Course Recommendation Method and System of Education Platform Based on Deep Learning

Jingbin Zhang(✉)

Xi'an FanYi University, Xi'an 710061, Shaanxi, China
ruibin321@163.com

Abstract. Course recommendation is a service mode that recommends personalized learning resources to users according to their historical behavior information. Traditional recommendation methods can not effectively use auxiliary information and its potential relationship, resulting in problems such as cold start and sparse data. With the growth of data and the progress of computing power, the combination of deep learning and course Recommendation provides a new idea to solve the above problems. The research on in-depth learning methods for education platform curriculum recommendation has become a research hotspot in the field of education. In view of the challenges faced by the current curriculum recommendation, this paper studies the curriculum recommendation method and system of the education platform based on deep learning.

Keywords: Deep learning · Education platform · Course recommendation

1 Introduction

With the popularity of the Internet, great changes have taken place in learners' learning methods, which helps to promote the development of personalized education and promote the reform of education system and education model. The individualized education concept that varies from person to person breaks the traditional education and teaching methods, so that the education and teaching goal of "teaching students according to their aptitude" is no longer just a slogan. In the Internet environment, a large number of high-quality curriculum resources are widely spread and shared, which provides convenience for learners. Learners can arrange learning according to their own time to meet their personalized learning needs.

At present, the development of course recommendation applications is relatively slow, and most of them use the traditional collaborative filtering recommendation algorithm. Collaborative filtering recommendation algorithm has some problems, such as cold start, data sparsity and scalability, which can not meet the needs of current education platform course recommendation. How to help learners obtain appropriate curriculum resources and carry out personalized learning has become an important research topic in the field of intelligent education. With the popularization of online education and the rapid development of education information system, a large number of education and

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2022, LNICST 465, pp. 406–415, 2023.
https://doi.org/10.1007/978-3-031-23950-2_44

teaching data can be presented and saved, including learners' majors, interests, learning conditions, curriculum feedback and other records [1]. By analyzing the historical web page information that learners have browsed, we can obtain learners' interest preference information. At the same time, comprehensively considering the factors such as learning background, personal interest and major, recommending suitable courses to learners to meet learners' personalized learning needs has important theoretical research significance and practical application value.

Personalized recommendation is based on users' historical access data for analysis and calculation. The personalized recommendation in the system is based on Web log data mining and is implemented by using ALS machine learning algorithm in spark. ALS algorithm is based on collaborative filtering for course recommendation.

The recommendation algorithm is mainly based on the relationship between users. The required data includes user ID, course ID and course score. However, there is no user rating data for the course in the web log data, only the course content accessed by the user. Therefore, the system uses the number of times the user watches the course video as the rating for the course.

This paper counts the times each user watches each video from the web log data, and reflects the user's preference for the course according to the user's access times to the course. The more you watch, the more you like it, and vice versa. ALS algorithm can find similar users according to their favorite course videos, and there must be an intersection between their favorite videos. Then, it analyzes the favorite course videos of similar users, and removes the course videos that users have seen to recommend them. It can be seen that the algorithm is based on other existing user preferences to predict the user's rating or liking for the video they have not watched, and can not make personalized recommendations to new users.

The user information is saved in the MySQL database. When logging in, the user checks the account and password by accessing the database.

On the other hand, since the log automatically generates 200000 lines using scripts, the effect recommended here is not very obvious.

Hot spot analysis includes two parts: hot word analysis and regional analysis. First, extract the course name from the web log data, and use the jiebar extension package for word segmentation. During word segmentation, it is necessary to establish a course dictionary and a stop vocabulary for the system, because the course name in the web log contains some special nouns, such as "the Belt and Road". Using the system dictionary, it will be divided into two words: "the belt and road" and also contains a large number of stop words, such as "de" and "yes". In this way, the accuracy of word segmentation can be effectively improved.

Hot word analysis is to recommend some keywords to users. Users can search and view through the search function, which is also a potential recommendation method.

This function module extracts the user's access IP from the web log. Through the crawler, you can use the API provided by the data science toolkit website to obtain the geographical location according to the IP. According to the geographic coordinates, the remaph function of remap extension package is used for geographic data visualization. Remap extension package can draw excellent geographic heat map and migration map.

The results returned by the API provided by the data science toolkit are stored in the form of Josn and need to be processed with regular expressions. R language regular expression gsub function can complete the replacement of regular expression. The results returned by the API include country, city, operator, geographic coordinates, etc. By observing the data, it can be found that only the geographical coordinates are of numeric type, and the rest are of string type. Therefore, the data processing process obtained by using the web crawler is relatively simple. You can directly replace the non numeric string with null. After the replacement, the rest are numeric types, that is, longitude and latitude. However, it seems that this API can not be accessed directly now. It is estimated that you need to climb over the wall to use it.

However, what I need to explain here is that the geographical heat map is drawn through the remap package. However, due to the conflict between the CSS of the remap package and that of shiny, there are some failures in the system interface, because I save the geographical heat map as a picture for display.

2 Relevant Theories and Technologies

When a new user or an unlisted user enters the learning platform, the system cannot know the specific information of the user, so it does not know the interests of the user, so it cannot use the conventional recommendation algorithm for recommendation. Generally, at this time, it is only before recommending items that are generally well reflected to users. In the face of this situation, the project extracts courses that have been viewed more times and recommends them on the home page through web log analysis.

Home page recommendation is to extract the course IDS accessed by users from the preprocessed data, then count the number of times these course IDS appear, sort them according to the number of times they appear, and recommend the most popular and visited courses to users.

The search function can help users quickly find all courses related to search keywords. These course name data are extracted from the web log using R language, so all courses existing in the web log can be searched. The search engine is based on regular expressions. By performing regular expression matching in the course name in the course data through the grepl function, you can search all courses containing keywords and obtain their serial numbers. When the search is completed, the serial number collection is returned. The found courses are displayed in the front end according to the serial number and ID number of the courses.

Users can use the search function to complete the course search. On the other hand, this function improves the hot word analysis function in hot spot analysis. Users can view hot words and then use the search function to view courses.

Course classification mainly uses manual methods to add labels to each course in the course information summary file, which can be displayed according to the labels in the front display. Course classification can help users view courses according to categories. The design includes 13 categories, such as “office”, “examination”, “foreign language”, “construction”, “finance”, etc.

At the beginning of crawling data, I didn’t expect that there would be a course classification function in the later stage, so no classification labels were added to the data, so the data classification labels were added manually in the later stage.

2.1 Recommended System Definition

Wikipedia interprets recommendation system as an application of information filtering. In his book "practice of recommendation system", Xiang Liang explained the recommendation system as an information filtering and information push tool that automatically contacts users and projects. Adomavicius et al. Gave a formal definition of recommendation system. Suppose u represents the user set and V represents the set of items (such as books, movies and restaurants), and the number and scale of users and items are large. F represents a utility function used to calculate the recommendation of item V to user u, then the utility function f can be defined as: F: u v r, where R represents a fully ordered set of real numbers or non negative integers in a limited range. The task of the recommendation system is to find the item V with the highest recommendation for any user u U. the formal definition formula is as follows:

$$u, U, v_u \, arg \, maxf(u, v) \tag{1}$$

By modeling user preferences, the recommendation system obtains the potential characteristics of users and projects, predicts users' scores on projects, or forms a recommendation list. Goldberg gives a general process of recommendation system, which is divided into three modules, as shown in Fig. 1.

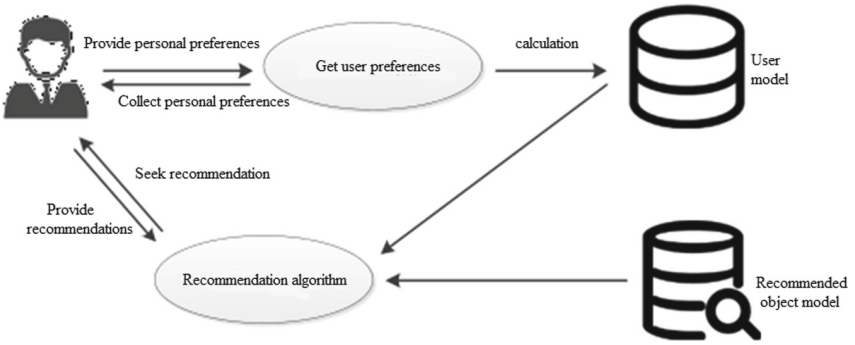

Fig. 1. General structure of recommendation system

2.2 Deep Learning Technology

In recent years, deep learning has made rapid development in the field of recommendation system. The performance of deep learning largely depends on the network structure. In order to adapt to different types of data and problems, various types of neural network structure models are applied to recommendation system. The following will describe in detail the two network model structures of multi-layer perceptron and long-term and short-term memory network.

1) Multilayer perceptron

Multilayer perception (MLP) is a basic neural network model. Because information propagates forward in MLP, and any two layers of neurons are connected with each other, MLP is also called forward propagation network and fully connected neural network. MLP network structure mainly includes input layer, hidden layer and output layer. The hidden layer can have one or more layers, and each layer is composed of several neurons, as shown in Fig. 2.

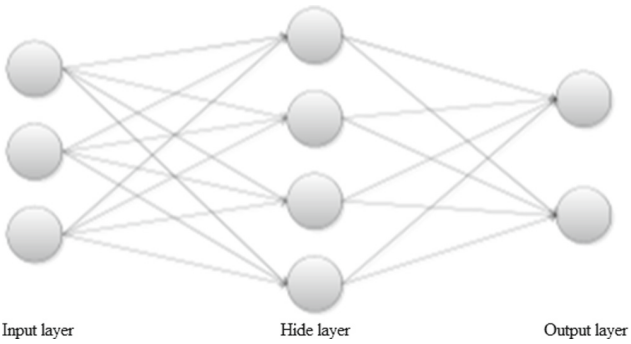

Fig. 2. Structure diagram of multi-layer perceptron

2) Cyclic neural network

Cyclic neural network is derived from Hopfield network proposed by saratha sathasivam. It is a kind of artificial neural network using sequence information. At present, it is mainly used to process and predict sequence problems. It is widely used in speech recognition, machine translation and timing analysis. Figure 3 and 4 show the basic structure of the recurrent neural network [2]. The output of the recurrent neural network at the current time depends on the input at the current time and the state at the previous time. We can use the following formula to descript it.

$$\begin{aligned} x &= \sum_{i=0}^{n} \sqrt{\frac{(X-\overline{X})^2}{n-1}} \\ x &= \sum_{i=0}^{n} \frac{(X-\overline{X})^2}{n-1}, \\ x &= \sum_{i=0}^{n} \frac{(X-\overline{X})(X-\overline{X})}{n-1} \end{aligned} \tag{2}$$

where x is the network layer.

In order to realize the neuron network with deep learning. The formula of the algorithm is following:

$$
\begin{aligned}
&T = (t_1, t_2, \ldots, t_n) \\
&i_t = \sigma\left(w_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \\
&f_t = \sigma\left(w_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \\
&O_t = \sigma\left(W_0 \cdot \left[h_{t-1}, X_t\right] + b_0\right) \\
&\tilde{c}_t = \tan h\left(w_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \\
&C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t \\
&h_t = o_t \cdot \tan h(c_t) \\
&u_t = \tan h(w_s h_t + b_s) \\
&a_t = \text{soft}\, max\left(u_t^T, u_s\right) \\
&v = \Sigma_t a_t h_t \\
&y = \text{soft}\, max(v)
\end{aligned}
\tag{3}
$$

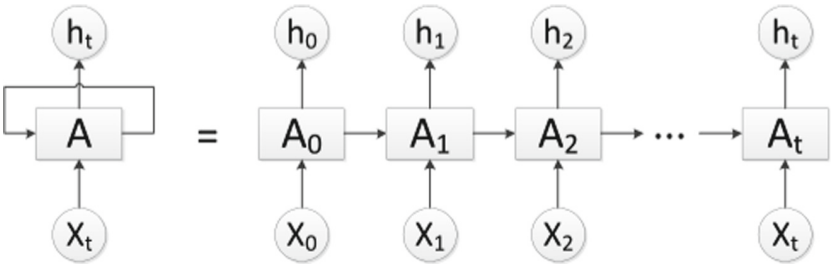

Fig. 3. Structure diagram of cyclic neural network

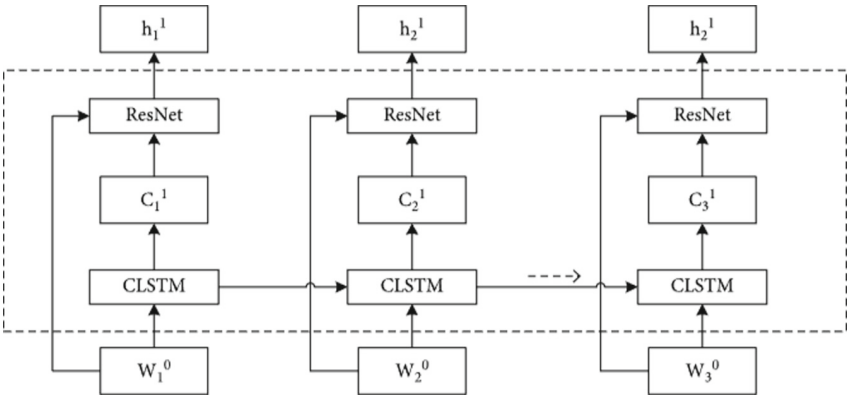

Fig. 4. Deep learning model

3 Curriculum Recommendation Method and System of Education Platform Based on Deep Learning

3.1 Personalized Course Recommendation Framework

As shown in Fig. 5, firstly, multi-source heterogeneous data fusion is carried out for book borrowing information, students' basic information, students' history course selection information and students' course selection behavior information, so as to absorb the characteristics of different data sources and enrich students' information, which is conducive to the expression of students' interest preference vector in model learning. Secondly, personalized course recommendation is carried out through deepafm model, and the combination of high-order and low-order features is integrated to improve the prediction ability of the model. Finally, the personalized course recommendation results are presented in the hot areas concerned by students to improve the course click through rate.

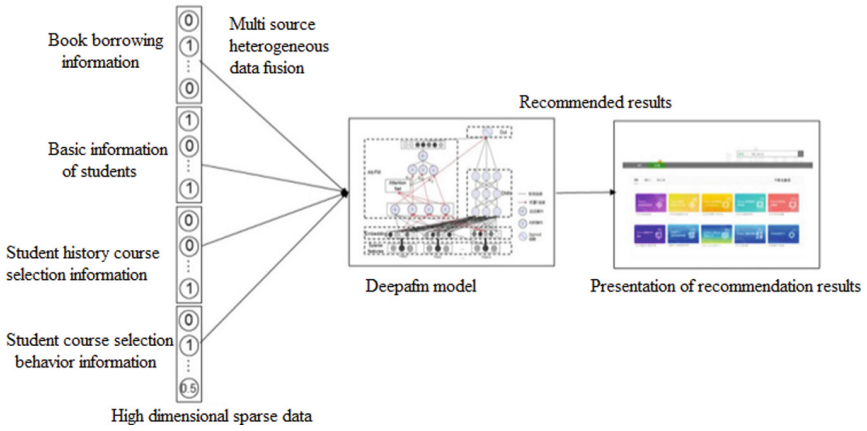

Fig. 5. Personalized Course recommendation framework

3.2 Application of Depth Recommendation System to Course Recommendation

Deep neural networks can learn potential connections and expressions in data. Usually, some information about courses and students can be obtained on the learning platform. Using this information can improve the understanding of courses and students, so as to provide better recommendation effect. Therefore, it is a wise choice to apply deep neural network to representation learning in course recommendation [3]. The application of in-depth recommendation system in course recommendation is mainly reflected in the following four levels:

(1) Improve the representation learning ability: the deep neural network can use the nonlinear activation function to model the nonlinearity in the data. This attribute can capture complex student course interaction patterns, and can effectively learn potential connections and useful feature representations from input data by using the powerful nonlinear fitting ability of deep learning, which is shown in Fig. 6.

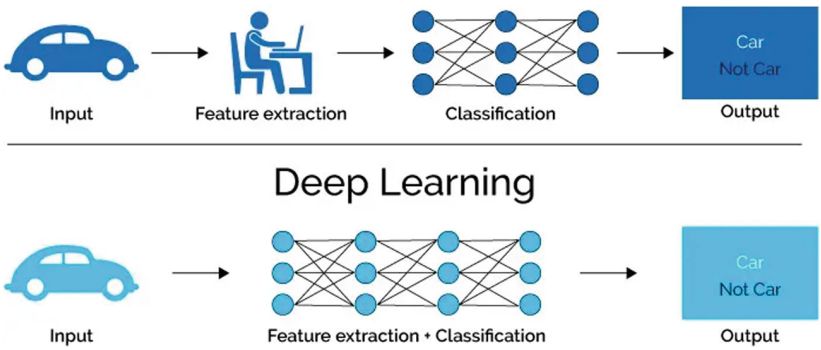

Fig. 6. Personalized Course recommendation deep learning network

(2) Deep collaborative filtering: the classical matrix decomposition model can be considered as a simple neural network. Through the interaction of the neural network of the deep learning model, more nonlinear feature interaction can be introduced to improve the fitting ability of the recommended model which is shown in Fig. 7.

(3) Deep interaction between features: in order to improve the accuracy of the model, some features will be artificially interactive in the traditional recommendation system to improve the expression ability of students' preferences, which requires a lot of artificial feature engineering, and the effect also has great limitations. The deep neural network can automatically learn feature interaction from the original data by unsupervised or supervised methods. It not only improves the efficiency, but also greatly improves the recommendation effect of the model.

(4) High flexibility: with the support of many popular deep learning frameworks, the model of deep learning has a high degree of flexibility. The deep learning model can flexibly adjust its own structure according to the course recommendation scenario and data characteristics to adapt to the application scenario [4]. For example, different neural network structures are usually combined to establish a powerful hybrid model or replace one module with other modules. Therefore, it is easy to build hybrid and composite recommendation models to capture different features and factors at the same time.

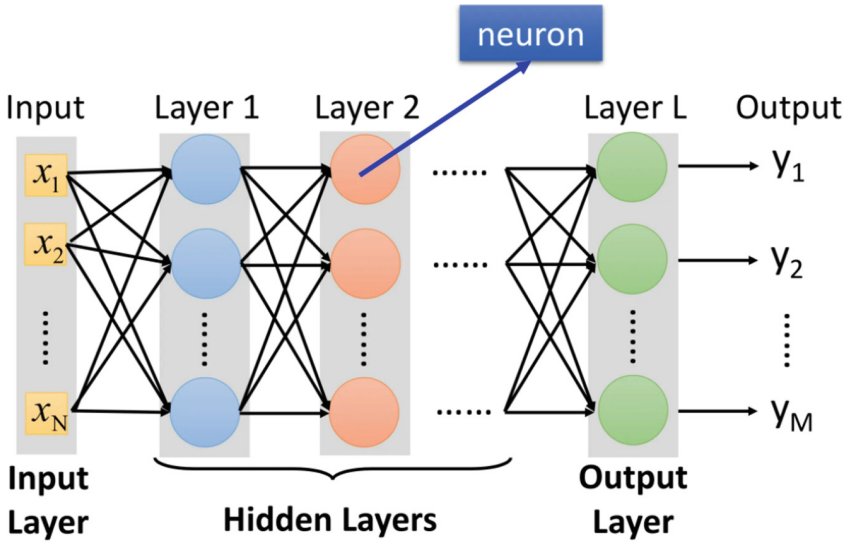

Fig. 7. Personalized Course recommendation neuron network

4 Conclusion

In the recommendation system based on deep learning, the commonly used deep learning technologies include neural networks such as RBM, MLP, AE, CNN and RNN. Among them, CNN, AE and their variants are often used in feature engineering to extract features from multimodal data such as text, image and video, RBM and MLP are often used in potential factor model, and RNN and its variants (such as long-term and short-term memory network, gated loop unit, etc.) are often used in recommendation based on sequence information. The curriculum recommendation method of education platform based on deep learning aims to better express how the time sequence of courses affects students' performance. From the initial method based on collaborative filtering, to machine learning method and matrix decomposition method, and then to the recommendation method based on deep learning, it is to realize personalized current course recommendation and gradually improve the effect of course recommendation.

Acknowledgements. 1. Xi'an Translation College 2021 University-level Online Curriculum Construction Project (ZK2143) "Introduction to Mao Zedong Thought and the Theory System of Socialism with Chinese Characteristics (Special School)".

2. The first batch of research projects and practical projects of online ideological and political work in Shaanxi universities (2021SPWSXM-Z-8) "Practice Innovation and Exploration of Shaanxi Local Party History Network Education in Colleges and Universities".

References

1. Li, Y.: Solving high dimensional partial differential equations based on deep learning method. Wang Lulu, Wuhan University (2019)

2. Yan, T.: Research on network representation learning method and application of multi-mode data fusion. Li Jiayi, Taiyuan University of Technology (2019)
3. Li, W.: Research on network representation learning method maintaining generalized structure. Zhu Sinan, Beijing Jiaotong University (2019)
4. Xi, X.: Research on emotion dictionary construction algorithm based on depth representation. Wang Leyi, Nanjing University of Technology (2018)