Recommendation Method of Ideological and Political Mobile Teaching Resources Based on Deep Reinforcement Learning

Yonghua Wang(✉)

Sanya Aviation and Tourism College, Sanya 572000, China
wangyonghua00060@163.com

Abstract. In order to improve the quality of ideological and political education and achieve the goal of effective management of mass mobile teaching resources, this paper puts forward a recommendation method of ideological and political mobile teaching resources based on deep reinforcement learning. Firstly, based on the theory of deep reinforcement learning, the recommendation model of ideological and political mobile teaching resources is constructed, and the recommendation method of ideological and political mobile teaching resources is extracted effectively.

Keywords: Intensive learning · Ideological and political mobile teaching · Mixed resources recommendation

1 Introduction

The mixed resources of ideological and political education are the digital resources used to support the teaching activities of ideological and political education based on information technology. With the development of Internet and Big Data technology, the access to learning resources is becoming more diversified. Different learners can choose appropriate learning resources according to their preferences to carry out learning and achieve personalized learning goals. Many technologies are widely used in mobile ideological and political learning [1]. The common recommendation algorithms of ideological and political mobile teaching resources include content-based recommendation, in-depth reinforcement recommendation and mixed recommendation, etc. Content-based recommendation algorithm constructs learner feature model and resource feature model by recognizing and extracting resource content feature, and recommends learning resource to learners. Deep reinforcement learning algorithm divides learners into groups based on different preferences by mining their preferences, and recommends similar learning resources to each group [2]. Deep learning method has strong learning ability, wide coverage and good adaptability. The neural network of deep learning has many layers and wide width, which can be mapped to any function in theory and can solve very complicated problems. Deep learning is highly dependent on data, and the larger the

W. Fu and L. Yun (Eds.): ADHIP 2022, LNICST 469, pp. 257–272, 2023.
https://doi.org/10.1007/978-3-031-28867-8_19

data, the better the performance. Deep learning can use TensorFlow, Pytorch and other frameworks. Therefore, the application of deep learning can effectively improve the recommendation effect of mixed resources in ideological and political mobile teaching.

Deep reinforcement learning algorithm can effectively reduce the complexity of model building, but there are some problems such as sparse matrix and cold start. Based on the introduction of the theory of heat conduction and material diffusion, such as Liu Zhongbao and so on, a bipartite graph method is proposed to recommend learning resources to learners. Generally speaking, the rationality and scientificity of resource recommendation model is always a difficult problem for traditional recommendation algorithms to be applied to the field of hybrid learning resource recommendation [3]. Therefore, this paper proposes a recommendation method of ideological and political mobile teaching resources based on deep reinforcement learning.

2 Recommendation of Ideological and Political Mobile Teaching Mixed Resources

2.1 Mobile Ideological and Political Learning Content Extraction

The realization of blended learning resources recommendation is essentially an analysis of the relationship between learners and learning resources. In order to provide personalized resource recommendation service, the key is to collect the original data of the learning platform, analyze and mine the data effectively, and finally recommend appropriate learning resources to the learners [4]. The hybrid learning resource recommendation model can utilize the historical aggregation information of learners' learning resources. This information can be represented by the mxn matrix on the left side of the graph, where R represents the learning resource and L represents the learner. The shaded portion indicates the learner's learning resources. The white blank space indicates the unlearned resources. The problem to be solved is how to realize the recommendation of mixed learning resources through this historical information matrix, that is, to obtain the recommended resources that meet the learners' needs from the new learning resources. The principles of recommendation of mixed resources for mobile teaching are shown as follows (Fig. 1):

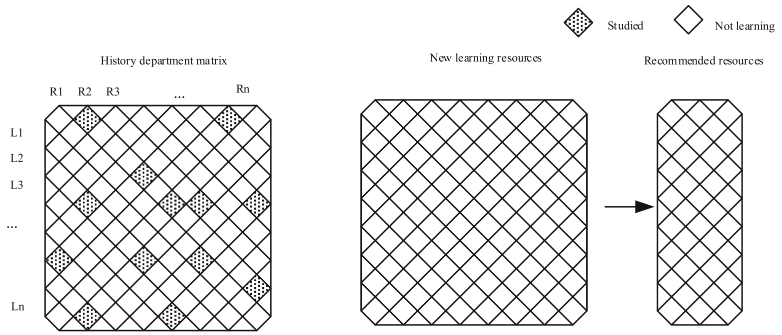

Fig. 1. Recommendation Principles of Mobile Teaching Mix Resources

The personalized recommendation method based on deep reinforcement learning can be summed up in two processes: model training process and resource recommendation proces. The model training process includes data processing of learning platform, algorithm design and so on, and the recommendation process is produced by the running of the recommendation model. The flow chart of teaching resource recommendation is shown in Fig. 2.

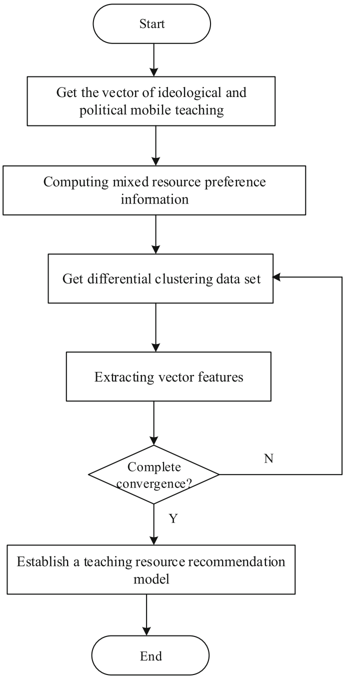

Fig. 2. Teaching resource recommendation process

The basic principle of the DSLL algorithm is as follows: a learner -learning resources scoring matrix is formed according to the learner's historical learning records, and the similarity measure method is used to mine the learner set similar to the target learner or the learning resources similar to the learner's historical preference by calculating the similarity between the learner and the learning resources, then a "neighbor" is formed based on the "neighbor" learner's scoring information to predict the target learner's predictive scoring value for the learning resources, and personalized recommendation is made for the target learner. If we strengthen the recommendation of deep learning, it will recommend the first N learning resources with the largest predictive score to the target learners. The specific steps of the content extraction method for ideological and political learning based on deep reinforcement learning are as follows (Fig. 3):

Mobile thought politics learning data contains learners' learning behavior records, there are also many implicit data, learning resource characteristics can also be obtained in the data [5]. There are irrelevant features or redundant features in the actual data

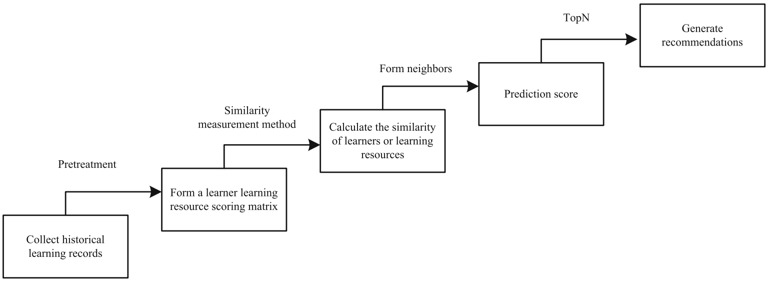

Fig. 3. Extraction steps of mobile ideological and political learning content

processing process, such as the student number or class number of a learner may be irrelevant features, and the home town, parent occupation, home address, etc. of the learner may also be recorded in some large open data sets, so it is necessary to screen or select features, and remove irrelevant features or similar redundant features to make the recommendation model more accurate and efficient [6]. It is necessary to select the features or data information which are closely related to the category in the practice of Deep Reinforcement Learning, and the feature selection method can be used. Generally, the performance of training data can be directly used to evaluate the characteristics, which is independent of subsequent algorithms and is fast [7]. Interactive information refers to the strength of the association between two random attributes or features. Judging the correlation between a single feature and the target category can reduce the redundancy of feature dimension. The method of feature selection based on mutual information (deep reinforcement learning) is selected, and the picture is an example of feature selection model based on deep reinforcement learning (Fig. 4).

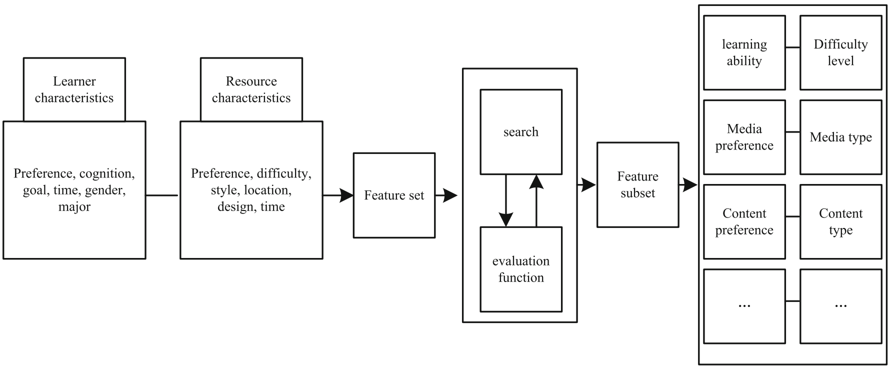

Fig. 4. Example of feature selection based on deep reinforcement learning

The Deep Learning Model consists of network nodes with multiple hidden layers. Networks with multiple hidden layers have powerful feature representation ability, and use multi-level nonlinear structure to abstract low-level features into high-level representation, so as to discover the internal relationship between data. The biggest difference between deep learning and multi-layer perceptron is that multi-layer perceptron needs to select feature input network manually. Deep learning can learn features autonomously, mine implicit feature representation from data, and depict intrinsic information of data. Deep learning generally constitutes a greedy hierarchical approach, with continuous learning and integration from the underlying inputs, and selection of effective features to improve the end result performance. For supervised learning tasks, deep learning analysis is similar to principal component analysis in that it converts data into compact intermediate classes, constructs hierarchical results, and eliminates redundant information. At present, deep learning has been applied in image recognition, intelligent speech, unmanned speech, natural language processing, medical health, etc.

2.2 Evaluation Algorithm of Ideological and Political Mobile Teaching Mixed Resources

Mixed-resource recommendation is one of the indispensable components of personalized learning, and it is also the key technology to realize personalized learning. Common personalized learning recommendation mainly realizes personalized recommendation process by establishing learner model, recommending algorithm processing and recommending result outputting. So, the structure of personalized learning recommendations, like this (Fig. 5).

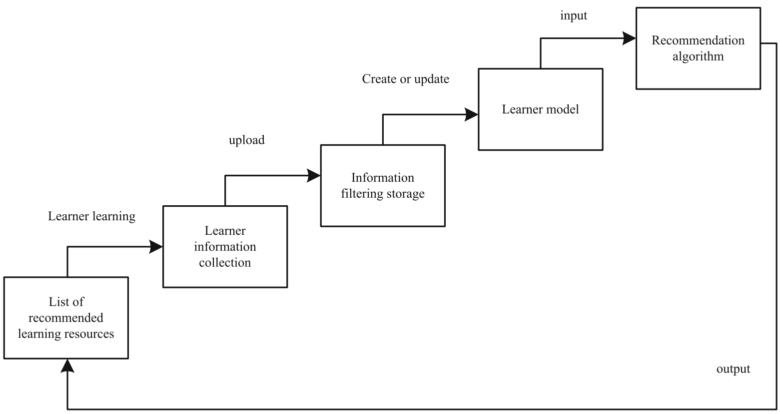

Fig. 5. Schematic of recommended structure for mixed resources

The recommendation algorithm of Deep Reinforcement Learning provides the core function support for recommendation, and the quality of recommendation results will directly affect the performance evaluation of recommendation. Based on the introduction of the main recommendation algorithms above, the advantages and disadvantages of these common recommendation techniques are compared as follows (Table 1).

The trust between learners in social network is difficult to be obtained by explicit way, because it takes the learners' time to input the trust of friends initiatively, and it can

Table 1. Common Recommended Techniques Comparison Table

Recommended technology	Advantage	Shortcoming
Recommendation algorithm based on Popularity	It is simple and easy, and there is no problem of cold start	Unable to make personalized recommendation
Collaborative filtering recommendation algorithm	It is simple and easy to operate with high recommendation accuracy	Rely too much on users' scores; cold boot; Low project coverage
Content based recommendation algorithm	No cold start problem	The recommended content is relatively simple
Model based recommendation algorithm	Fast and accurate, especially for businesses with high real-time performance	The model needs to be maintained frequently

not be replaced by fixed value because the trust between learners is in dynamic change. To some extent, the interaction between learners and friends in social networks reflects the trust relationship between learners and friends, which mainly includes comments and forwarding. The more times the learners comment on each other's messages, the more familiar they are with each other, and the more times the learners forward each other's messages, the more they agree with each other and trust each other. As the complexity of recommendation becomes higher and higher, and the single recommendation technology has different defects, which affects the performance of recommendation. Therefore, the majority of recommendations are using hybrid recommendation technology, according to different application scenarios, using two or more than two kinds of recommendation algorithm, select the appropriate mix strategy, the formation of hybrid recommendation technology. This technology can take advantage of each algorithm, avoid the shortcomings of each algorithm, so as to provide more efficient personalized recommendation services. According to the research, there are seven combinations of hybrid recommendation algorithms, such as (Table 2).

Table 2. Different hybrid recommendation strategies

Mixed strategy	Main advantages
Weighted mixing	The recommendation results obtained by all the recommendation algorithms used in the system are weighted and divided to obtain the final and optimal recommendation results, and then recommended to users
Switch mixing	When the recommendation system adopts the recommendation algorithm, it selects different recommendation algorithms according to different application scenarios and different users

(*continued*)

Table 2. (*continued*)

Mixed strategy	Main advantages
Cross mixing	In the recommendation system, the recommendation results obtained by various recommendation algorithms are mixed for recommendation
Feature combination	In the process of recommendation, the features of different data sources are combined to form vector features, and then the similarity is calculated and recommended
Feature Augmentation	The feature model of a recommendation algorithm is input into other recommendation algorithms as features to improve the recommendation performance
Graded mixing	The model constructed by one or more recommendation algorithms is used as the input of other recommendation algorithms, and then the recommendation results are generated
Series mixing	Multiple recommendation algorithms are used in series, that is, the later ones further screen and optimize the previous recommendation results to obtain the final recommendation results

In fact, learners usually comment on messages that they are interested in or unclear, but do not necessarily forward them, and the messages they forward are often self-selected and highly identified, so the weight of forwarding should be slightly greater than that of the comments. Inspired by the scholar Hu Xun and others on the method of computing the trust between mobile users, the trust between learners in the network environment can adopt the following formula

$$t_{u_1,v} = \frac{c_{u_1,v}}{\max c_{u_1,u}} + (1-\alpha) \times \max f_{u_1,u} \tag{1}$$

In the formula, $f_{u_1,u}$ represents the trust between the learner u1 and the learner u; α represents the set of learners who interact with the learner, and $c_{u_1,v}$ the number of comments the learner has made on the message published by learner V. Since the forwarding behavior should be slightly more weighted than the commenting behavior, the $t_{u_1,v}$ value is 0.4. In the feature selection method based on deep reinforcement learning, information metric evaluation function is very important. Although the functions are various in form, the aim is to select the subset of features that are most relevant to the category, which is recorded as $g(C, f, S)$. The generalized information metrics evaluation function can be expressed as

$$J(f) = AB - g(C, f, S) - \delta t_{u_1,v} \tag{2}$$

Among them, A is the selected feature, δ is the candidate feature, B is the category, and the function is the information between C, F and S, that is, the correlation between the candidate feature and the category, β is the adjustment coefficient, which is used to adjust the information brought by the addition of W, and x is the punishment factor.

The simplest and most intuitive information metrics evaluation function can therefore be expressed as follows:

$$T = g(C,f) - \beta \sum_{s \in S} Wx - J(f) \tag{3}$$

The selected features indicate other features that will affect the learner's choice of resources, such as the knowledge content of the resources, the length of time of study, etc. Candidate features represent other features that are not known for the time being, such as the learner's age, professional background, gender, etc. The category is the extraction of the selected features, which is used to measure the relevance between the selected features and the candidate features. Finally, the evaluation function is constructed to judge the influence of candidate features on the evaluation results, and to filter out some redundant features so as to reduce the workload of deep learning training.

2.3 Realization of Mixed Resource Recommendation of Ideological and Political Education

Ideological and political education resources recommend to provide learners with a learning environment in line with their individual characteristics by analyzing the differences among learners. The meaning of personalized learning mainly includes the following aspects: First of all, personalized learning refers to taking flexible, appropriate and pertinent teaching resources, teaching strategies, teaching evaluation and other learning services according to the students' learning ability, learning style, learning attitude and other learning characteristics, so that students can get all-round development and progress. Secondly, the learning process advocated by personalized learning is to help students find their own strengths, highlight their own personality, the pursuit of self-realization. Finally, personalized learning embodies the essential difference between learning and education: learning is through the acquisition of knowledge and skills to achieve self-improvement. Education is to teach students some knowledge and skills, in accordance with the training objectives step-by-step training of students, to a certain extent, buried in the initiative and creativity of learners. In the study of recommendation of learning resources, the classical deep reinforcement learning algorithm can effectively train the history learning data, but the traditional deep reinforcement learning algorithm or simple network model can not meet the actual needs, and can not guarantee the convergence to an optimal solution. This paper designs a deep reinforcement learning model (see figure) to judge whether or not the learner learns a certain learning resource or how much attention is paid to the learning resource (Fig. 6).

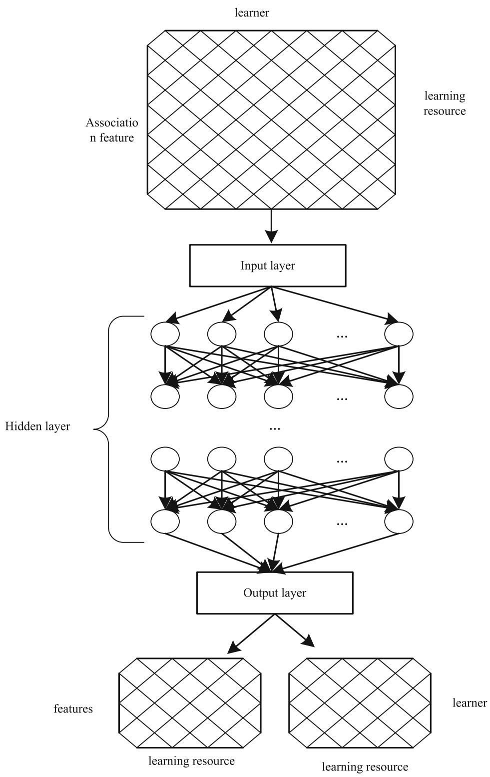

Fig. 6. Deep reinforcement learning model

The key of applying DSLL to the recommending scene of learning resources lies in modeling the history learning records of learners, mining the implicit features of the original data, and then standardizing the input layer and output layer in constructing the training model. Above, feature selection model based on DLL and learner-resource bipartite graph association model effectively solve the input and output of DLL. In the model training of deep reinforcement learning, a large number of transformations (s, a, r) will be obtained by interaction with the environment, but updating the parameter value of Q network according to each transformation will make the network fall into local optimization, and then forget the previous learning experience. At the same time, it takes a long time for the network to back-propagate and update the parameters, and the training efficiency of the whole model becomes low. So, in the deep reinforcement learning algorithm, experiential playback technology is used to solve this problem to break the dependence of time sequence and avoid the local optimization of the network. At the same time, it can accelerate the training by random microbatch data, break the similarities between samples, and accord with the learning and improving behavior of human using past experience. The use of experiential playback simplifies the debugging and testing of the algorithm, and makes the training task of the network model more similar to the common supervised learning mode.

In the model training of deep reinforcement learning, a large number of transformations (s, a, r) will be obtained through interaction with the environment. However, if the parameter values of Q network are updated according to each change, the network will be trapped in local optimization, and the previous learning experience will be forgotten. At the same time, it takes a long time for the network to back-propagate and update the parameters, and the training efficiency of the whole model becomes low. So, in the deep reinforcement learning algorithm, we use the experience replay technology to solve this problem, use the experience replay mechanism to break the time dependence, avoid the network falling into the local optimum, and can accelerate the training through the random microbatch data, break the similarity between the samples, at the same time, it accords with the human's learning and improving behavior using the past experience. The use of experiential playback simplifies the debugging and testing of the algorithm, and makes the training task of the network model more similar to the common supervised learning mode. The transformation is stored in the experience playback pool as experience, and a random number of transformations are randomly sampled from the experience playback pool for each run to form a random microbatch data for training the network (Fig. 7).

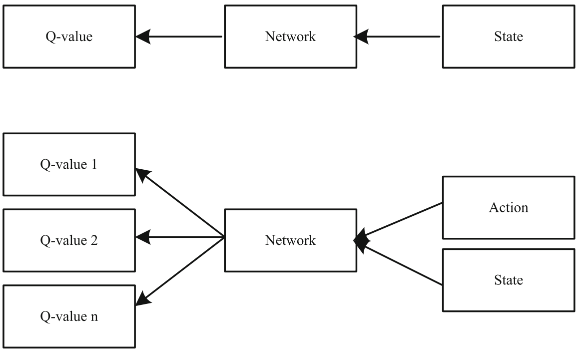

Fig. 7. Training network of teaching resources

Resource management subsystem is the management and maintenance part of adaptive learning. It is mainly used by teachers, educational administration and administrators. Specifically, teacher users shall realize the management of addition, deletion and check of curriculum related resources (courses, courseware, examinations, etc.); educational users shall realize the management of addition, deletion and check of curriculum resources and announcement information; and administrative users shall realize the management of addition, deletion and check of resources (users, logs, data, etc.). The resource management subprocess, shown in Fig. 8.

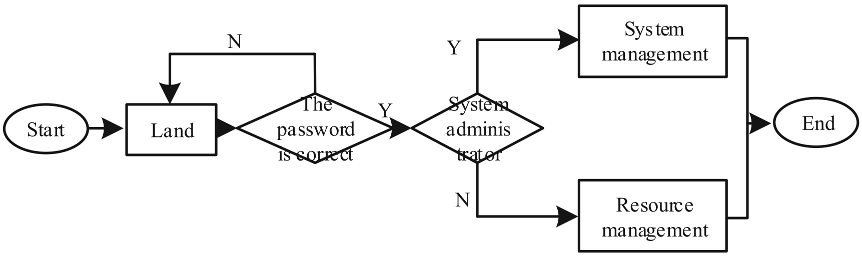

Fig. 8. Resource Management Subflowchart

For the users who use the resource management subsystem, the login verification of user account and password is carried out first, and then the identity is determined after the success. For the administrators, the management of all resources can be realized, while for other users, the corresponding part of the resource management is carried out according to the identity. Ideally the algorithm should be able to give moderately difficult resources as reflected in the answer rate that eventually converges to a fixed number. Let the learner do the resource search item N each time, the number of correct answers N each time, the construction deviation measurement index is.

For the users who use the resource management subsystem, the login verification of user account and password is carried out first, and then the identity is determined after the success. For the administrators, the management of all resources can be realized, while for other users, the corresponding part of the resource management is carried out according to the identity. Ideally the algorithm should be able to give moderately difficult resources as reflected in the answer rate that eventually converges to a fixed number. Let the learner do the resource search item N each time, the number of correct answers N each time, the construction deviation measurement index is

$$err = \left(\frac{n}{N} - \delta\right)^2 \tag{4}$$

The actual experiment looked at the mean of recommended deviations per turn (epoch or designated segment). Deep reinforcement learning is a learning method that combines deep learning and reinforcement learning to solve the problems in control and decision field and to realize one-to-one correspondence from perception to action. Deep learning is used to analyze environmental information and extract features from it, and reinforcement learning is used to further analyze environmental features and select corresponding actions to achieve target return. For the complex decision making problem of learner's cut-set selection strategy, the deep reinforcement learning method is used to extract effective information from the learner's historical query data environment, and the decision-making control is realized by combining the environmental information and the learner's response. The model is constructed as shown in Fig. 9.

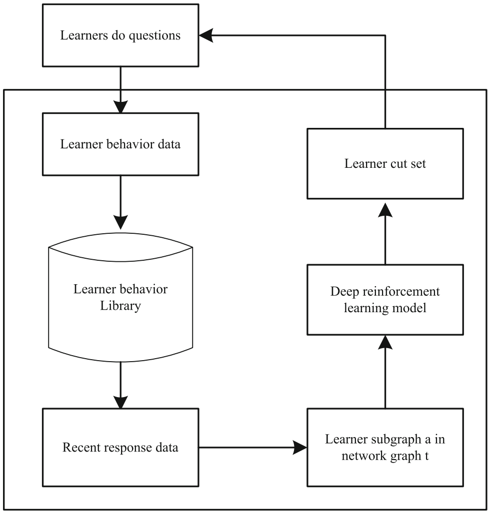

Fig. 9. Guided resource recommendation model based on deep reinforcement learning

With the development of "Internet Plus Education", the research on personalized learning recommendation algorithm in the field of education has become a hot spot in recent years. However, the current recommendation algorithms in the field of education rely more on the traditional linear recommendation methods, not taking into account the relevance between knowledge points and the important characteristics of the subject. At the same time, the current personalized learning recommendation algorithm can not recommend moderate difficulty for users, facing the recommendation efficiency is not high, the user answer rate is not stable. Due to the complexity of learners' learning ability level and the variability caused by the learning process, the processing ability and learning ability of deep reinforcement learning in large state space provide a new direction for learners to recommend learning resources.

3 Analysis of Experimental Results

In order to verify the conformity of personalized learning resources recommended by hybrid learning resources recommendation method based on deep reinforcement learning with learners' needs, a series of experiments were conducted. The experimental data includes not only the data of learning resources, but also the historical data of learners' learning. In the existing public datasets, such as edX, World Uc, and others, dozens of attributes are provided, including course data, learner information, and learner behavior data. Form the experimental data set. Guided resource recommendation model based on deep reinforcement learning algorithm doesn't know how to select learning area for users at the beginning of setting questions for users, and the parameters of deep reinforcement learning network are randomly initialized according to Gaussian distribution. Therefore, the selection of topics of deep reinforcement learning model at the beginning is also random. The model strengthens the ability of selecting test strategy in the process of giving test questions to users, updates the weight value in the network according to the return value returned from the user's historical data, and learns to choose the strategy of learning area questions for users after repeated training. This experiment sets 20000 rounds of training for each user, each model gives 10 questions to the user. The model parameter settings in the experiment are shown in the Table 3.

Table 3. Experimental parameter settings

Parameter	Size/content	Describe
Q	20000	Number of experiment rounds per user
E	110	Take the current status and answer the previous question
T	Comprehensive difficulty value	Comprehensive difficulty value of the first e question
Y	12	Number of topics selected each time
K	5	Number of topics with the highest ranking
B	6	Number of topics in the middle
L	5	Number of topics with the lowest ranking
M	0.9	Optimal Recommendation Index
P	5	Return value of err = 0
U	0	Return value at the end of turn

The learner -resource features, i.e., the subset of features to be input in the whole process, can be obtained by using the model processed by feature selection method based on deep reinforcement learning. Many deep reinforcement learning tasks have the characteristics of table discretization, as shown in the table. For example, a study record shows that the study resources belong to the computer category, the difficulty is easy, the media type is video format, and the learners study at 9: 00 AM. Description of resource association characteristics and their values are (Table 4):

Table 4. Part Description and Numerical Value of Resource Association Characteristics

Features	Numerical representation	Significance
Subject attribution/preference	1, 2, 3, …, 9	Including computer, economic management, literature and history, life science, art and design, etc.
Media type/preference	1, 2, 3, 4, 5, 6	Including video, audio, text, pictures and slides

(*continued*)

Table 4. (*continued*)

Features	Numerical representation	Significance
Difficulty level	1, 2, 3, 4, 6	Including easy, easy, moderate, difficult and difficult
Content type/preference	1, 2, 3, 4, 5	It includes concept explanation, test questions, cases, introduction and course review
Study time	1, 2, 3, 4, 5	Starting from the morning, every 5 h is a time period
Equipment terminal system	1, 2, 3, 4	ISO,Android,Windows

Accuracy refers to the proportion of the resources recommended to the learner's interest in the total resources recommended to the learner, while recall refers to the proportion of the resources recommended to the learner's interest in the total resources. The recommended list of recommendations is N = 5, 10, 15, 20, 25, 30, 35, 40, respectively. The recommended accuracy and recall rates of the traditional methods and the proposed methods are measured, as shown in the Fig. 10.

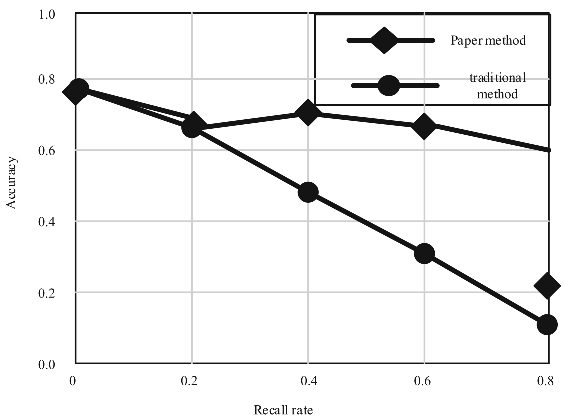

Fig. 10. Accuracy and recall of the two methods in different cases

In this paper, a scoring matrix of learners' learning resources is established according to learners' historical learning records, and the similarity measurement method is used to mine the target learners of similar learners. Therefore, the recall rate of this method is high (Fig. 11).

Because this method adopts hybrid recommendation technology, it can improve the mobile resource integration method according to different application scenarios, so it can effectively avoid the shortcomings of the algorithm and provide efficient personalized recommendation service. Therefore, the method of this paper carries out the coverage

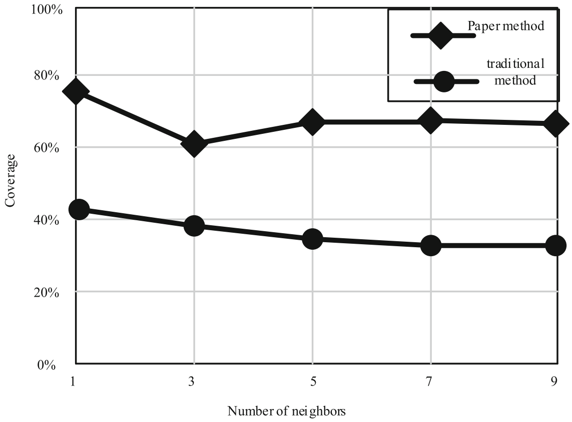

Fig. 11. Coverage of two methods with different number of neighbors

experiment under the condition of 9 neighbors. Experimental results show that this method is superior to traditional methods.

It can be seen that in the same environment, the accuracy and recall rate of the proposed method are higher than those of the traditional method, which shows that the proposed method has better performance and can recommend more accurate and more interesting learning resources for learners.

4 Conclusion

With the rapid development of "Internet Plus" education, the scale of mobile ideological and political learning resources expands rapidly, which makes it more difficult for learners to choose suitable resources. How to help learners acquire appropriate learning resources to carry out personalized learning has become a major research topic of intelligent learning. The key to realize the recommendation of mixed learning resources is to explore and mine the value of data application of mobile ideological and political learning platform. This method is superior to the traditional deep reinforcement learning algorithm in classification and regression ability evaluation index, which shows that it can provide better recommendation service of mixed learning resources in big data environment.

Fund Project. 1. Project of teaching reform of Sanya Aviation and Tourism College in 2020: On the teaching mode of "micro ideological politics" in Colleges and Universities(Project No.:SATC2020JG-12)

2. General topics of Educational Science Planning in Hainan Province: Investigation and Research on labor values of Higher Vocational College Students in the new era (Project No.: QJY20201018). Funded project for the construction of "double leaders" teachers' Party branch secretary studio in Colleges and universities in Hainan Province.

References

1. Jing, K.L., Si-, F., Ya-fu, Z.: Model predictive control of the fuel cell cathode system based on state quantity estimation. Comput. Simul. **37**(07), 119–122 (2020)
2. Su, J., Xu, R., Yu, S., et al.: Idle slots skipped mechanism based tag identification algorithm with enhanced collision detection. KSII Trans. Internet Inf. Syst. **14**(5), 2294–2309 (2020)
3. Su, J., Xu, R., Yu, S., et al.: Redundant rule detection for software-defined networking. KSII Trans. Internet Inf. Syst. **14**(6), 2735–2751 (2020)
4. Zhang, H., Chen, L., Tan, L., et al.: Research on co-construction and sharing of virtual simulation experimental teaching resources among provincial colleges and universities. Exp. Technol. Manag. **38**(05), 26–28 (2021)
5. Shi, Y., Zhang, J.: Design of information based teaching resources sharing system based on multimedia technology. Modern Electron. Tech. **44**(20), 32–36 (2021)
6. Luo, P., Yuan, J., Chen, M., et al.: Mean penalty random forest nonstationary time series prediction method for teaching re-sources. J. Chin. Comput. Syst. **42**, 2089–2094 (2021)
7. Sun, Y.: Design of practice teaching network platform for ideological and political teaching resources in colleges and universities. Microcomput. Appl. **36**(02), 152–155 (2020)