the time constraint conditions, and finally the conclusion of whether resources conflict is obtained [5–7]. The resource conflict is detected based on the rule conflict of bit vector intersection operation. The algorithm is based on asbv algorithm and uses the divide and conquer idea and bit vector technology. Only one bit vector intersection operation is required for one-dimensional rule component processing, which greatly improves the processing efficiency [8].

Reinforcement learning, also known as reinforcement learning, is one of the paradigms and methodologies of machine learning. It is used to describe and solve the problem that agents maximize rewards or achieve specific goals through learning strategies in the process of interaction with the environment. The common model of reinforcement learning is the standard Markov decision process [9]. According to the given conditions, reinforcement learning is divided into model-based reinforcement learning (model-based RL) and model-free reinforcement learning (model-free RL). Reinforcement learning also includes more complex research directions, such as reverse reinforcement learning, hierarchical reinforcement learning and Multi-Agent Reinforcement Learning. The basic principle of reinforcement learning is to let the agent interact with the environment continuously, update the strategy by using interactive samples and feedback information, and use the strategy to finally obtain the optimal strategy [10, 11].

To sum up, this paper proposes a human resource scheduling control method based on deep reinforcement learning, which combines human resource scheduling with deep reinforcement learning to improve the management efficiency of enterprises.

2 Human Resource Scheduling Control Method Based on Deep Reinforcement Learning

2.1 Construction of Human Resource Scheduling Control Model

The purpose of human resource management is to improve production efficiency, reduce enterprise costs and increase profits. In order to achieve the established production indicators, it is necessary to assign manpower to each stage of task implementation, and allocate manpower scientifically and reasonably. The optimal allocation and effective utilization of manpower need to carry out targeted training to improve the ability of employees to deal with various professional management work and improve their comprehensive quality. In terms of quantitative overall arrangement and application analysis of various resources, many industries have been using linear programming to realize the optimal allocation of limited resources such as human, financial and material resources in the economic management system. Among the research branches of operations research, linear programming has the characteristics of early exploration, rapid progress, easy understanding, universal application and mature algorithm research [12, 13].

As long as we engage in economic activities such as economic management, transportation, industrial and agricultural production, there will be a need to improve economic benefits. There are usually two ways to improve economic benefits: first, technological improvement. For example, select new materials, new equipment and improve the production process. Second, scientifically allocate human, financial and material resources and optimize production plans.

As a mathematical method to assist scientific management of practical problems, linear programming studies how to complete the optimal allocation of various resources and improve economic benefits according to the limitations of various conditions. Objective function, decision variables and constraints are the three elements of linear programming. Usually, the linear programming problem is to solve the maximum or minimum value of the objective function under linear constraints. The feasible solution is the solution that meets the linear constraints. A feasible region is a set containing all feasible solutions. Integer programming usually needs to round the set variables (part or all). If the variables in the linear model are limited to integers, the integer programming is called integer linear programming. Its general form is: list the objective function and constraints; Draw the feasible region (when there are many variables, it can be analyzed with the help of programming software). The model is established and solved within the feasible range to obtain the optimal value and optimal solution of the objective function. The calculation formula is as follows:

$$Maxz = \sum_{j=1}^{n} c_j x_j \tag{1}$$

where, z represents the optimal solution of the programming function; j represents linear variable of objective function; c represents integer linear programming vector; n represents the feasible region constant.

Establishing a mathematical model to solve practical problems usually has the following three steps: first, put forward assumptions, find decision variables, and determine them according to the factors affecting the purpose to be achieved; Determine the objective function, which is determined by the functional relationship between the decision variable and the goal; The constraints to be satisfied by setting the decision variables are determined by the constraints on the decision variables. Before the application and popularization of electronic computers, the solution and calculation of linear programming were quite complex. With the maturity of computer application technology, it is more and easier to use the powerful computer processing systems to solve linear problems. As a tool for solving linear and nonlinear optimization problems, LINGO software itself has a simple and easy to learn built-in modeling language. It can express complex logical relations with the help of various internal functions, making it possible to round decision variables (i.e. integer programming), which is convenient to use and has high running speed. In terms of data interaction, the connection between the model and external data (such as Excel spreadsheet) can be realized by using input and output interface functions.

Enterprises adapt to the fierce competitive market, formulate real-time, accurate and reasonable personnel allocation schemes, and optimize the allocation of personnel, which all put forward practical requirements for the application of linear programming knowledge. In the past, the planning of enterprises was more complex. On the one hand, it was necessary to comprehensively consider the needs of customers of the construction party, on the other hand, it was also necessary to balance the profits of enterprises. Manual calculation was inaccurate and time-consuming. It was much easier to use linear programming and then use computers to complete the calculation and solution, which could improve the reliability and scientificity of enterprise decision-making. The decision-making theory is formed on the basis of strict theory, which can be obtained by analyzing and applying large-scale accurate basic data and strict mathematical solution. This model introduces the human capital theory, based on the concept of project management, divides the supervision projects of a certain scale, and realizes the multi project management under the restriction of human resources. Next, we use linear programming to obtain the optimal scheme for personnel scheduling from the aspects of the company's human resources arrangement in the cost standard of professional supervisors, project progress, project scale and customer needs, and the relationship between projects and supervisors, so as to achieve the goal of reducing the human cost of supervision projects.

In order to ensure that the model framework of this study is in line with practical application and scientific and reasonable, the following assumptions are put forward: due to the different supervision work contents and supervision contract requirements, different project supervision work needs to be carried out, and different supervision personnel need to be configured. The time is in months. The number of project personnel per month will change due to the progress plan of each project. It is necessary to use linear programming software to solve all kinds of supervisors required by the enterprise. The enterprise supervisors have the background of civil engineering, electromechanical or other housing and municipal public works, have learned relevant professional knowledge, and the work content is consistent with their major. For unrestricted work, all supervisors can exchange posts without obstacles, and there are only differences in work execution efficiency. It is necessary to assume that similar supervisors have the same monthly cost and work output value, calculate the monthly cost of various supervisors and the monthly output value of each type of supervisors respectively, and then summarize the total labor cost. The supervision work of this model does not consider the difference of project benefits and overtime factors. If it is necessary to refine this aspect, new constraints shall be added. If the parameters or constraints change due to irresistible factors, it is necessary to set new conditions to replace the human resource scheduling.The flow chart of human resource scheduling control model is shown in Fig. 1.

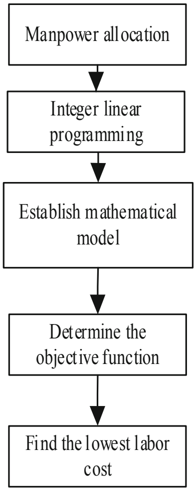

Fig. 1. Flow chart of human resource scheduling control model

The objective function expression of the human resource scheduling control model constructed in this paper is:

$$Minz = \sum_{j=1}^{n} c_j x_j zj \tag{2}$$

$$WPij(i, j, k) = WPi(i, k) \times Ujk(j, k) \tag{3}$$

$$XPij(i, j, k) = WPij(i, j, k)/W_j(j) \tag{4}$$

$$Xij(i, j) = \sum_{k=1}^{e} XPij(i, j, k,) \tag{5}$$

where, i refers to the i month of planned work execution; j refers to the j supervisor; k refers to e projects in total in the k project, and e refers to the total number of supervision projects to be implemented within the planned time; Xij refers to the total number of supervisors of type j required for all projects in the i month; $XPij$ refers to the number of j supervisors required for the k project in the i month. The constraint expression for building the model is:

$$Xi(i) = \sum_{j=1}^{n} Xij(i, j) \tag{6}$$

$$Xj(j) = \sum_{i=1}^{n} Xij(i, j) \tag{7}$$

The purpose of this study is to find the minimum labor cost under various constraints. In order to carry out the supervision work of each project, different kinds of supervisors must be dispatched to carry out different supervision work under the condition of meeting various constraints.

2.2 Algorithm Design Based on Deep Reinforcement Learning

After the completion of the human resource scheduling control model constructed above, in terms of the agent, the required status information includes the agent type and whether the agent has selected the target. Here, considering the time factor, when the agent has selected the target, it is assumed that the movement of the agent has been completed. The agent's non motion attribute value is 1 and the already motion attribute value is 0. The status information required by the target includes the target type, target residual value and whether the target currently has obstacle areas. The target obstacle area attribute value is 1, and the non-existent attribute value is 0. In this way, there are 8 agents in total, each agent needs 2 values to represent the state, there are 9 targets in total, and each target needs 3 values to represent the state.

Considering the time factor, the definition of action selects the m-th target for the current n-th agent and moves, where $n \in [1, 8]$ and $m \in [1, 9]$, so the size of action space is 72.Because this action space is too large, the design here is divided into two actions. The space size of action 1 is 8, indicating which agent selects the target, and the space size of action 2 is 9, indicating which target the agent selects. In this way, the size of action space can be reduced from 72 dimensions to 17 dimensions, reducing the parameters of deep reinforcement learning, making the calculation of deep reinforcement learning network faster and making the design of action more meaningful [14]. After each action decision, the environment needs to use the path planning intelligent decision model for decision-making, calculate the arrival success rate according to the determined path, and then interact according to the environmental success rate. Therefore, the success rate and value coefficient cannot be directly used to calculate the score, and the real score under the current episode can only be calculated according to the real environment interaction. According to the success rate, when the agent can reach the specified goal, the reward is set to obtain a score multiplied by 100, and when it cannot reach the specified goal, the reward is set to - 10. Considering the time factor, there will be a special case, that is, the agent determined by the current action has selected the target and completed the motion, so the current action is meaningless. When the decision makes such an action, set the reward to - 10. In this way, actions are continuously selected according to the state, and each episode ends until all agents move.

Fitness function is the evaluation of individuals. An individual is a set of objective allocation solutions. Therefore, the final score obtained by the agent is directly used as the fitness function to evaluate the individual. The score is calculated by the target allocation scheme corresponding to the individual, the movement success rate of the agent and the value coefficient. Deep reinforcement learning combines the perception ability of deep learning with the decision-making ability of reinforcement learning, which can be controlled directly according to the input image. It is an artificial intelligence method closer to human thinking mode. It can solve more complex tasks that are closer to the current situation. It uses a depth network to represent the value function. According to

Q-learning in reinforcement learning, it provides the target value for the depth network and updates the network continuously until convergence. A fixed target network with better training stability and convergence.

The flow based on the deep reinforcement learning algorithm is shown in Fig. 2.

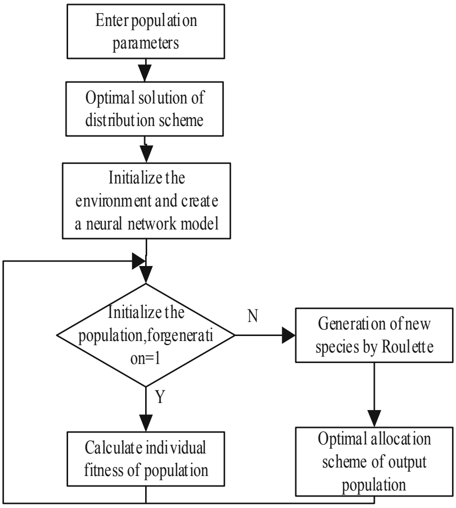

Fig. 2. Flow chart of deep reinforcement learning algorithm

As shown in Fig. 2, input population size, population size, crossover rate, mutation rate, genetic algebra, DNA size, disease and gene location boundary. Optimal solution of output allocation scheme; Initialize the environment, and create a neural network model according to the parameters of the environment initialization algorithm; Initialize the population pop, forGeneration = 1, translate the individuals in the population into allocation scheme, calculate the score in the environment according to the allocation scheme corresponding to the individuals as the fitness of the population individuals, and record the optimal individual fitness of the population; According to the fitness function, the roulette method is used to select and generate a new population pop, generate random numbers, and randomly select cross loci; Output the current population pop optimal individual corresponding allocation scheme. The policy trajectory of scheduling all job sets is obtained through the initialization model; Calculate the baseline at each time of all job sets; The difference between baseline and score function is used to update the parameter weight of the model. Next, the implementation details of imitation learning are described in combination with the code.

```
Time register level (void * ARG)
struct multiboot_ uinfo*mb= (struct multiboot_ Uinfo *) parameter;
EDF_ uregister_ ulevel (EDF_ Uenable all); / / level 0: EDF
CBS_ uregister_ ulevel (CBS_ Uenable all, 0); / / level 1: CBS
RR_ uegister_ ulevel (RRTICK, RR_ .MAIN_ Yes, MB); / / level 2:
loop
dummy_ uregister_ Ulevel(); / / Level 3: Virtual
Register module ();
1 / resource access protocol
CABS_ uregister_ umudule ();
//Resource access protocol
Warning sound;
```

The multi project scheduling problem with limited human resources involves several parallel projects and a shared human resource database, which contains human resources with limited supply (a kind of renewable resources). Due to the complexity of several projects involved, it is assumed that the projects are independent of each other except for sharing the human resource database, there is no tight relationship between projects and between different project tasks, and there is no constraint relationship on other types of resources except human resources, that is, Competition for limited human resources is the only relationship between these parallel projects.

After the above implementation steps of in-depth reinforcement learning, the scheduling algorithm of specific scheduling objectives can be designed for the human resources scheduling control center.

2.3 Evaluation System of Human Resource Scheduling Management and Control Based on Model

The evaluation of multi project human resource scheduling based on the model refers to the evaluation of multi project human resource scheduling based on the model and suitable for the model. It is an evaluation matched with the model, which aims to find the situation of enterprise multi project human resource scheduling. It is not difficult to see that the evaluation is in the position of the supervisor of the implementation of the model.

The scientific principle of the evaluation index system means that the evaluation content should be scientific and standardized. The concept of each index should be scientific, accurate, with accurate connotation and extension, and the calculation scope should be clear without ambiguity; The index system should reflect the essential characteristics of the evaluation object as comprehensively and reasonably as possible. It is not a simple combination of indexes, but should reflect the overall effect and internal relationship of the project and the quantifiable degree of each index; The establishment of index system should minimize the subjectivity of evaluators and increase objectivity. Therefore, expert opinions should be widely solicited; The establishment of the index system must be guided by advanced scientific theory, which can reflect the objective and actual situation of the evaluation object. No matter what evaluation method is adopted and what evaluation mathematical model is established, the index system must be an abstract description of objective reality.

The principle of system optimization requires that the number of indicators and the structural form of the index system should be based on the principle of comprehensively and systematically reflecting the evaluation objectives, and the evaluation index system should be established from the overall perspective. The indicator system shall be composed of several indicators. Indicators of different types in the system structure shall not be combined with each other, and the main indicators and accompanying indicators shall not be juxtaposed. Systematization first requires to avoid the index system being too complex to avoid the evaluation being difficult to implement, and to avoid too few indicators ignoring some important factors and difficult to reflect the internal essence of the project. Therefore, we should neither be all inclusive nor lose sight of one or the other. We should build a reasonable index system with fewer indicators as far as possible to achieve the purpose of optimizing the overall function of the index system. Secondly, it is required to take into account the relationship between current and long-term, overall and local, qualitative and quantitative.

The principle of comparability requires that the indicators must reflect the common attributes of the evaluated items, and the quantities of different indicators must be transformed into the same unit before they can be compared and calculated. The stronger the comparability, the more credible the evaluation results will be. In order to make the evaluation indicators universal, the designed evaluation indicators should adopt domestic and international standards or recognized concepts as far as possible. In the first mock exam, the factors that should be eliminated and the influence of environmental factors under specific conditions should be eliminated. The factors that can not be compared to the factors can be transformed into comparable factors, and the data of evaluation can be transformed into a unified equivalent value or dimensionless value, so that the evaluation indexes can be compatible with the same model and make them comparable.

The principle of practicability requires that the index system must have clear meaning, standardized data, moderate complexity and easy calculation. The requirements specified in the evaluation index shall conform to the actual situation of the evaluated object, that is, the specified requirements shall be appropriate, that is, they shall not be too high or too low. In order to facilitate practical use, a specific and measurable index system must be designed to characterize the main aspects of the goal. There should be enough information for the contents specified around the indicators, and there must be practical quantitative methods for use. The practicability also requires that the established indicators should have levels and key points, the qualitative indicators can be quantified, and the quantitative indicators can be measured directly, so as to make the evaluation work simple, save time and cost, and facilitate computer processing.

The independence of evaluation indexes means that the indexes in the evaluation index system should be independent of each other and can not be subordinate to and overlap with each other. This is because if the two indicators in the index system reflect the same attribute of the evaluated object, it will cause the repetition of the evaluation content, increase the workload of the evaluation, and even reduce the feasibility of the evaluation. Moreover, the repeated calculation of the content indicators twice is equal to increasing the weight of the evaluation indicators of the project, which will naturally affect the scientificity of the whole evaluation.

The goal oriented principle means that the designed evaluation index system must fully reflect the evaluation objectives and fully reflect the basic principles based on the goal center, which requires that all indicators in the evaluation index system should be consistent with the objectives.

The core of multi project management is how to reasonably allocate various resources among various projects; The goal of multi project management is to solve the competition between multiple projects in terms of capital, time and resources by allocating enterprise resources and optimizing enterprise resource allocation, so as to reduce the project cost and improve the profit margin of the enterprise. The goal of multi project human resource scheduling in software development enterprises is to reasonably schedule all kinds of human resources among multiple projects to maximize the interests of the enterprise. In order to keep consistent with the model, the goal of maximizing enterprise interests is still adopted here.

From the four elements of modern project management, the goal of enterprise multi project human resources scheduling should cover three elements: quality, progress and cost, and satisfy all stakeholders. Therefore, the criterion layer of the evaluation index system is: quality, progress, cost and stakeholders, as shown in Fig. 3.

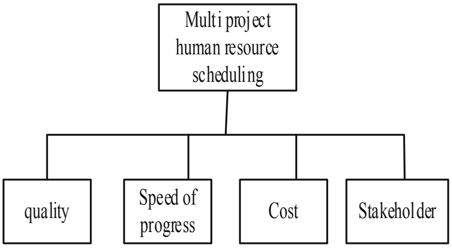

Fig. 3. Original index architecture

As shown in Fig. 3, it is the original index system structure established in this paper. The progress goal of the model established in this paper is the progress goal on the premise of ensuring the quality: the ranking of the relative importance of multiple projects makes the relationship between the main stakeholders (customers, teams, governments and enterprises) and some expenses (breach penalty) It is also included in the progress. Here, the indicators of multi project human resource scheduling evaluation of software development enterprises are simplified to the progress indicators that include the relevant interests of the enterprise (reflected in the different relative importance of multi projects). Comprehensively considering the relative importance of multiple projects and the model, the following evaluation scales are determined, as shown in Table 1.

Table 1. Evaluation items and corresponding symbols

Evaluation items	Corresponding symbol
Proportion of completed projects with weight coefficient of 9 to the total number of projects with weight coefficient of 9	N9
Proportion of completed projects with weight coefficient of 8 to the total number of projects with weight coefficient of 8	N8
Proportion of completed projects with weight coefficient of 7 to the total number of projects with weight coefficient of 7	N7
Proportion of completed projects with weight coefficient of 6 to the total number of projects with weight coefficient of 6	N6
Proportion of completed projects with weight coefficient of 5 to the total number of projects with weight coefficient of 5	N5
Proportion of completed projects with weight coefficient of 4 to the total number of projects with weight coefficient of 4	N4
Proportion of completed projects with weight coefficient of 3 to the total number of projects with weight coefficient of 3	N3
Proportion of completed projects with weight coefficient of 2 to the total number of projects with weight coefficient of 2	N2
Proportion of completed projects with weight coefficient of 1 to the total number of projects with weight coefficient of 1	N1

As shown in Table 1, each evaluation item is replaced by corresponding symbols, and each value corresponding to each evaluation item is the lower bound of the completion proportion range corresponding to the corresponding evaluation scale. After the evaluation scale of the evaluation project is determined, the weight of the evaluation project needs to be calculated. Here, the weight is calculated by the pair by pair comparison method. Like the determination of evaluation scale, the weight of evaluation items also has a certain subjectivity. The evaluation of comprehensive evaluation value is divided into two cases: one is the evaluation of single multi project comprehensive evaluation value, and the other is the evaluation of multiple multi project comprehensive evaluation values. The evaluation of single multi project comprehensive evaluation value refers to the evaluation of the scheduling results of a group of multi projects with human resource conflicts. The evaluation system is set as follows: if all projects are completed on schedule, the comprehensive evaluation value is 6.507. If 90% of all projects are completed on schedule, the comprehensive evaluation value is 5.508. If 80% of all projects are completed on schedule, the comprehensive evaluation value is 4.536. If 70% of all projects are completed on schedule, the comprehensive evaluation value is 3.591. If 60% of all projects are completed on schedule, the comprehensive evaluation value is 2.7.

The comprehensive evaluation value can be associated with the advantages and disadvantages of multi project scheduling. Due to the great difficulty of multi project practice management, in order to encourage enterprises to adopt more suitable scheduling methods, 4 + 1 evaluation indexes are adopted here: excellent +, excellent, good, medium and poor, as shown in Table 2.

Table 2. Correlation between comprehensive evaluation value and multi project schedule

Serial number	Comprehensive evaluation range	Advantages and disadvantages of multi project team progress
1	5.508–6.507	Excellent +
2	4.536–5.508	Excellent
3	3.591–4.536	Good
4	2.7–3.591	Medium
5	0–2.7	Subalternation

As shown in Table 2, it is the correlation table between comprehensive evaluation value and multi project progress. The evaluation of multiple multi project comprehensive evaluation values can be completed by comparing their comprehensive evaluation values. The multi project human resource scheduling with large comprehensive evaluation value is better than that with small comprehensive evaluation value.

3 Experiment and Analysis

3.1 Experimental Preparation

In order to verify the effectiveness of the human resource scheduling and control method based on deep reinforcement learning, the following experimental operations are carried out. In this experiment, the depth reinforcement learning PPO algorithm is used, the attenuation rate is 0.99, the target KL divergence parameter is 0.01, game lambda is 0.97, and the shear objective function parameter is 0.2. The experimental algorithm model is divided into actor network and critical network. In this paper, a 4-layer fully connected neural network is used to build the actor network. The size of the state space of the input layer and the model is the same as 43 dimensions. The hidden layer has 2 layers and 64 neurons. The relu6 activation function is used. Because there are two actions in action modeling, the output layer of actor network has two outputs with dimensions of 8 and 9 respectively. Four layers of fully connected neural network are used to build a critical network. The input layer is the same as 43 dimensions, the hidden layer is the same as actor network, and the output layer is only 1 dimension. The output value estimates the value of the actions output by actor network. During network update, both actor network and critical network are optimized by Adam optimizer. The learning rate of actor network is 0.00002 and critical network is 0.00005. In this experiment, 10000 epochs are trained, and 1000 decisions are made in each epoch, so the batch size of the neural network is 1000. During the training, the reward sum of each episode and the loss function at each update are saved. After the trained intelligent decision model is obtained, because the score expectation cannot be calculated directly, the environment is allowed to interact as much as possible and use the model for decision-making to calculate the score expectation of the decision model.

3.2 Result Analysis

The experiment in this paper adopts the deep reinforcement learning PPO algorithm, which does not plan according to the model, but trains the model through the samples generated by the environment. The score expectations of the deep reinforcement learning algorithm proposed in this paper are compared with those of ergodic method, genetic algorithm and microbial genetic algorithm. All the experimental results of target allocation are shown in Table 3.

Table 3. Comparison of experimental results of target allocation

Method	Is time taken into account	Score expectation
Ergodic method	No	5.3895411
Genetic algorithm	No	6.1582693
Microbial genetic algorithm	No	6.0214698
Deep reinforcement learning algorithm	Yes	4.1023458

As shown in Table 3, the following conclusions can be obtained by comparing and analyzing the experimental results of genetic algorithm and deep reinforcement learning. Genetic algorithm is very fast and can find the optimal solution of target allocation. However, the time factor is not considered, so there may be a better solution when the time factor is considered. Secondly, the genetic algorithm directly calculates the optimal solution according to the problem. When changing a problem, it needs to be recalculated, so it can not generate a general intelligent decision model. Deep reinforcement learning can save the decision model through neural network. When the state modeling includes location information, it can train a general intelligent decision model to guide different problems for target allocation. Considering the time factor, it can obtain a better solution, the number of samples required for training is less, the time is shorter, and the ideal effect is achieved.

In order to further verify the convergence of this method to human resource scheduling, test the performance of the algorithm, and the comparison results are shown in Table 4.

It can be seen from the above table that the number of iterations of this method is only 4 and 5, and the response time is 1.36 s and 2.45 s. Under different experimental conditions, this method can always respond in a short time. It shows that the optimization time of human resource scheduling in this method is short and the response ability is improved.

Table 4. Comparison of algorithm results

Method	Parameter performance	Experiment 1	Experiment 2
Ergodic method	Number of iterations Response time/s	14 3.56	15 12.45
Genetic algorithm	Number of iterations Response time/s	11 3.86	14 13.45
Microbial genetic algorithm	Number of iterations Response time/s	17 5.56	18 18.45
Deep reinforcement learning algorithm	Number of iterations Response time/s	4 1.36	5 2.45

4 Conclusion

The human resource scheduling and control method based on deep reinforcement learning proposed in this paper not only helps project managers identify the key human resources that determine the success or failure of the project, but also enables project managers to understand the overall ability level of existing human resources and formulate corresponding training and development plans. Build the human resource scheduling control model, set new conditions and replace the human resource scheduling. Considering the time factor and based on in-depth reinforcement learning, the scheduling algorithm with specific scheduling objectives can be designed for the human resource scheduling control center.When managing each link of optimal allocation of human resources in multi project environment, the more mature management model in each link is used for reference, which improves the efficiency and operability of management. In the evaluation of project priority, the index value established in this paper is limited, and there is no index of financial dimension. In the actual project management, capital is also an issue that must be considered. Therefore, in the future research, the index database will be continuously enriched to make the project priority management more perfect.

References

1. Otoo, F.: Human resource management (HRM) practices and organizational performance. Empl. Relat. **41**(5), 949–970 (2019)
2. Chen, J., Tong, S., Xie, H., et al.: Model and algorithm for human resource-constrained R&D program scheduling optimization. Discrete Dyn. Nat. Soc. **2019**, 1–13 (2019)
3. Hwang, K.J., Kang, P.K., Dong, J.J.: The Association between human resources investments in the internal accounting control system and unfaithful disclosure designations. Korean J. Manag. Account. Res. **20**(2), 55–78 (2020)
4. Imran, M.: Influence of human resources, information technology, internal control system and regional financial supervision on value of financial reporting information. Al-Kharaj J. Islamic Econ. Bus. **2**(1), 49–65 (2020)

5. Ali, M.C., Islam, K., Chung, S., et al.: A study of green human resources management (GHRM) and green creativity for human resources professionals. Int. J. Bus. Manag. Future **4**(2), 57–67 (2020)
6. Zaichenko, Y.O.: The current state of human resources management at enterprises: challenges and prospects. Bus. Inform **11**(514), 436–441 (2020)
7. Krpk, G.: The concept of employee's happiness in human resources management. Bus. Manag. Stud. Int. J. **8**(3), 2750–2775 (2020)
8. Paun, B., Raevi, I., Raketi, O.: IT solutions for human resources management. Serbian J. Eng. Manag. **4**(1), 15–21 (2019)
9. Parsehyan, B.G.: An overview of organizational behavior within the framework of human resources management in the art institutions. Turk. Online J. Des. Art Commun. **9**(2), 104–110 (2019)
10. Tahiri, A., Kovai, I., Krasniqi, A.: Human resource management, performance management and employee performance appraisal by SME managers in Kosovo. Int. J. Econ. Bus. Adm. **VIII**(Issue 4), pp. 288–298 (2020)
11. Huralska, V., Sharkova, H., Skrypnyk, N.: Adaptation and development of human resources as an instrument of increasing the competitiveness of the organization. Econ. Finances Law (6/2), 21–23 (2021)
12. Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. Mob. Netw. Appl. **24**(1), 5–17 (2018). https://doi.org/10.1007/s11036-018-1134-8
13. Liu, S., Bai, W., Liu, G., et al.: Parallel fractal compression method for big video data. Complexity **2018**, 2016976 (2018)
14. Zhao, N., Cheng, Y.Q., Liu, Z.H., et al.: Deep reinforcement learning-based channel intelligent access method and NS3 simulation. Comput. Simul. **385**, 5 (2021)