International Journal of Information Technology and Web Engineering
Volume 18 • Issue 1

A Course Recommendation Algorithm for a Personalized Online Learning Platform for Students From the Perspective of Deep Learning

Zhengmeng Xu, Industrial and Commercial Bank of China, China
Hai Lin, Hebei Finance University, China
Meiping Wu, Guangzhou Health Science College, China*

ABSTRACT

This paper mainly studies the content of the recommendation algorithm of learning resource courses in online learning platforms such as MOOC and mainly introduces the automatic encoder neural network that integrates course relevance to realize the personalized course recommendation model. The authors first introduce how to embed a course relevance decoder in an autoencoder neural network. Secondly, the proposed confidence matrix method is introduced to distinguish the recommendation effect of the learned to the unlearned courses, and the training process of the model is introduced. Then, the design content of the experiment is introduced, including the model structure, comparative experiments, parameter settings, and evaluation indicators. Finally, the experimental results are analyzed in detail from the horizontal and vertical aspects. It is hoped that this research can provide a reference for personalized recommendation of learning resources based on deep learning technology and big data analysis.

KEYWORDS

Autoencoder, Big Data Analysis, Deep Learning, Personalized Recommendation, Recommendation Model

INTRODUCTION

With the continuous development of internet technology, we have entered an era of information explosion, which provides unprecedented opportunities and challenges for online learning. The advancement of internet technology has driven the rapid development of online education, enabling more and more people to realize their desire for remote online learning and access rich learning resources without leaving their homes (Tzavara et al., 2023). However, the era of big data is also accompanied by the problem of information overload, making it difficult for learners to find suitable content from massive resources.

DOI: 10.4018/IJITWE.333603
*Corresponding Author

In this context, recommendation systems have emerged as one of the effective tools for dealing with information overload issues. The recommendation system can efficiently screen out important learning content and provide personalized online course recommendation services for learners (Bang et al., 2023). The research in this article focuses on learning resource course recommendation algorithms in online learning platforms, with the aim of improving the effectiveness of recommendation systems to meet learners' personalized recommendation needs.

Meanwhile, personalized learning has become an important goal and trend in the field of education, receiving widespread attention and support. Online education platforms are constantly expanding their learning resource pool to meet the diverse needs of different learners (Johnson et al., 2023). We can now easily access various learning resources, promoting the convenience of self-directed learning and problem-solving.

Deep learning technology, as a part of neural networks, provides an opportunity to build more accurate personalized learning resource recommendation systems to provide more intelligent recommendations based on user attributes and learning resource types. This article will explore how to apply deep learning technology and build a personalized learning resource recommendation system based on deep learning to meet the growing demand for personalized learning.

This article consists of five parts, and the main content of each part is arranged as follows:

The Introduction provides an overview of the research background of the course recommendation system, as well as the main content and significance of this article. Finally, here is an introduction to the organizational structure of this article. The Literature Review introduces the relevant information of deep neural networks and automatic encoder networks. The Methodology describes course relevance encoders and integrating relevant information on course relevance, model design, and experimental design. It also explains the relevant theoretical technologies, such as recommendation systems, word vector embedding, and neural networks, and applies them to course recommendation models. Finally, experimental design was conducted to verify that the constructed recommendation model can improve recommendation effectiveness and provide more satisfactory personalized recommendation services for online learners. The Result Analysis and Discussion verifies the results and the superiority of this model and explains the corresponding reasons through experimental comparison of the baseline model. Two classic recommendation algorithms, User-Based Collaborative Filtering (UBCF) and based collaborative filtering (BCF), were selected to verify the objectivity of the results. Through variant experiments, it was verified that the Self-Attention-based Course Recommendation (aeCrdRec) model applied a self-attention mechanism encoder to capture learners' preferences and decode course relevance to obtain the hierarchical relationship of courses, which to some extent, improved the evaluation performance of the deep neural structure model. The Conclusion mainly summarizes the main content of the paper and discusses the research direction of future course recommendation systems.

LITERATURE REVIEW

Deep Neural Networks

Machine learning is a research hotspot in the field of artificial intelligence technology, and its purpose is to allow computers to simulate human learning activities (Adedoyin & Soykan, 2023). Deep learning technology is an important part of machine learning (Wong, 2023). Compared with traditional machine learning, deep learning technology can extract the deep features of the data, thus solving a problem. A shallow model cannot, leaving the features to be designed manually (Kang & Zhang, 2023). Deep learning technology can analyze and predict new similar data after learning from existing data. In recent years, it has been widely used in natural language processing, personalized recommendations, image analysis and other fields, and has achieved good results (Sofi-Karim et al., 2023). The further application of deep learning technology in the field of education will definitely promote the innovation and reform of education, and it will also lead to changes in teaching mode, teaching method, teaching conten (Elglaly & Liu, 2023). The education field will make full use of the advantages of deep learning technology, which will certainly inject new vitality into personalized recommendations and learning and promote the realization of personalized learning goals (Bharadiya, 2023). This is the value and development opportunity brought by deep learning technology to education.

Deep neural networks are developed from artificial neural networks and are the basic model of deep learning technology (Kuadey et al., 2023). Artificial neural networks simulate the principle of the human brain perceiving the outside world, with neurons, also known as perceptrons, receiving input from external information and outputting results after internal processing (Amini & Rahmani, 2023). Artificial neural networks achieve various functions by connecting multiple perceptrons to each other. According to the signal processing flow, it is divided into an input layer, a hidden layer, and an output layer. Deep learning can well show the characteristics of data, the hierarchy of the model, the number of parameters, and the capacity (Sanusi et al., 2023).

The deep learning model can achieve better results on large-scale training data. The deep learning models mainly include deep confidence network, deep Boltzmann machine trestle automatic encoder, convolutional neural network, etc.

The convolutional neural network can process multidimensional input data and retain the original local spatial information of the input data, and its feature extraction ability is also very strong. It is mainly used in image processing and speech recognition tasks; the recurrent neural network has a "memory" function due to the interconnection of nodes between its hidden layers. The idea of a convolutional neural network is a simulated functional structure designed for two-dimensional data. The deep network is constructed through multiple convolution and pooling processes. The training of the network contains the characteristics of "weight sharing" and "sparsity." The learning parameter process is similar to a Back Propagation (BP) algorithm. The advantages of a convolutional neural network model are strong learning ability, good effect in feature extraction, and high classification accuracy (Martins & Gresse Von Wangenheim, 2023).

Autoencoder Network

The idea of the automatic coder is a deep network constructed by combining multiple automatic coding networks (Cardona et al., 2023). First, greedy pretraining is performed layer by layer, and the learning process of each layer is consistent with the BP algorithm. After the pretraining is completed, the network is expanded into a deep-level forward network, and then the BP algorithm is used for overall microadjustment. The advantage of the automatic coder model is to solve the feature learning process of tagless data, which is widely used in pattern recognition (Al-Emran et al., 2023).

As shown in Figure 1, the function-based autoencoder is mainly composed of three layers of input layer; its loss function can be expressed as:

$$h = f(x) = s_f(W_x + b)\,. \tag{1}$$

Among them, *s_f* is the activation function in the network. The dimension of the weight matrix *W* in the encoder parameters is *mxn*, and the offset vector *b R^m*.

Reconstructing the implicit representation *h* as $\widehat{X}_x = \left\{\widehat{X}_1, \widehat{X}_2, \cdots, \widehat{X}_n\right\}$, the loss function can be expressed as:

$$x' = g(h) = s_g(W'h + b')\,, \tag{2}$$

where s_g represents the activation function of the decoder. Likewise, the dimension of the decoder's parameter weight matrix *W'* is *nxm,* and the offset vector $b \in R^n$. s_f and s_g are usually nonlinear activation functions, such as hyperbolic tangent and sigmoid functions. In contrast to principal component analysis (PCA) methods, nonlinear activation functions can help autoencoder neural networks learn more useful object features.

The goal of training the autoencoder neural network is to minimize the error value between *X* and the reconstructed *X* . The reconstructed error value has two calculation methods: squared error and cross-inheritance. The specific calculation method can be expressed as follows:

Squared error:

$$E_{AE}(x,\hat{x}) = \left\|x - \hat{x}\right\|^2 . \tag{3}$$

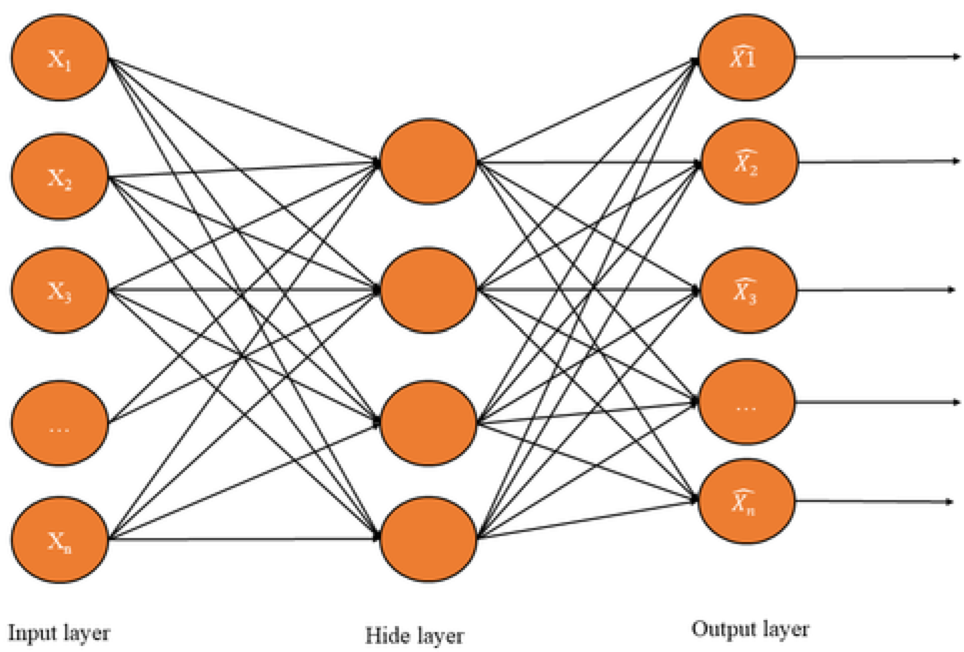
Cross entropy:

$$E_{AE}(x,\hat{x}) = -\sum_{i=1}^{n}(x_i \log \hat{x}_i + (1-x_i)\log(1-\hat{x}_i)) . \tag{4}$$

When calculating the reconstruction error of an autoencoder neural network, adding a regularization parameter constructs a loss function:

The personalized learning resource recommendation system based on deep learning can meet the needs of personalized learning and precise learning. Therefore, it is necessary to build a personalized learning resource recommendation system based on deep learning. Integrating deep learning into the learning resource recommendation system will help improve the utilization rate of learning resources, improve the accuracy of recommendation results, and improve the learning effect of learners.

Figure 1. Autoencoder neural network



METHODOLOGY

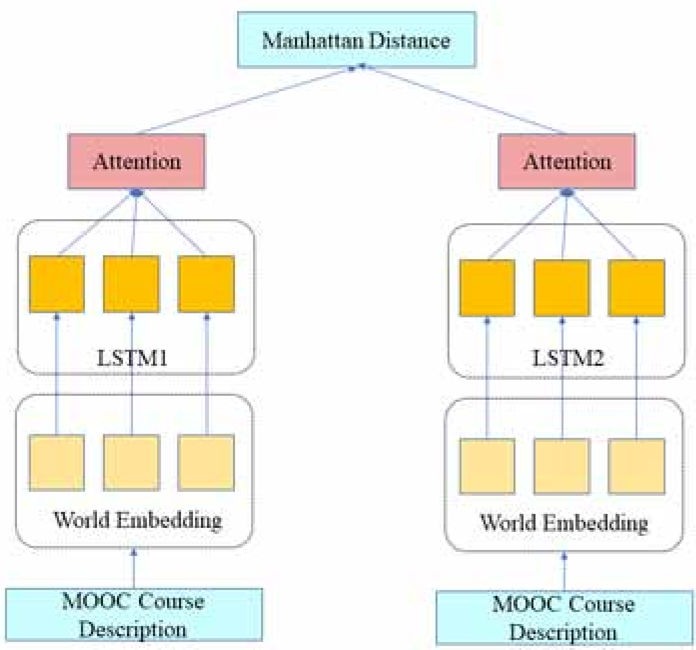
Course Relevance Encoder

This paper proposes an attention-based Manhattan twinning long-term short-term memory network (AMSLSTM) to predict the semantic similarity of course description sentences, thereby capturing the relevance of courses (Hussain & Khan, 2023). The network enriches the attribute characteristics of courses so that the course recommendation system has the following advantages: firstly, it introduces the semantic association between courses, which helps to obtain the interactive information of implicit courses; secondly, it can reasonably expand the learner's preferences and increase the recommended courses abundance; and finally, for the calculated course relevance factor, the learner's historical learning course records can be traced to perform an interpretable analysis of learner preferences.

The AMSLSTM network structure is shown in Figure 2, which is mainly composed of three steps:

First, it is necessary to convert the descriptive text information of the course into word vectors that the neural network can learn. Since a Recurrent Neural Network (RNN) is a neural network model suitable for sequence data $(x1, \cdots, x_t)$ and Long Short-Term Memory Network (LSTM) not only inherits the advantages of RNN, but also considers the long-term dependent state of words, it is more suitable for discovering the contextual semantics of long sentences than RNN. Therefore, an LSTM-based network is used for text training. An LSTM learns to map in the vector space of variable-length sequences. Specifically, each sentence (represented as a sequence of word vectors) passes $x1, \cdots, x_t$ to the LSTM neural network, and then changes at the place of each sequence index. Its implicit representation is *h_t*, and the end of the sentence is represented by an implicit code at the end of the neural network. This paper represents the sentence candidate pairs described in the course as two embedding matrices, which are input to the twin long-term short-term memory network (essentially the same LSTM network), and the final state after each sentence output is a 64-dimensional

Figure 2. The structure of the curriculum relevance encoder



vector. These vectors catch after training, and obtain the semantic information of course sentences. At each time step $x \in 1, \cdots, t$, the hidden state vector *h_t* can be updated by the following equation:

$$h_t = \sigma(W \cdot h_t + U \cdot h_{t-1}) . \tag{5}$$

Secondly, the attention mechanism is embedded in the long-term memory network, which can better capture the hierarchical relationship of the implicit courses in the course description information:

$$a_c = soft\max(\tanh(W_l \cdot h_t + b_l)) . \tag{6}$$

Among them, *W_l, b_l*, represent the weight value and offset value of the attention layer, respectively. *a_c* is the attention weight value corresponding to the hidden state at each time t.

$$h_i = \sum\nolimits_t a_c \cdot h_t . \tag{7}$$

Finally, a distance function—Manhattan distance (used to compare the final hidden states of the two LSTM layers) is learned on the final output layer of the Siamese network so that the two A coded sentence is measured for relevance:

$$D\left[K_s\right] = \exp\left(-\left\|h_1 - h_2\right\|\right) \tag{8}$$

h_1 and *h_2* represent the left and right subnetwork outputs. Since the exponential function is calculated on negative numbers, the resulting predicted value is a floating-point number between 0 and 1.

Integration of Curriculum Relevance

Embedding Curriculum Relevance Factors

The personalized learning resource recommendation system mainly consists of a user module, a learning resource module, and a recommendation system module. Among them, the user module mainly includes a user and user model (Kubsch et al., 2023). The user is the key to obtain the learning resource recommendation information experience through the recommendation algorithm, the core of the whole recommendation system model, and the learner who realizes the deep learning experience. In the user model, the user's main information can be stored, such as the user's age, gender, and preference. Statistical information includes information such as education level, age group, background, and learning motivation. After the user's personal preference information is extracted, data analysis can be carried out, and the resource information required by the user can be pushed by the data analysis recommendation algorithm. The learning resources module includes information on subject areas, difficulty levels, forms.

After a learner has finished a course, there is a high probability that they will continue to study a course related to the previous one so as to expand their knowledge and achieve the purpose of consolidating the knowledge points of the learned course (Jurgensmeier et al., 2023). Since learners are more inclined to study courses that are related to the history courses they have learned, the existence of correlation in the course relationship will affect the learner's decision to choose the corresponding course so the recommendation effect will be affected by the course correlation factor (Guleria & Sood, 2023). The SaeCrdRec model proposed in this paper is able to discover the influence of learned courses on unlearned courses. Given the set $Ks = (k_1, \cdots, kn)$ of the learner's registered courses, then the influence factor of the learned courses on the unlearned courses is:

$$L_s = W_4 \cdot W_1 \left[K_s\right]. \tag{9}$$

Among them, $L_s \in R^{N \times n}, W_4 \in R^{N \times H}$. Each column in L_s represents the influence of each course on other courses (where the influence on its own course is set to 0).

Although formula (9) adopts the calculation method of dot product to obtain the basic correlation between two courses, it needs to deeply consider the relationship between the predecessor and the subsequent prerequisites between courses in the model. In order to integrate the hierarchical relationship of courses, the influence of course relevance factors on the recommendation results is considered. For the influence of the courses studied on the courses not studied in other aspects, the matrix integrating the relevant factors of courses can be expressed as:

$$L_s = \left(W_4 \cdot W_1 \left[K_s\right]\right) \otimes D\left[K_s\right]. \tag{10}$$

Among them, $D\left[K_s\right] \in R^{N \times n}$ calculated above, $D\left(k_i, k_j\right)$ and ⊗ are the element product.

To calculate the impact of all courses on unlearned courses, summing each row of *L_s* gives $L\_s \in R^N$:

$$l_s = \sum_{i=1}^{N} \sum_{j=1}^{n} L_s^{(i,j)} . \tag{11}$$

Among them, i and j represent the index values of the row and column in the *L_s* matrix, respectively. Then, after embedding the course relevance factor, the output vector (x_s)ˆ of the autoencoder integrating the course relevance in the reconstruction space is obtained, which can be expressed as:

$$\hat{x}_s = f_4(W_4 \cdot y_3^s + l_s + b_4) . \tag{12}$$

Among them, $W_4 \cdot y_3^s$ is the learner preference feature representation obtained from the attention mechanism encoder, *l_s* is the curriculum relevance calculated by the curriculum relevance decoder, and b_4 represents the offset value.

Model Training

This paper adopts the implicit feedback recommendation method of the behavioural data of learners registering to learn courses and studies the behavioural data records of learners' learning courses. It can be found that the main content is the relevant information of the courses they have studied. If a student likes one of the courses, the frequency of their study of this course will increase accordingly, so this paper considers that the interaction data of the student enrolling in the course of study represents the data record of the course of study as the training positive example during the training process.

Another problem is that the behavioural data of these learners lacks negative examples, which makes it difficult to obtain more real preferences of learners from the registered learning course data. Therefore, this paper adopts the weighted method to distinguish the courses that the learners have learned and the courses they have not learned, that is, all the unlearned courses of the students are regarded as the negative examples of training, and these negative examples are uniformly set as the same weight value. When the frequency of learners learning courses increases, the corresponding weight value will also increase, and the way of increasing shows a proportional trend. According to this method, the confidence matrix *C* of the learner's preference can be proposed, wherein the weights of the elements in the matrix are set as follows:

$$c_{s,c} = \begin{cases} 1 + \alpha \log(1 + f_{s,c} \ / \ \varepsilon) \\ 1 \end{cases}, \tag{13}$$

where α and ε represent confidence weight parameters. *f_(s,c)* is the weight value of the positive example calculated according to the first formula in formula (13) when the frequency of the learner's course is greater than 0. In other cases, the weight value of $C_{s,c}$ is set as 1.

After embedding the confidence matrix, the objective function of the model proposed in this paper is as follows:

$$Loss = \sum_{s=1}^{M} \sum_{c=1}^{N} \left\| C \otimes (X_{S,C} - \hat{X}_{S,C}) \right\|_2^2 . \tag{14}$$

Since the regularization method is not only robust but also has a higher generalization effect, which can well mitigate noisy data. Therefore, after embedding the regular term, the model objective function proposed in formula (14) can be rewritten as follows:

$$L = L\cos + \lambda(\left\|W_i\right\|_2^2 + \left\|W_a\right\|_2^2 + \left\|W_t\right\|_2^2 . \tag{15}$$

Among them, *λ r*epresents the regularization parameter, and *i*∈(1, 2, 3, 4) in W_a, w_t represents the corresponding weight matrix, respectively. W_a, w_t are the learned parameters in the corresponding attention layer and pooling layer, respectively.

Whether it is a course recommendation or a recommendation in other fields, how to solve the model quickly is the key point of research (Alsabhan, 2023). Due to the ever-increasing amount of data in the era of big data, it is a challenge to process models quickly. In this paper, the website data sources are used as input, and the text information of course descriptions is integrated into the designed model. Although it provides powerful data support for the course recommendation system, it increases the complexity of the model. Since stochastic gradient descent (SGD) has the characteristics of reduced computational complexity and good parallelism, it has been widely used in the training of recommendation models. Therefore, this paper adopts the gradient descent method of backpropagation to update all parameters of the model and calculate the partial derivatives of all parameters to minimize the objective function.

Gradient descent is a first-order optimization algorithm. To find the local minimum of a function by gradient descent method, an iterative search must be carried out to the specified step distance point in the opposite direction of the gradient (or approximate gradient) corresponding to the current point on the function. If the search iterates in the positive direction of the gradient, it will approach the local maximum point of the function, and this process is called the gradient ascent method.

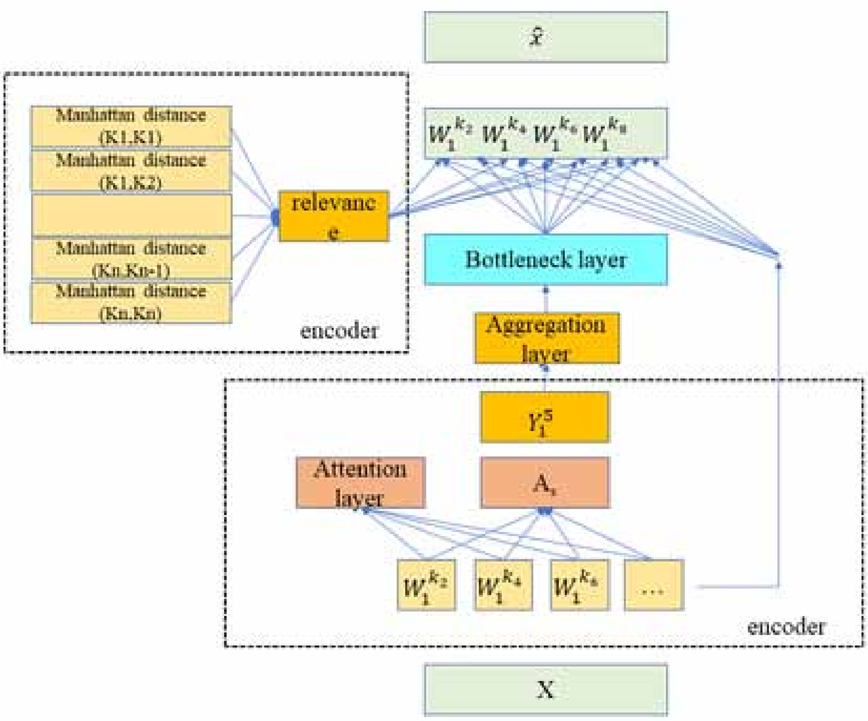
Experimental Design

This section will explain the experimental design from four aspects and verify the feasibility of the proposed SaeCrdRec model. The model experiment will use the Microsoft Windows operating system as the running condition of the experiment and the programming language Python and the Anaconda integrated development environment.

1) Model structure: As one of the unsupervised neural networks, the auto-encoder neural network has a network structure composed of an encoder and a decoder. The former uses an activation function on the input data to obtain new information in the hidden space. The latter uses an activation function on the feature representation vector to obtain the output in the reconstruction space. Figure 3 shows the framework of the SaeCrdRec model designed in this chapter.

The input data in the encoder are the binary rating matrix X converted from the historical interaction data of the learner and the course, and W_1 is the embedding matrix of the course, which is similar to the word embedding matrix of the word2vec model. The learner preference vector A_S expressed in the form of a multihot vector is input, and the embedding vector $W_1\left[K_s\right]$ of the corresponding learning course is obtained through the corresponding operation. Correspondingly, the input data in the decoder Y_I are to convert the text information of the course description into the word vector representation. In the twin long-term memory networks, the word vector containing the course description is learned to capture the implicit feature representation of the course, and the Manhattan distance is used. The method precalculates course pairs and finally obtains the relevance of courses.

Figure 3. Schematic diagram of the general framework of the model



Finally, the course relevance and learner preference features are integrated, and the predicted course scores are output in the reconstruction space of the autoencoder neural network. The ranking list of the recommended courses is obtained according to the predicted course recommendation probability. Specifically, when predicting scores, the model takes the binary rating vector Xs of each learner for the course as input, and outputs the corresponding rating vector Xs after reconstruction as the prediction result, which will not be in the training set, but the predicted score is top k courses that are recommended for target learners.

2) Comparative experiments: In order to verify the interpretability and effectiveness of the model designed in this paper, the following related recommended baselines will be compared experimentally:

User-Based Collaborative Filtering (UBCF): This method is a recommendation algorithm based on user similarity (Chen & Ding, 2023). Using the similarity preference between two learners, the courses are obtained from the set of learned courses of adjacent learners with high similarity, and the courses for which they are not registered are provided to the target learners.

Item-Based Collaborative Filtering (IBCF): This method is a recommendation algorithm based on the similarity of courses. Using the similarity between courses, the target learners are provided with courses that have been learned by adjacent learners and have a high degree of similarity.

Bayesian Personalized Ranking Matrix Factorization (BPRMF): This method is an implicit feedback recommendation algorithm (Al Shloul et al., 2023). The BPRMF model is trained by maximizing the probability of ranking pairs, where the user's ranking of one item should be higher than the user's ranking of other items. This method allows the model to learn from the implicit behavior of users and generate personalized recommendation lists.

Model parameters: In order to make the experimental results more reasonable, it is stipulated that each learner in the data set records the data of the learned course as a positive example of the training data and considers the unlearned course as a negative example. In addition, in order to achieve the best recommendation effect, the corresponding parameters are set in the research data set to train the model. For the parameters set by the SaeCrdRec model: the minimum batch value is 128, the learning rate value is 0.001, the dropout value is 0.5, the values of α and ε in the confidence matrix are 2 and 10^{-5}, respectively, and the dimension of the attention layer is set to 20, and the three-layer hidden layer dimensions are set to [200, 50, 200], respectively.

The way of dividing the training set and the test set in the neural network will have a certain impact on the recommendation results. As shown in Figure 4, it can be found that dividing the training data with different percentages does have an impact on the course recommendation method.

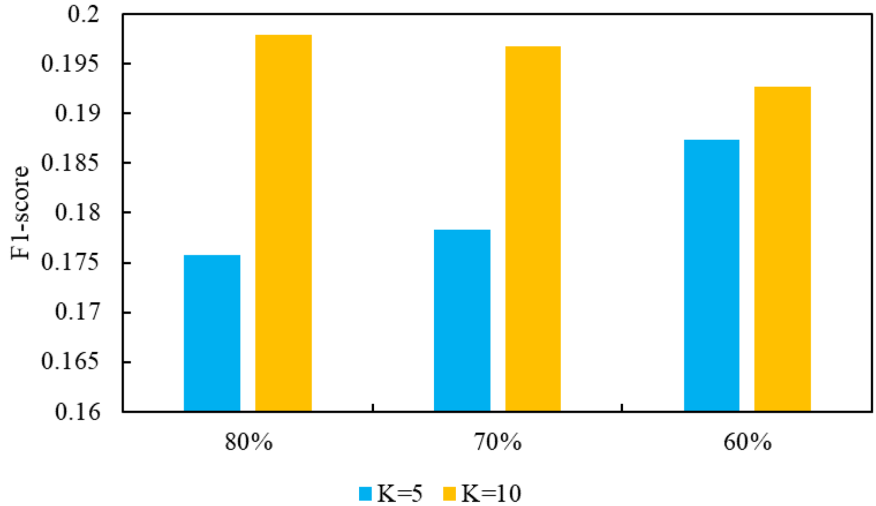
RESULT ANALYSIS AND DISCUSSION

Baseline Experiment

This section reports on the experimental comparison and validation of the baseline model. As shown in Figure 5, @k in the table represents the recommendation of the top k courses, which clearly proves that the SaeCrdRec model proposed in this paper outperformed other comparative baseline methods (in FI -score@10, it is improved by 2.54% to 8.28%). The following main points can be drawn:

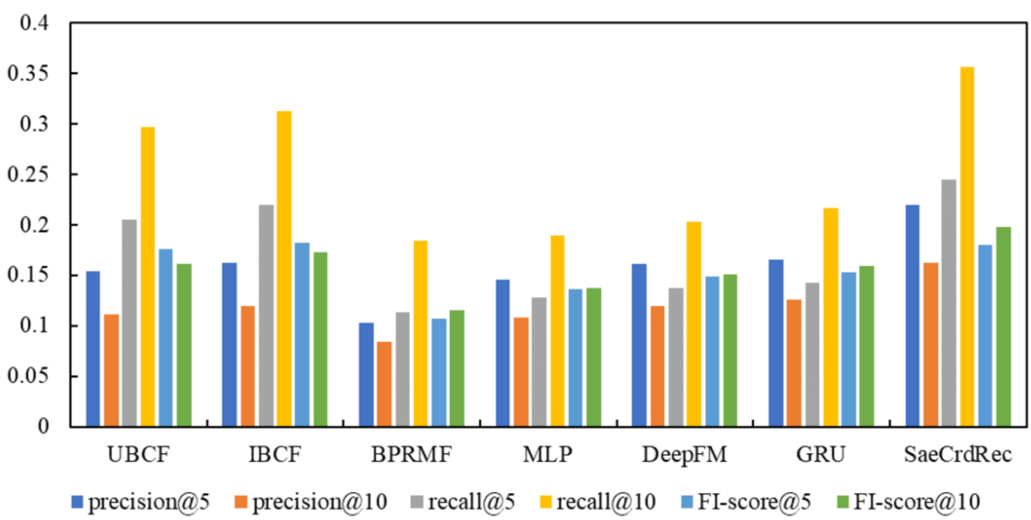
The BPRMF recommendation performance was inferior to other models. The main reason was that the BPRMF model conducts a pairwise ranking of courses according to learner preferences and does not take other factors, such as course relevance, into account. The SaeCrdRec model integrates course relevance factors, thereby further improving the effect of course recommendation. At the same time, it illustrated the importance of the SaeCrdRec model to deal with the impact of curriculum relevance.

Figure 4. FI-score under different data set divisions



The SaeCrdRec model performed the best in all benchmark experiments and had a more obvious advantage as the k value changes. The main reasons were: (1) In the Xuetang online data set studied, most learners only took a few courses, resulting in that many learners' embedded feature vectors could not be sufficiently learned from these sparse data; and (2) In all the baseline methods, the average and summation method was used for each historical learning course, and the ability of learners to represent their preferences for different courses was limited to a certain extent. The SaeCrdRec model not only made use of the nonlinear interaction between learners and course data but also used an attention mechanism to distinguish the importance of different courses to learners and then excavated learners' preferences for different courses. This showed that the use of a nonlinear interaction coefficient and attention mechanism in a neural network had a significant effect on recommendation.

Figure 5. Baseline effect comparison



It can be seen that the above analysis shows that the combination of a self-attention mechanism encoder and course relevance decoder could improve the accuracy of recommendation results for online course recommendation.

The authors have reported that adopting the implicit feedback method has a promoting effect on the recommendation in this study. As shown in Figure 6, this paper selected two classic recommendation algorithms—user-based collaborative filtering and based collaborative filtering (BCF) recommendation algorithms. Comparing and evaluating the performance of the SaeCrdRec model under different k values, it can be seen that the implicit feedback method was more objective than the learner's preference level obtained directly based on the learners' displayed ratings for the courses.

- Implicit feedback factor: For the evaluation of the performance of two different collaborative filtering recommendation algorithms under different k values, the x-axis represented the recommendation of the top k courses, and the y-axis represented the corresponding precision, recall, and FI-score, respectively. From this, the following main points can be drawn:

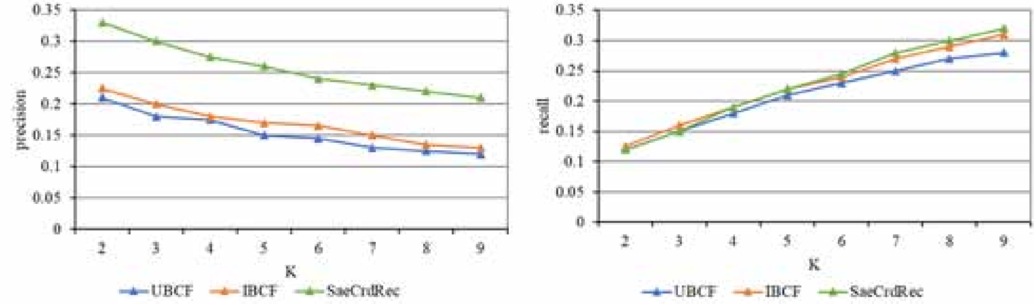
The SaeCrdRec model performed the best in the precision, recall, and FI-score metrics because the model addressed the "cold-start" problem commonly found in recommender systems. In the case of different k values, the accuracy of SaeCrdRec model was 8.9%~12.01% higher than UBCF, and 8.07%~10.26% higher than IBCF. Although the recall rate of the model proposed in this paper was not very different from the two collaborative filtering models at the beginning, as the value of k increases, the recall rate will gradually widen the gap with other models.

The SaeCrdRec model was much better than other evaluation metrics of UBCF and IBCF models. This is because in most explicit methods, the learner usually makes the final score after using the item once, but in online course learning, the preference may change over time. Therefore, the traditional explicit scoring method will lead to a large difference between the learner's scoring or evaluation information and the real preference, and even cannot reflect the learner's real preference characteristics.

As shown in Figure 7, this paper evaluates the experimental comparison between the neural network with the self-attention mechanism encoder and the curriculum relevance decoder and other neural network models.

The experimental results show that compared with other neural network baseline methods, the SaeCrdRec model had the best recommendation performance. For example, the SaeCrdRec model improves the FI-score@10 indicator by 3.91%~8.28%, which showed that the curriculum relevance decoder method is very effective in course recommendation to better capture the attribute characteristics of courses from course description text information.

Figure 6. Comparison experiment with recommendation algorithms based on collaborative filtering: (a) precision@k, (b) recall@k



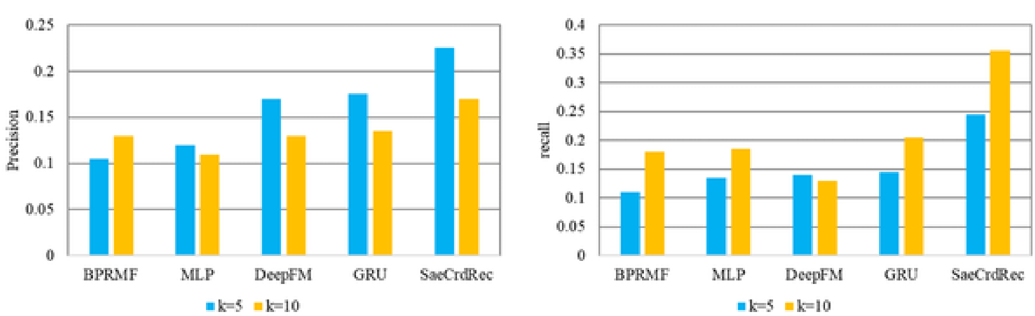
The performance improvement of a matrix factorization model mainly depends on the learners' ratings of courses, and BPRMF belongs to a matrix factorization method in implicit feedback recommendation (Safarov et al., 2023). The characteristic of Multilayer Perceptron (MLP) is multilayer, that is, it needs to set a large number of hidden layers and use linear data for training (Dolezal et al., 2023). Although MLP can learn from learner interaction data, it extracts learning from implicit feedback nonlinear interaction data. User preference is not outstanding. In addition, these data-feature time-item methods, such as BPRMF and MLP, have fewer courses when some learners only learn evidence-based features, and some sparse data cannot make these methods adequately capture the learner's feature embedding vector. The Gated Recurrent Unit (GRU) performs worse than the designed neural network because it treats all history learning courses equally, so it has a limited ability to capture the learner's preference level (Liang et al., 2023). Deep Factorization Machine (DeepFM) belongs to click-through rate (CTR) prediction, which is used to calculate the probability of learners clicking on recommended courses, but it does not integrate other attributes of courses, such as course relevance information. In conclusion, combining the nonlinear interaction between learners and courses and the influence of course relevance has a significant effect on improving the accuracy of predicting the rating of personalized recommended courses.

Variant Experiments

Regardless of the evaluation metric, in these variant comparison experiments, the SaeCrdRec model had the best performance results. The comparison results of the variant experiments will be analyzed in detail below. In addition, the performance of different components of the model is shown in Figure 8. It can be obtained:

1) Deep Autoencoder (DeepAE): The performance of the DeepAE was not bad. It was even better than using the implicit feedback method. The DeepAE uses a nonlinear activation function optimization in the deep network structure, and it can extract more complex feature relationships for nonlinear interactive data, which was better than the BPRMF model.
2) Deep Autoencoder with Attention (DeepAE-ATT): The DeepAE-ATT was better than the DeepAE because using the self-attention mechanism encoder, it can effectively distinguish the importance level of different courses to learners, better reflect the different preferences of learners, and be more fine-grained, thereby improving the effect of recommendation.
3) Deep Autoencoder with Course Correlation (DeepAE-COR): The recommendation results of COR were better than those of DeepAE-ATT and DeepAE. The reason was that DeepAE-COR excavated the relationship between the courses that have been learned and the courses that have not been learned, and integrated the characteristics of course attributes of course relevance,

Figure 7. Comparison experiment with the neural network baseline: (a) precision@k, (b)recall@k



which proved that considering the impact of course relevance on the course recommendations are necessary.

Other observations include that DeepAE performed better than BPRMF. On the one hand, the ranking optimization model generally uses the inner product calculation method to predict students' preferences and cannot effectively mine the nonlinear interaction between learners and courses. On the other hand, even if DeepAE is an implicit feedback neural network that can use weighted loss, its performance was not as good as the SaeCrdRec model. This is because DeepAE only uses a fully connected neural network structure to model course data without considering distinguishing different historical courses, learner preferences, and course correlation effects. The experimental results verified that the SaeCrdRec model applies the self-attention mechanism encoder to capture the learner's preference and course relevance decoding to obtain the hierarchical relationship of the course, which can improve the evaluation performance of the deep neural structure model to a certain extent.

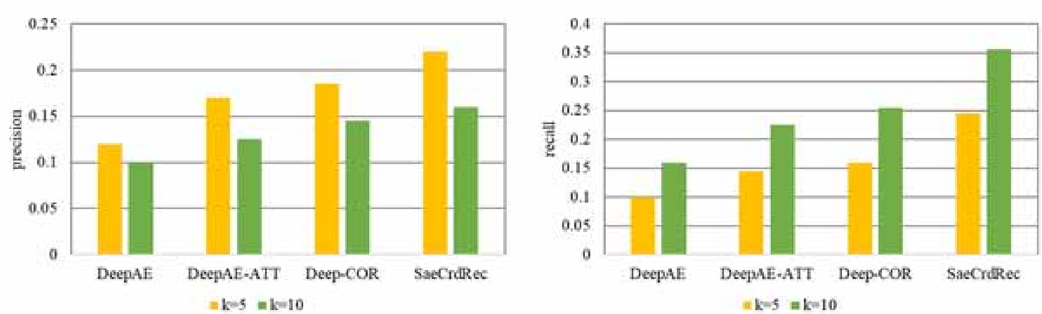
Deep learning can deeply describe the internal information of massive data and can well show the characteristics of data. The model has many levels, parameters, and sufficient capacity. A personalized learning resource recommendation system is an effective way to improve the learning effect and realize personalized learning, which can fully reflect the learner-centered concept.

CONCLUSION

The progress of internet technology has also promoted the development of online education. More and more users are utilizing online platforms to achieve their desire for remote online learning and can access a large number of learning resources without leaving their homes. However, in the era of big data, information overload is prone to occur. As one of the effective methods to deal with information overload, recommendation systems can efficiently screen important content from a large number of learning resources and provide personalized online course recommendation services for learners.

This study highlights the challenge of information overload through in-depth research on the continuous development of internet technology and the rise of online education. We focus on the key role of recommendation systems in solving this problem and apply them to online learning platforms. By combining techniques, such as recommendation systems, word vector embedding, and neural networks, we have successfully constructed an efficient course recommendation model. The experiment showed that the model significantly improves the recommendation effect and provides learners with more satisfactory personalized online course recommendation services. However, there are still some future research directions worth exploring. Future research can further optimize the performance of recommendation systems, while considering more factors such as learners' personal interests and learning styles. In addition, research can also be conducted on how to cope with the

Figure 8. Variant model experimental comparison: (a) precision@k, (b) recall@k



increasing number of online learning resources to ensure that learners can easily access high-quality content. These efforts will help promote the development of online learning and provide learners with broader and richer learning opportunities. In the information age, recommendation systems will continue to play a crucial role in the field of education, filling learning gaps and promoting knowledge accumulation for individuals and society.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING STATEMENT

This work was supported by the Project Approval Authority: Practical research on improving students' ability of career planning under the background of employment orientation (project number: 20220611).

REFERENCES

Adedoyin, O. B., & Soykan, E. (2023). Covid-19 pandemic and online learning: The challenges and opportunities. *Interactive Learning Environments*, *31*(2), 863–875. doi:10.1080/10494820.2020.1813180

Al-Emran, M., Al-Nuaimi, M. N., Arpaci, I., Al-Sharafi, M. A., & Anthony, B. Jr. (2023). Towards a wearable education: Understanding the determinants affecting students' adoption of wearable technologies using machine learning algorithms. *Education and Information Technologies*, *28*(3), 2727–2746. doi:10.1007/s10639-022-11294-z

Al Shloul, T., Javeed, M., Gochoo, M., Alsuhibany, S. A., Ghadi, Y. Y., Jalal, A., & Park, J. (2023). Student's health exercise recognition tool for E-learning education. *Intelligent Automation and Soft Computing, 35*(1), 149–161. https://doi.org/

Alsabhan, W. (2023). Student cheating detection in higher education by implementing machine learning and LSTM techniques. *Sensors (Basel)*, *23*(8), 4149. doi:10.3390/s23084149 PMID:37112489

Amini, M., & Rahmani, A. (2023). Agricultural databases evaluation with machine learning procedure. *Australian Journal of Engineering and Applied Science*, *8*, 39–50. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4331902

Bang, H. J., Li, L., & Flynn, K. (2023). Efficacy of an adaptive game-based math learning app to support personalized learning and improve early elementary school students' learning. *Early Childhood Education Journal*, *51*(4), 717–732. doi:10.1007/s10643-022-01332-3

Bharadiya, J. P. (2023). Machine learning and AI in business intelligence: Trends and opportunities. *International Journal of Computer*, *48*(1), 123–134. https://www.researchgate.net/publication/371902170_Machine_Learning_and_AI_in_Business_Intelligence_Trends_and_Opportunities

Cardona, T., Cudney, E. A., Hoerl, R., & Snyder, J. (2023). Data mining and machine learning retention models in higher education. *Journal of College Student Retention*, *25*(1), 51–75. doi:10.1177/1521025120964920

Chen, S., & Ding, Y. (2023). A machine learning approach to predicting academic performance in Pennsylvania's schools. *Social Sciences (Basel, Switzerland)*, *12*(3), 118. doi:10.3390/socsci12030118

Dolezal, J. M., Wolk, R., Hieromnimon, H. M., Howard, F. M., Srisuwananukorn, A., Karpeyev, D., Ramesh, S., Kochanny, S., Kwon, J. W., Agni, M., Simon, R. C., Desai, C., Kherallah, R., Nguyen, T. D., Schulte, J. J., Cole, K., Khramtsova, G., Garassino, M. C., Husain, A. N., & Pearson, A. T. et al. (2023). Deep learning generates synthetic cancer histology for explainability and education. *NPJ Precision Oncology*, *7*(1), 49. doi:10.1038/s41698-023-00399-4 PMID:37248379

Elglaly, Y. N., & Liu, Y. (2023). Promoting machine learning fairness education through active learning and reflective practices. *ACM SIGCSE Bulletin, 55*(3), 4–6.

Guleria, P., & Sood, M. (2023). Explainable AI and machine learning: Performance evaluation and explainability of classifiers on educational data mining inspired career counseling. *Education and Information Technologies*, *28*(1), 1081–1116. doi:10.1007/s10639-022-11221-2 PMID:35875826

Hussain, S., & Khan, M. Q. (2023). Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. *Annals of Data Science*, *10*(3), 637–655. doi:10.1007/s40745-021-00341-0

Johnson, C. C., Walton, J. B., Strickler, L., & Elliott, J. B. (2023). Online teaching in K-12 education in the United States: A systematic review. *Review of Educational Research*, *93*(3), 353–411. doi:10.3102/00346543221105550

Jurgensmeier, K., Till, S. E., Lu, Y., Arguello, A. M., Stuart, M. J., Saris, D. B., Camp, C. L., & Krych, A. J. (2023). Risk factors for secondary meniscus tears can be accurately predicted through machine learning, creating a resource for patient education and intervention. *Knee Surgery, Sports Traumatology, Arthroscopy*, *31*(2), 518–529. doi:10.1007/s00167-022-07117-w PMID:35974194

Kang, X., & Zhang, W. (2023). An experimental case study on forum-based online teaching to improve student's engagement and motivation in higher education. *Interactive Learning Environments*, *31*(2), 1029–1040. doi:10.1080/10494820.2020.1817758

Kuadey, N. A., Mahama, F., Ankora, C., Bensah, L., Maale, G. T., Agbesi, V. K., Kuadey, A. M., & Adjei, L. (2023). Predicting students' continuance use of learning management system at a technical university using machine learning algorithms. *Interactive Technology and Smart Education*, *20*(2), 209–227. doi:10.1108/ITSE-11-2021-0202

Kubsch, M., Krist, C., & Rosenberg, J. M. (2023). Distributing epistemic functions and tasks—A framework for augmenting human analytic power with machine learning in science education research. *Journal of Research in Science Teaching*, *60*(2), 423–447. doi:10.1002/tea.21803

Liang, Z., Liu, Z., Shi, H., Chen, Y., Cai, Y., Hong, H., Liang, Y., Feng, Y., Yang, Y., Zhang, J., & Fu, P. (2023). SPOC learner's final grade prediction based on a novel sampling batch normalization embedded deep neural network method. *Multimedia Tools and Applications*, *82*(7), 9843–9853. doi:10.1007/s11042-022-13628-y

Martins, R. M., & Gresse Von Wangenheim, C. (2023). Findings on teaching machine learning in high school: A ten-year systematic literature review. *Informatics in Education, 22*(3), 421–440.

Safarov, F., Kutlimuratov, A., Abdusalomov, A. B., Nasimov, R., & Cho, Y. I. (2023). Deep learning recommendations of E-education based on clustering and sequence. *Electronics (Basel)*, *12*(4), 809. doi:10.3390/electronics12040809

Sanusi, I. T., Oyelere, S. S., Vartiainen, H., Suhonen, J., & Tukiainen, M. (2023). A systematic review of teaching and learning machine learning in K-12 education. *Education and Information Technologies*, *28*(5), 5967–5997. doi:10.1007/s10639-022-11416-7

Sofi-Karim, M., Bali, A. O., & Rached, K. (2023). Online education via media platforms and applications as an innovative teaching method. *Education and Information Technologies*, *28*(1), 507–523. doi:10.1007/s10639-022-11188-0 PMID:35791317

Tzavara, A., Lavidas, K., Komis, V., Misirli, A., Karalis, T., & Papadakis, S. (2023). Using personal learning environments before, during and after the pandemic: The case of "e-me". *Education Sciences*, *13*(1), 87. doi:10.3390/educsci13010087

Wong, R. (2023). When no one can go to school: Does online learning meet students' basic learning needs? *Interactive Learning Environments*, *31*(1), 434–450. doi:10.1080/10494820.2020.1789672

Zhengmeng Xu was born in Yancheng, Jiangsu Province, P.R. China. He is a postdoc researcher at Industrial and Commercial Bank of China. His current research interests mainly focus on deep learning, reinforcement learning, and quantum computing algorithms.

Hai Lin was born in Baoding, Hebei Province, P.R. China. He is an undergraduate at the School of Big Data Science, Hebei University of Finance. His research interests include reinforcement learning, deep learning, quantum computing, and financial technology.