Research on Personalized Course Recommendation Algorithm Based on Att-CIN-DNN under Online Education Cloud Platform

Xiaoqiang Liu\textsuperscript{1,2,*} and Feng Hou\textsuperscript{1,2}

Abstract

A personalized course recommendation algorithm based on deep learning in an online education cloud platform is proposed to address the challenges associated with effective information extraction and insufficient feature extraction. First, the user potential preferences are obtained through the course summary, course review information, user course history, and other data. Second, by embedding, the word vector is turned into a low-dimensional and dense real-valued vector, which is then fed into the compressed interaction network-deep neural network model. Finally, considering that learners and different interactive courses play different roles in the final recommendation and prediction results, an attention mechanism is introduced. The accuracy, recall rate, and F1 value of the proposed method are 0.851, 0.856, and 0.853, respectively, when the length of the recommendation list K is 35. Consequently, the proposed strategy outperforms the comparison model in terms of recommending customized course resources.

Keywords

Attention Mechanism, Big Data, Course Resource Recommendation, Deep Learning, Online Education Cloud Platform, Personalized Recommendation

1. Introduction

Premier Li Keqiang proposed the new requirements of “developing 'Internet + education' and promoting high-quality resource sharing” in the 2019 Government Work Report. Well-known domestic platforms include China MOOC University, Tencent Classroom, and NetEase Open Class. Numerous well-known platforms have also been established abroad, such as Coursera, edX, and TED [1-6]. As of June 2019, the number of online courses in China exceeded 125 million, with more than 232 million learning users. The scale and application range of online education in China has gradually increased to the first position globally [7-9].

Online learning platforms offer numerous educational courses, and students independently select their courses based on their interests. Enormous online courses have resulted in course selection problems [10-12]. The completion rate of the Coursera platform is less than one-tenth, and the maximum completion rate is only 40% [13-18].

\hspace{1cm} (*) This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

\hspace{1cm} * Corresponding Author: Xiaoqiang Liu (luxiaqiang@smxpt.edu.cn)

\hspace{1cm} \textsuperscript{1} College of Applied Engineering, Henan University of Science and Technology, Sanmenxia, China (luxiaqiang@smxpt.edu.cn, 694788839@qq.com)

\hspace{1cm} \textsuperscript{2} Sanmenxia Polytechnic, Sanmenxia, China (luxiaqiang@smxpt.edu.cn, 694788839@qq.com)
Although online learning platforms have developed rapidly over the past decade, relevant technology has continuously improved. Relevant websites, course resources, and online interaction methods have become increasingly rich. However, the quality and efficiency of online learning are generally low and several factors may account for this. First, because the number of existing online platforms has increased, the types of learning resources provided by each platform are rich, the degree of difficulty of the course resources differs, and the quality of the courses is uneven, learners find it difficult to choose from the numerous available resources. Second, the existing online learning resource system generally transfers classroom activities to the internet in a mechanical manner, without analyzing users' learning situations or considering learners' personalized needs. Thus, these platforms offer generalized activities and the rationality of learning content design and hypertext link jumps between pages is rarely considered. Third, a large number of resources are available for each learning platform. Learners seeking to discover learning resources that meet their intentions must examine the directory of each platform, which complicates the search for curriculum resources that are strongly related to learners' knowledge backgrounds [19,20].

In summary, this study makes three contributions:

1) Current online learning resource recommendation algorithms do not consider the fact that different courses have different degrees of influence on the prediction results. Most existing studies only consider the low-order and linear relationships between projects but do not consider the nonlinear and high-order relationships between courses.

2) To highlight important features, we introduce an attention mechanism that focuses on important information with high weights, ignores irrelevant information with low weights, and constantly adjusts the weights so that important information can also be selected in different situations, thus providing higher scalability and robustness.

3) To examine the relationship between features, this study uses a combination of compressed interaction network (CIN) and deep neural network (DNN) to generate feature interactions not only explicitly within a certain order but also implicitly for arbitrary low- and high-order feature interactions.

The remainder of this paper is organized as follows. Section 2 discusses the current state of existing literature. Section 3 introduces the proposed Att-CIN-DNN. Section 4 presents a comprehensive experimental comparison of the proposed models. Finally, Section 5 summarizes the study and discusses future research.

2. Related Research

The singular value decomposition (SVD)-based personalized recommendation algorithm is essentially a matrix decomposition operation. However, SVD has two fatal shortcomings. First is the complexity associated with sparse scoring matrix and high-dimensional matrix decomposition calculations. Deep learning creates nonlinear network topologies with numerous hidden layers to approximate complex functions.

Deep learning technology provides new ideas and solutions for the further improvement and development of personalized recommendation systems. In [21], the authors use a superimposed denoising self-coder for feature extraction. Naumov et al. [22] develops a deep learning recommendation
model that uses data parallelism to perform full-connection-layer calculations. Wang et al. [23] provides a multi-task feature learning method for knowledge map augmentation recommendations using a knowledge map as the source of supplemental information. A knowledge map is used to embed tasks to help recommend a deep end-to-end task framework. Nisha and Mohan [24] proposes a hybrid method for a depth self-coder. Guo et al. [25] proposes a neural network framework integrating a factorization machine and DNN. The relevance of the course is calculated in [26] by assessing the external attribute tolerance and internal attribute quality value, and a linear discriminant analysis user interest model is subsequently constructed to compute the user's choice for the topic. In [14], the authors presents a personalized online course recommendation click-through rate model. Amane et al. [27] proposes a dynamic ontology-based e-learning recommendation system. The proposed recommendation method semantically describes the course and learner and is integrated into collaborative and content-based filtering techniques to generate the top-N recommendations using clustering methods. The results demonstrate the effectiveness of the proposed approach in the recommendation process.

However, the techniques mentioned above can only extract shallow information from data sources. An online education cloud-platform-based personalized course recommendation system built on deep learning was suggested as a solution to these problems. The word vector is transformed into a real-value vector using data such as course summary, course review information, and historical courses learned by users, and is then input into the CIN-DNN method using embedding technology. This significantly enhances the system's capacity for feature extraction. An attention layer is introduced to highlight the significance of various past interactions, considering that learners and various interactive courses play diverse roles in the final recommendation prediction results of the model.

### 3. Proposed Personalized Course Recommendation Algorithm

#### 3.1 Att-CIN-DNN Personalized Course Recommendation Model

![Att-CIN-DNN model architecture.](image)

**Fig. 1.** Att-CIN-DNN model architecture.
Compared with the traditional deep collaborative filtering (DCF) model, the proposed Att-CIN-DNN first introduces an attention layer to distinguish the significance of different historical interactions. For example, the course records of embedded principles and hardware development that exist in the historical interactions of learners who primarily study database technology are relatively less important for the recommendation results. The combination of CIN and DNN is then utilized to generate feature interactions not only explicitly within a certain order but also implicitly for arbitrary low- and high-order feature interactions. In addition, considering that learners and different interaction courses play different roles in the final recommendation and prediction results, the Att-CIN-DNN method, as shown in Fig. 1, is employed. This indicates that the major difference between the Att-CIN-DNN model and the DCF model is the addition of a layer of attention mechanism network and CIN-DNN network after the pooling layer.

3.2 Attention Model

Data mining, image segmentation, target recognition, recommendation systems, and other fields have benefited from the application of the attention model (Att), which first emerged in the field of machine translation and has been steadily applied to deep learning in recent years. Modeling user behavior data is the most frequently used attention method in recommendation systems. Typically, the attention module’s input comprises a value, key, and query; its output is the weighted sum of the similarities between the key, query, and value. The attention mechanism assigns varying weights to feature interactions depending on their significance. The three elements are identical and comprise the features of previous user interactions. The structure of the attention mechanism is shown in Fig. 2.

$$Attention(Q, K, V) = softmax\left(\frac{QR^T}{\sqrt{d}}\right)V,$$

where the query matrix is called $Q$; the key matrix is called $K$; and the numerical matrix is called $V$.

3.3 CIN

The feature interaction sequence is determined by the number of network layers, which makes the CIN unique. To enable the output unit to view the feature interaction mode in various orders, each hidden layer is linked by pooling with the output layer. In contrast to the CIN, which uses fixed additional input data, the recurrent neural network (RNN) uses additional variable input data. Fig. 3 summarizes the macro framework.
3.4 DNN

As shown in Fig. 4, a deep neural network was expanded based on the perceptron model by including multiple inputs, hidden layers, and outputs.

4. Experiment and Analysis

4.1 Experimental Environment

This study uses the BertBase Chinese model, which contains 110 million parameters, a 12-layer transformer network structure, 12-head attention mode, and 768-dimensional hidden unit. Other training parameters were as follows: learning rate of 0.001, drop-out rate of 0.1, 10 iterations, and a mini-batch
of 32. In the dataset, the average number of words in the course outline description was 240, whereas that of the course review text was 154. Because CIN and DNN networks require fixed input dimensions, BERT fine-tuned the vector of text information conversion to 200 dimensions. The parameter settings are listed in Table 1.

Table 2 shows the primary hardware and software setup used in this investigation.

### 4.2 Experimental Dataset

Information from the course selection experiment utilized in this study was obtained from cmooc.com. The course selection data of students only within the range of \([10, 510]\) were considered to reduce the effect of noisy data produced by students who selected fewer courses and too many (no choice) courses in the experiment. Table 3 presents the sampling results.

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>Parameter value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning rate</td>
<td>0.001</td>
</tr>
<tr>
<td>Mini-batch</td>
<td>1024</td>
</tr>
<tr>
<td>Iterations</td>
<td>20</td>
</tr>
<tr>
<td>DNN layers</td>
<td>3</td>
</tr>
<tr>
<td>Neurons per layer</td>
<td>[120,120]</td>
</tr>
<tr>
<td>Activation function</td>
<td>ReLU</td>
</tr>
<tr>
<td>CIN layers</td>
<td>4</td>
</tr>
<tr>
<td>Neurons in each layer of CIN</td>
<td>[120,120,60]</td>
</tr>
<tr>
<td>CIN activation function</td>
<td>Identity function</td>
</tr>
<tr>
<td>Output dimension</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experimental environment</th>
<th>Specific information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating system</td>
<td>Windows</td>
</tr>
<tr>
<td>Graphics card</td>
<td>GTX 1080Ti</td>
</tr>
<tr>
<td>Memory</td>
<td>64 GB</td>
</tr>
<tr>
<td>Language</td>
<td>Python3.6</td>
</tr>
<tr>
<td>Development platform</td>
<td>PyTorch</td>
</tr>
<tr>
<td>TensorFlow</td>
<td>2.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total number of interactions</th>
<th>Average number of interactions</th>
<th>Sparsity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse dataset</td>
<td>17,967</td>
<td>43</td>
</tr>
<tr>
<td>Relatively sparse dataset</td>
<td>21,309</td>
<td>49</td>
</tr>
<tr>
<td>Relatively dense dataset</td>
<td>30,981</td>
<td>61</td>
</tr>
<tr>
<td>Dense dataset</td>
<td>53,982</td>
<td>120</td>
</tr>
</tbody>
</table>

### 4.3 Evaluating Indicator

The accuracy, recall rate, and F1 values were used to assess the performance index of the method.
Precision = \frac{TP}{TP + FP}, \quad (2)
Recall = \frac{TP}{TP + FN}, \quad (3)
F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \quad (4)

where TP denotes that the prediction is positive and correct; FP denotes a positive but inaccurate forecast; FN denotes a prognosis that is negative but incorrect; and TN denotes a forecast that is both negative and accurate.

In addition, online learning resource recommendation mainly presents the most relevant course resources to learners, namely, the top-k course list recommendation. To evaluate the recommendation list, we focus on the following two points: first, place the results that learners are interested in at the top of the list because learners are used to browsing from top to bottom; second, ensure that the entire recommendation list has a strong correlation with learners’ interests. Therefore, this study also used hit rate (HR) and normalized discounted cumulative gain (NDCG) as evaluation indicators for the algorithm.

The HR indicator formula is shown in Eq. (5).

\[ HR@K = \frac{NH@K}{|GT|} \times 100\% \] \quad (5)

where \( NH@K \) represents the sum of the predicted number of test sets in the top-k list of each learner and \( |GT| \) represents the hit ratio.

The NDCG formula is shown in Eq. (6).

\[ NDCG_k = \frac{\sum_{i=1}^{k} 2^{rel_i} - 1}{\sum_{i=1}^{k} \frac{1}{\log_2(i+1)}} \] \quad (6)

where, \( rel_i \) represents the correlation of the recommendation results in the position \( i \), and \( k \) is the value in top-k. The range of value for \( NDCG_k \) is between [0, 1].

4.4 Model Training

4.4.1 Influence of regularization coefficient and iteration number on the loss rate of model

For the model to acquire anti-noise capacity and improve its prediction performance for unknown samples, regularization is adopted; however, it can inhibit overfitting to some extent. In this study, we used loss minimization to direct the neural network prediction to maintain its approximation in the real world.

The algorithm requires a high initial attribute value setting, which has a direct impact on the effectiveness of the recommendations and the training time of the algorithm. The results of the recommendation process are significantly influenced by the regularization coefficients and iterations. Both the loss rate and recommendation effect were best when lambda of 0.01. The loss rate gradually increases with the number of repetitions, and the recommendation effect is correspondingly affected. As
shown in Fig. 5, the recommendation effect of the algorithm is optimal when parameters are fixed, such as rank of 15.

![Fig. 5. Effect of regularization coefficient and iteration number on the loss rate of model.](image)

4.4.2 Effect of datasets with different sparsity on model performance

The impact of test datasets with various levels of sparsity on the model performance is depicted in Fig. 6. The sparse dataset had the lowest values under the evaluation indexes F1, HR, and NDGG, with values of only 0.67, 0.50, and 0.25, respectively. The model performs poorly under the sparse dataset, and the model's performance gradually improves with a gradual increase in the data density; however, the increase in the data does not mean an increase in effective information. Excessively dense data sets lead to a decrease in the model index. The model is effectively learned with denser datasets and with F1, HR, and NDGG reaching 0.751, 0.70, and 0.352, respectively. The sparsity of the dataset had a significant effect on the model.

![Fig. 6. Effect of datasets with different sparsity on model performance.](image)
4.5 Experimental Comparison

4.5.1 Comparison of accuracy and recall of different algorithms

To demonstrate the efficacy of the proposed personalized recommendation algorithm, this section compares the literature [14,26,27] with the proposed personalized curriculum resource recommendation method. A comparison of the recall and accuracy rates of the four models is shown in Figs. 7 and 8, respectively. Figs. 7 and 8 show that the proposed method significantly enhances the recommendation accuracy owing to the multiple types of input features, strong correlation, and the deep confidence network’s feature learning ability. However, this study believes that learners and different interactive courses play different roles in the final recommendation and prediction results of the model and hence, introduces the attention layer to distinguish the importance of different historical interactions.

![Fig. 7. Overall accuracy comparison.](image)

![Fig. 8. Overall recall comparison.](image)

4.5.2 Comparison of F1 values of different models

In practical applications, classification accuracy is affected by the $K$ value (length of the recommendation list); therefore, it is necessary to select multiple $K$ values to evaluate the recommendation system. The range of $K$ values used in this study was [10,20,30,40]. Through multiple evaluation parameters, the precision and recall were compared under different parameters, and a better model was obtained. Because errors occur when the recommendation system is evaluated solely by recall or precision, precision and recall are combined to verify the recommendation performance.

The F1 value is also used to evaluate the recommendation algorithm. Combined with the comprehensive index F1 value, the hybrid recommendation method designed in this study is effective and has a good
effect when $K > 10$. The recall rate and accuracy were higher than those of the other two models. The mixed recommendation model reached the optimal value at $K = 30$, and the accuracy rate at this time was an F1-score of 0.853.

The deep learning-based tailored course recommendation system performed superiorly at $K = 30$ with a higher F1 value than the other three recommendation models. The outcomes are shown in Fig. 9.

Fig. 9. Comparison of F1 value for the recommended model.

Fig. 10. Comparison of HR indicators of different algorithms.

Fig. 11. NDGG index comparison of different algorithms.
4.5.3 Comparison of HR and NDCG of different algorithms

Figs. 10 and 11 show that the method in [26] did not perform well for the HR and NDCG assessment indicators. The reasons for this analysis may be as follows. First, convolutional neural networks (CNNs) have more advantages in image processing, and second, the most significant feature of convolution in Conv NCF is translation invariance. It uses the outer product operation to replace Neu CF’s point multiplication to extract linear features, forms a two-dimensional interactive map as the CNN's input, and then utilizes a CNN to obtain the predicted value, which has certain limitations on model learning interactive sequence features. Third, this model may have some limitations in terms of transfer ability and generalization ability and is not suitable for personalized recommendations of online learning resources. The method in this study not only considers the auxiliary information and the relative order of learning but also combines the attention mechanism to further distinguish the contribution of different interactive courses to the recommendation results, which is higher than the HR and NDCG indicators in literature [14,26,27]. This implies that by combining the more fine-grained attention mechanism, the method in this study can further distinguish the different contributions of historical interactive records to decision-making, which is conducive to learning more features.

![Fig. 12](attachment:fig12.png)

(a) Precision, (b) recall, and (c) F1 value of different algorithmic recommended models.

4.5.4 Influence of recommended courses $K$ on various indicators of different algorithms

This section uses the MOOC userlabel08rl dataset of China University to compare the references [14,26,27] with the proposed personalized curriculum resource recommendation algorithm. Figs. 12 and 13 show that the suggested algorithm’s index was the best under the length of all recommendation lists on the userlabel08rl dataset. When the recommendation list length $N$ was 35, the proposed algorithm had a precision rate of 0.851, a recall rate of 0.856, and an F1 value of 0.853, all of which were greater than
those in the references. This is because the proposed algorithm considers that learners and different interactive courses play different roles in the final recommendation prediction results of the model and introduces an attention layer to distinguish the importance of different historical interactions, which improves the accuracy of personalized course recommendations.

![Fig. 13. Comparison of evaluation indicators of different algorithms: (a) HR and (b) NDCG.](image_url)

### 4.5.5 Ablation experiment

Table 4 presents the results of the ablation experiments performed in this study. The experimental results indicate that the Att-CIN-DNN model can effectively model users with at least 0.053 and 0.011 improvements in recall and NDCG metrics, respectively. In addition, the results also shows that the attention network can assign more attention weights to relatively important short-term interaction sequences. This indicates that the proposed algorithm can efficiently mine hidden information in the interactive sequence and improve the weight of the courses in the sequence, which causes the model to accurately recommend personalized courses according to the current status of users, making the recommendations more timely and dynamic.

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall</th>
<th>NDCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIN-DNN</td>
<td>0.759</td>
<td>0.312</td>
</tr>
<tr>
<td>Att-CIN</td>
<td>0.776</td>
<td>0.334</td>
</tr>
<tr>
<td>Att-DNN</td>
<td>0.803</td>
<td>0.341</td>
</tr>
<tr>
<td>Att-CIN-DNN</td>
<td>0.856</td>
<td>0.352</td>
</tr>
</tbody>
</table>

### 5. Conclusion

This study proposes a personalized course recommendation algorithm based on deep learning for an online education cloud platform to address the challenges associated with effectual information extraction and insufficient feature extraction. Experiments validated the efficacy of this approach. The conclusions drawn from the analysis and description presented above are listed below.

1) Using a CIN-DNN network can help the model explore the potential links between features and learn explicit and implicit higher-order feature interactions.
2) The use of attention mechanisms can rationalize the different contributions in the process of predicting targets and improve the accuracy of recommendations. Although the system considers numerous user interactions as the foundation for analysis, additional implicit and significant interactions are present between users and websites, such as frequent clicking behaviors and the length of time spent watching a course. The model can gather and analyze this component of behavior to assist with future research. However, as consumers gain knowledge and experience, their interests and preferences are not static but rather dynamic; therefore, to anticipate user preference changes, the proposed approach must be able to track them dynamically.

Conflict of Interest

The authors declare that they have no competing interests.

Funding

None.

References


Research on Personalized Course Recommendation Algorithm Based on Att-CIN-DNN under Online Education Cloud Platform


Xiaoqiang Liu  https://orcid.org/0000-0003-3403-5562

He has a master's degree in computer science and is now an associate professor. He graduated from Wuhan University in 2009. He once worked in College of Applied Engineering, Henan University of Science and Technology. His research interests include network system algorithms and applications.

Feng Hou  https://orcid.org/0000-0002-2917-8203

She is a master of Computer Science with the title of professor. She graduated from Henan University of Science and Technology in 2010. She used to work in College of Applied Engineering, Henan University of Science and Technology. Her research interests include big data analysis and computer application.