Innovative Application of Big Data Technology in Network Teaching Model of University Courses

Changdong Xu(✉) and Song Wang

School of Management, East University of Heilongjiang, Haerbin 150080, China
xuchangdongyouxian@163.com

Abstract. Facing the huge amount of learning resources, how to push the appropriate learning resources to the learners has become a major problem. Therefore, in this context, the application of big data technology has obvious advantages. Based on the above analysis, we will study the big data technology in the university curriculum network teaching model innovation application, to achieve the big data technology expansion network teaching model function goal. After analyzing the current situation of networked teaching model of university courses, the data mining technology is used to obtain user preferences. Through clustering algorithm to build a user image, using collaborative filtering algorithm personalized recommended teaching resources, model optimization is completed. Compared with the teaching models before and after optimization, the accuracy rate of the recommended resources is higher than 95%, the score is higher and the teaching effect is better.

Keywords: Big data · College curriculum · Network teaching · Innovation of teaching model · Data mining · Clustering algorithm

1 Introduction

Network has been input into everyone's work and life. It is an effective combination of network and education to apply network in education and promote the development of network. Colleges and universities not only have abundant computer and network resources, but also use network courses as the carrier of teaching activities, which can enrich teaching methods and ways. With the development of network information technology, curriculum resources will be shared in today's university curriculum network, so that teachers will be able to use these materials as teaching topics, and students will have the opportunity to discuss these materials with each other [1]. Curriculum networking not only inherits the advantages of traditional distance education that is not limited by time, space and place, but also overcomes the shortcomings of traditional distance education that lacks of communication and interaction. It can provide learners with rich and colorful human-machine interface with pictures and texts, and provide a large scale

W. Fu and G. Sun (Eds.): eLEOT 2022, LNICST 453, pp. 578–588, 2022.
https://doi.org/10.1007/978-3-031-21161-4_44

of knowledge base and information base, which can stimulate learners' interest in learning, so that learners can take the initiative to build knowledge, and achieve their ideal goal of acquiring knowledge, self-renewal and even innovation. Therefore, it is of great significance to continuously innovate and optimize the networked teaching model for improving the quality of networked teaching. There are many researches on networked teaching model in foreign countries. In the process of implementing network education, Germany pays attention to the development and utilization of educational information resources and devotes its eyes to the all-around development mode of the whole people, not only for schools, but also for families and society. The French government has a very early policy of applying information technology widely to education. Most universities in the world are competing to develop online courses for networked teaching. Reference [2] An online teaching model of program design experiment course is designed with the support of deep learning, and the distance experiment course is realized. But this model needs a lot of teaching resources and cannot make full use of them when teaching online.

WEI et al. [3] proposed a multi-source multi-modal large data retrieval method based on Mapreduce. Aiming at semantic congruence discrimination, hash and dictionary learning are introduced to construct objective evaluation. In order to guarantee the approximation of feature pairs in low dimensional space, matrix decomposition combined with intermediate variables is adopted to transform the objective evaluation into a non-convex problem. The hierarchical structure is designed as the node adjacency matrix, and the multimodal retrieval results are calculated according to the approximation. ZHANG et al. [4] constructed an intelligent recommendation model of big data learning resources. The model consists of six parts: data source module, learner analysis module, learning resource analysis module. The model not only pays attention to the analysis of big data on the level of knowledge, learning behavior, learning style and learning interest of learners, but also to the analysis of big data on the attributes, types, efficacy and evolution of learning resources. It also pays attention to the collaborative work of learning resources retrieval, matching agent, management agent, algorithm optimization agent and recommendation agent. The model has the characteristics of adaptability, personalization and usability. Although the above methods have achieved good results at this stage, but the depth of the application of big data is not enough to push the appropriate learning resources to learners.

Networked teaching model is implemented by using computer technology, new media means and various educational resources. The application of networked teaching model provides a large number of learning materials and teaching resources for students, but how to use these resources effectively is a new problem. With the continuous development and maturity of information technology, the big data era of university curriculum networking has come. Using big data technology to optimize the networked teaching model of college courses can effectively improve the performance of teaching model and achieve better teaching results. Based on the above analysis, in order to study the big data technology in the university curriculum network teaching model, this paper will use big data technology to optimize the university curriculum network teaching model. To sum up, the arrival of big data era, strongly promoting the development of education information, big data will change the face of traditional computer education. Compared with the existing recommendation and processing methods of network teaching resources, the

accuracy rate of the proposed method is higher than 95%, and the application advantage is obvious.

2 Analysis on the Current Situation of Network Teaching Model of Courses in Colleges and Universities

At present, the networked teaching model of college courses mainly has the following defects:

(1) The application of networked teaching model provides a large number of learning materials and teaching resources for students, but the application of networked teaching model is a kind of distance education. Although it can provide various teaching videos, the online Q&A link and classroom interaction link always flow into formalism, which cannot achieve the real interactive effect in traditional teaching and cannot fully mobilize students' enthusiasm. Moreover, the networked teaching model cannot create an attractive interactive environment in the teacher-student interaction, and cannot give full play to the teacher-student relationship in teaching and learning. Each student's intelligence has certain difference, the network teaching model application cannot achieve truly pays attention to each student, cannot solve the different stratification promptly sensitively miscellaneous disease, the student often is in "the passive class" the condition.

(2) The application of networked teaching model focuses on theoretical guidance and lacks practical experience. Because the students who use the network teaching model come from different places and their situation is different, it is difficult to achieve complete unity of practice teaching. In the long run, students no longer pay attention to the real meaning of practice and lack interest in practice teaching, which goes against the educational idea of combining theory with practice. Nowadays, the application of networked teaching model cannot provide personalized practical teaching according to the differences of different students, resulting in the lack of practical experience.

(3) The network teaching model of college courses still has the shortcomings of inadequate application of teaching resources and the slow updating of resources and high redundancy in the teaching model. At present, the providers or instructors of online courses in colleges and universities can be divided into three categories: in-school teachers, out-of-school teachers and those willing to provide shared resources. Under circumstances. In terms of the rigor of educational communication, open courses are offered by highly respected or well-established teachers in a particular subject area: first, by being accountable to the students; and second, by making the course itself more attractive [5]. But at the same time, the choice of courses will be greatly limited, but also makes the course more professional. As a result, the number of people who have heard of such courses has increased. In addition, in the process of network courses in colleges and universities, teachers mainly know the excellent courses through the links in the school websites, search engines and recommendations of colleagues; students mainly know the courses through the links in the school websites and teachers. This shows that colleges and universities ignore

the promotion and utilization of quality curriculum, which makes the actual role of quality curriculum has not been played accordingly.

In view of the defects of the current college curriculum network teaching model analyzed above, this paper will optimize the network teaching model by using big data technology.

3 Research on the Innovative Application of Big Data Technology in College Curriculum Network Teaching Model

3.1 User Preference Acquisition of Course Network Teaching Model

Explicit mining and implicit mining can be used in learning preference mining. Explicit retrieval, i. e. direct acquisition of user learning preference information. This approach requires direct user participation, and it is done by the learner registering information or filling out a preference form. Implicit mining, that is, by mining the historical behavior of students browsing the Web to obtain students' learning preferences.

According to the students' browsing behavior and content, we need to process the data before constructing the students' learning preference model. Preprocessing of data for data cleansing, data integration, data specification and data transformation as the scale of user behavior data is growing much faster than the growth of computing power. For user behavior data, the main preprocessing techniques are: similarity measure, SVM, association rules extraction, main components analysis, information gain and so on.

There are many methods to measure similarity. Euclidean distance is the most common one:

$$ED(a,b) = \left(\sum_{i=1}^{n}(a_i - b_i)^2\right)^{1/2} \tag{1}$$

Among them, n represents the number of attributes that dimension, a_i and b_i data a and b respectively, the i-component that attribute value. Another common method is to compute the Pearson correlation coefficients between item feature vectors to measure the similarity between items. The Pearson correlation coefficient is calculated as follows:

$$P(a,b) = \frac{\sum(a,b)}{C_a \times C_b} \tag{2}$$

Among them, C_a and C_b represent the standard deviation of a and b respectively, and $C_a \times C_b$ calculates the covariance between a and b. In practice, cosine similarity or Pearson correlation coefficient is the most common measure of similarity in data analysis systems. The formula for calculating the cosine similarity is as follows:

$$\cos(a,b) = a \cdot b(\|a\|\|b\|)^{-1} \tag{3}$$

Here, $a \cdot b$ represents the dot product between two vectors, and $\|\|$ is the length of the vector.

The sparsity of data is one of the common problems in the field of big data. Sparsity of data is the collection of data set, there are a large number of elements value of zero data, these zero-valued data sometimes even more than the number of valid data. If these invalid data are not processed, the huge computational resources in the training and detection process will be useless. Data dimension reduction technology provides a good solution to the problem of data sparsity. In this paper, the singular value decomposition method is used to reduce the dimension of data. In real life, most of the data matrices we deal with are not square matrices. In this paper, we assume that there are a total of N users of networked course teaching model, each user participated in the M courses, and the matrix $N \times M$ formed is not a square matrix. For ease of study, the singular value decomposition technique provides a decomposition method:

$$J = Q \sum W^T \tag{4}$$

where, J is a matrix of $N \times M$; Q represents a square matrix of $N \times M$ (where the vector is orthogonal, and the vector in Q is called a singular vector), $\sum$ represents a matrix related to $N \times M$ (except the diagonal, other elements are 0, and the elements on the diagonal are called singular values), W^T (transpose matrix of W) is a matrix of moments of $N \times N$, where is an orthogonal vector, and the vector in W is called a right singular vector. Singular values are similar to eigenvalues. In a matrix, Singular values are arranged in descending order, and they decrease very rapidly. In most cases, the sum of the first 10% or less Singular values accounts for more than 99% of the total Singular values. So, in most studies, only the top 10% of the eigenvalues need to be studied to get good results. In the process of user data behavior analysis, the collected data can be missing, abnormal and so on. Gaussian transform is used to reduce the noise in the data.

3.2 Get Model User Preferences

First of all, the long-term preference data is mined by judging the type of students by the data accumulated in the teaching model. Second, students' short-term learning preferences are implicitly mined by tracking and observing their browsing behavior (including cookies, web server logs, and favorites). Finally, students' learning preferences are updated by the information of learning resources and feedback. Based on this, a model framework of student learning preferences is established, as shown in Fig. 1.

Educational big data, both broadly and narrowly, has four dimensions, namely, 4V: Volume, Velocity, Variety, and Value. There are 8 dimensions in the big data, such as learning time, forum posting, real-time discussion and homework. Each dimension reveals some information of the students, which can be used as the basis for teachers to evaluate students.

Study time shows students' enthusiasm for learning, students' attitude towards learning and their efforts in learning. Forum response shows students' initiative, whether there is any positive thinking, the degree of interaction and participation among students, the degree of harmony among students. Homework can reflect students' attitude towards learning, the degree of mastery of classroom knowledge, whether they listen carefully in class, the initiative of learning and so on.

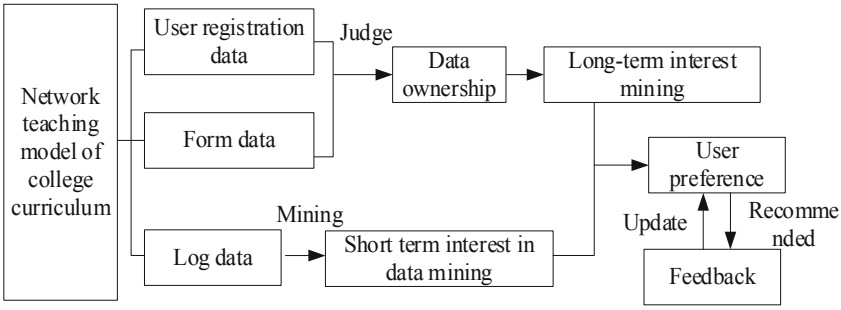

Fig. 1. Framework of student learning preference model

Pearson coefficient is used to calculate the similarity between teaching resources. The average score in the traditional Pearson coefficient formula can reduce the recommendation error of the algorithm. The user preference matrix is established to call teaching resources from the database quickly and update the user's preference in time. Assume that the user's preference matrix for the resource is RM:

$$\begin{cases} RM = [r_1, r_2, \cdots, r_m]^{\mathrm{T}} \\ r_k = \left[r_{k,1}, r_{k,2}, \cdots, r_{k,m}\right] \end{cases} \tag{5}$$

In formula (2), r_k is the user's preference for teaching resource k, and $r_{k,m}$ is the user m's preference for teaching resource k. According to the similarity of teaching resources calculated according to the following formula, the recommendation degree $r_{q,i}$ of student q to teaching resources i can be calculated according to the following formula [6]:

$$r_{q,i} = \frac{\sum_{k=1}^{m} sim(i,j)p_{q,k}}{m} \tag{6}$$

In the formula, $p_{q,k}$ is the degree of q's preference for k, m is the number of similar resources in the database, and $sim(i,j)$ is the degree of similarity between i and j. Based on the above process, the learning and use preferences of the users of networked teaching model are obtained. Users are clustered according to their preferences when they use the networked course teaching model.

3.3 User Clustering

Nowadays, if we want to promote development, we need to enhance the individualized part of the teaching model. Individualized teaching is a kind of teaching mode to show individuality, which is oriented to students. The new student-centred teaching model should be tailored to each student's individual circumstances (personality preferences, gender and age competencies, physical and mental development and personality traits, etc.) so as to fully tap their potential. Adopt diversified content, flexible structure and diversified forms, adopt appropriate teaching methods, so that every learner's potential can be fully tapped and developed. Therefore, in order to make full use of teaching

model data, it is necessary to cluster model users. The aim is to classify a large number of unknown datasets according to the intrinsic similarity of the datasets, so that the data within the datasets are more similar and the data outside the datasets are smaller.

In this paper, an improved differential evolution algorithm is used to cluster users. Although the standard differential evolution algorithm has better global optimization ability, it also has obvious disadvantages: firstly, the algorithm converges slowly in the later stage of running, so it cannot converge to the optimal solution quickly in less iterations; secondly, the DE algorithm is very sensitive to the values of contraction factor and crossover probability, and the settings of contraction factor and crossover probability are often different for different problems. In order to solve some defects of the standard differential evolution algorithm, this paper improves the algorithm from three aspects: introducing the adaptive adjustment strategy of parameters, speeding up the convergence speed of the algorithm and improving the termination conditions of the algorithm. The flow chart of the concrete steps of the improved differential evolution clustering algorithm (IDE Cluster) is shown in Fig. 2 below [7].

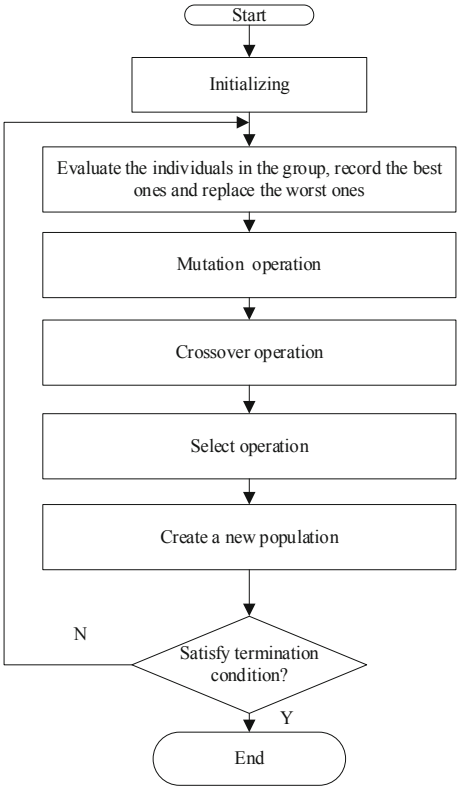

Fig. 2. Flow chart of improved differential evolution clustering algorithm

The specific steps of clustering algorithm are as follows:

Step1: randomly select k different data from the data set to form a chromosome, repeat N times, to get the initial population size of N;

Step2: Evaluate each chromosome in the current population, calculate its fitness function value, and select the individual with the smallest fitness value as the current best individual;

Step3: DE/current-to-best/1/exp approach to mutation and crossover operations to produce intermediate individuals, norm factors and crossover probability by using the formula;

Step4: Evaluate the intermediate individuals, calculate their fitness value, select them, generate a new generation of population, and update the best individuals, and then replace the worst individuals with the best fitness individuals;

Step5: If the current number of iterations plus 1 is set and the termination condition is satisfied (the number of iterations reaches the specified maximum number $Dg_{\max}$ or the improvement value of the objective function is less than the threshold ζ in the continuous $Dg_{\max}$ iterations), switch to Step6, otherwise switch to Step3;

After the model users are clustered, the big data technology is used to construct the user portrait and discover the user needs, so as to optimize the network teaching model.

3.4 Realizing the Optimization of Network Teaching Model of College Courses

User portrait modeling is to standardize and normalize multi-dimensional data of User portrait, to realize unified expression and description, to match user requirement and product function, to realize effective management of resource and requirement, to characterize, normalize and systemize complex data, and to realize effective management of data, in order to ensure user portrait of knowledge and knowledge process. According to the information characteristics of user portrait elements, the following principles shall be followed in user portrait modeling:

(1) Must be able to describe key characteristic information of the main entities in the user's portrait. Because the user portrait information comes from different entities, it is not necessary and impossible to obtain all the entity information completely. Therefore, in the process of building user portrait model, we must be able to select the most relevant user portrait entity accurately, and be able to describe its main characteristics accurately to describe the knowledge user portrait.
(2) The interaction between different user portrait entities must be reflected. Different entities in user's portrait usually interact with each other. Only by establishing the relationship between different entities can the user's portrait description be complete and accurate.
(3) The entity and its features in a user portrait shall be identifiable. The aim of constructing user portrait model is to realize the integration of knowledge and manage the knowledge effectively. So the elements of user portrait must be recognized automatically by computer, or the user can provide explicit information of the elements of user portrait.

According to the above user's clustering processing, determine the user's business scene using the network teaching model of college courses, and then get the user's portrait. In order to optimize the performance of the networked teaching model of university courses, the corresponding teaching resources are recommended according to the user portrait guidance.

User based collaborative filtering algorithm mainly studies the similarity between users. The implementation process is as follows: firstly, measure the similarity between users. If it is necessary to formulate a top-N Course Resource Recommendation sequence for users, calculate the similarity between users or users is the first task to be completed. Through the ranking of similarity between projects, select V courses with the highest similarity, Finally, n items with the highest frequency are selected as the final recommendation results and recommended to the target user and users with high similarity with the user.

Collaborative filtering is the most popular type of collaborative filtering, and the related algorithms cannot be listed. If there are m resources, n users data, only some users and some resources between the score data, the other part of the score is blank, need to use the existing part of sparse data to predict the score relationship of those blank resources, find the highest score of resources, and recommend it to the corresponding users. The following is a probability model algorithm:

$$Y_{x,s} = \sum k \times p\left(Y_{x,s} = 1\right) \Big| Y_{x,s} \tag{7}$$

Among them, $Y_{x,s}$ is the probability that a user with a user portrait label of x will be recommended for resources in Category s, and k is the total number of resources in Category s. Convenience of data collection is the advantage of the strategy of collaborative filtering recommendation, and it can get better recommendation performance over time. This strategy allows researchers to process complex unstructured data without having to understand the relevant domain knowledge and at the same time meet the requirements of personalized recommendations. According to the recommendation result of recommendation algorithm, the teaching model can recommend corresponding resources for users. Thus, the big data technology is used to optimize the teaching model. The above contents are all the research of applying big data technology to the networked teaching model.

4 Experimental Study

4.1 Experimental Contents

In order to verify whether big data technology can effectively optimize the networked teaching model of college courses, experimental research will be carried out. This experiment evaluates the effect of the data technology applied in this paper by comparing the proposed method, the MapReduce based multi-source and multi-modal big data retrieval method proposed in reference [3] and an intelligent recommendation method for learning resources under the big data threshold proposed in reference [4]. Among them, students, teachers and experts are used to evaluate the use of the model. A total of 300 evaluation subjects were invited, which were randomly divided into 6 groups, each group had

weighted evaluation scores. The weight of students is 0.4, and the weight of teachers and experts is 0.3.

4.2 Experimental Result

Table 1 shows a comparison of the accuracy of the recommended resources under different methods.

Table 1. Comparison of accuracy of model recommended resources /%

Group	Reference [3] method	Reference [4] method	Proposed method
1	79.1	80.2	95.8
2	79.3	79.3	96.1
3	76.3	82.6	95.4
4	76.1	80.8	95.3
5	79.6	79.6	95.6
6	77.2	81.5	96.7

Analysis of the data in Table 1 shows that the accuracy rate of teaching model resources recommendation after big data optimization is higher than 95%, obviously better than before big data technology.

Figure 3 shows the evaluation of the model by different evaluation subjects.

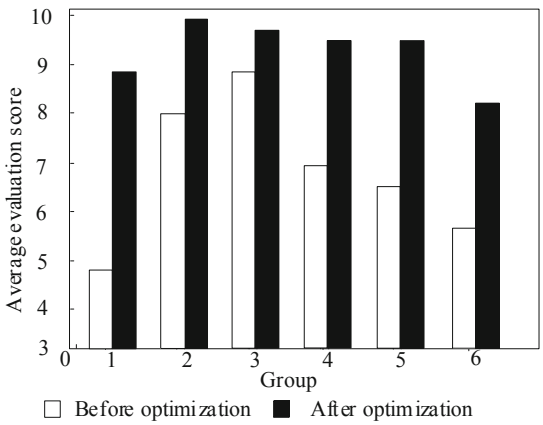

Fig. 3. Average score of evaluation before and after model optimization

As can be seen from Fig. 3, the model score after optimization is much higher than that before optimization. This is because the optimized model makes effective use of

teaching resources and gives good feedback to users. To sum up, the application of big data technology in teaching model can improve the model performance and teaching effect.

5 Conclusion

The networked curriculum in colleges and universities has the advantages of free learning environment, good interaction and cooperation, and has become an important way to cultivate talents and promote scientific research and education. With the continuous development and popularization of network, online teaching will become an irresistible trend in the development of modern curriculum in all aspects of educational research. This paper optimizes the networked teaching model of university courses by using big data technology, and studies the innovative application of big data technology in the networked teaching model of university courses.

Fund Project. 1. General Research Project of Heilongjiang Province's 2020 Higher Education Teaching Reform: Innovative Research on the Training Model of Applied Talents in Private Undergraduate Universities from the Perspective of Big Data (SJGY20200579);

2. Heilongjiang Province's 2020 Higher Education Teaching Reform Key Project: Innovation Research and Practice of Applied Talent Cultivation Model in Transforming Colleges and Universities (SJGZ20200138).

3. Heilongjiang Province Philosophy and Social Science Research Planning Project: Research on the Construction of Applied Undergraduate Colleges in Heilongjiang Province under the "Double First-Class" Strategy (18EDE501);

4. Heilongjiang Province Educational Science "Thirteenth Five-Year Plan" 2020 Key Project: Research on the Innovation and Entrepreneurship Development Mechanism of Application-oriented College Students from the Perspective of Resource Collaboration (GJB1320276).

5. School-level scientific research project of Heilongjiang Oriental College: Research on the construction of agricultural product supply chain ecosystem from the perspective of blockchain innovation coupling (HDFKY210205).

References

1. Zhao, Y., Li, J., Qu, P.: Construction of network, sharing and open basic course platform for electrical specialties. Experimental Technology and Management **36**(10), 225–228 (2019)
2. Tan, M., Yuan, Y., Feng, J., et al.: Online teaching model for programming experiment courses based on deep learning. Research and Exploration in Laboratory **40**(10), 162–165 (2021)
3. Wei, X., Zhao, H.: Research on multi-source and multi-modal big data retrieval method based on MapReduce. Computer Simulation **38**(04), 422–426 (2021)
4. Zhang, J., Ye, Q.: Construction of intelligent recommendation model of learning resources from the perspective of big data. J. Hunan Univ. Sci. Technol. (Social Science Edition) **22**(04), 178–184 (2019)
5. Guo, W., Cui, X., Li, J., et al.: Discussion and practice of surveying online teaching for non-surveying and mapping majors. Bulletin of Surveying and Mapping (09), 151–156 (2021)
6. Zhao, G., Feng, Y.: Teaching reform and innovation of C++ programming course based on the "Internet+" and big data technology. Journal of Tangshan Teachers College **41**(06), 123–126 (2019)
7. Zhai, X.: Research on the connection and innovation of big data technology and database series curriculum system. Wireless Internet Technology **16**(20), 96–97 (2019)