IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 February 2024, accepted 3 June 2024, date of publication 11 June 2024, date of current version 18 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3412114

RESEARCH ARTICLE

Research on Knowledge Concept Recommendation Algorithm With Spatial–Temporal Information

ZHAOYU SHOU[1,2], YIXIN CHEN[1], HUIBING ZHANG[3], AND JIANWEN MO[1]

[1]School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China
[2]Guangxi Wireless Broadband Communication and Signal Processing Key Laboratory, Guilin University of Electronic Technology, Guilin 541004, China
[3]School of Computer and Information Security, Guilin University of Electronic Technology, Guilin 541004, China

Corresponding author: Zhaoyu Shou (guilinshou@guet.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62177012,Grant 61967005, and Grant 62267003; in part by Guangxi Natural Science Foundation under Grant 2024GXNSFDA010048; and in part by the Project of Guangxi Wireless Broadband Communication and Signal Processing Key Laboratory under Grant GXKL06240107.

ABSTRACT Online learning is an important complementary form of offline classroom learning, in order to meet the personalized learning needs of students in the online learning process and improve the effectiveness of online learning, this paper proposes a knowledge concept recommendation algorithm based on spatio-temporal information. The algorithm models students from both the spatial and temporal dimensions to accurately recommend a personalized knowledge concept learning list. Spatially, a heterogeneous information network (HIN) based on the MOOC platform is constructed, and an improved graph convolutional network TSA-GCN is used to learn the representation of students and knowledge concepts under the meta-paths, which can adaptively aggregate the information of neighboring nodes through the trainable self-weight adjacency matrix, fully taking into account the differences in the student's perceptual ability, and using the attention mechanism to fuse the information under multiple meta-paths to ensure the information integrity. Temporally, Attention RNN (Attention RNN, ARNN) is used to learn students' temporal learning behaviors, mine the interest offset features in the learning process, and predict students' current learning interests. The spatial and temporal information is fed into the extended matrix factorization model to generate the final knowledge concept recommendation list. Experiments on publicly available MOOC datasets show that the method proposed in this paper can more accurately predict and recommend the knowledge concepts that students are interested in compared to the latest proposed methods.

INDEX TERMS Knowledge concept recommendation, graph convolutional networks, recurrent neural networks, online learning.

I. INTRODUCTION

While the spread of higher education promotes the modernization of education, it also stimulates the exploration and implementation of high-quality education and teaching models. The construction of information technology-based online learning spaces and the use of information to measure educational outcomes and the teaching process have become powerful means of improving the quality of higher education.

The associate editor coordinating the review of this manuscript and approving it for publication was Mehedi Masud.

Massive Open Online Courses (MOOCs) are widely used by a large number of students for their geographical and time-independent features [1]. However, the diversity of MOOC courses and their frequent changes bring problems such as knowledge disorientation and information overload to online learning [2]. In order to stimulate students' interest and meet the needs of students with different backgrounds for customized personalized learning materials and continuous online learning, establishing an accurate personalized recommendation system on MOOC platforms has become the focus and hotspot of current research [3]. Diao et al. [4]

proposed a personalized learning path recommendation method based on weak concept mining to discover valuable learning paths from online learning data and provide effective online learning references for subsequent learners. Tian and Liu [5] extended the multidimensional item response theory (MIRT) as a capacity tracing model for the first time to integrate into recommendation models in MOOCs, which improves the effectiveness and interpretability of MOOCs. Chuang et al. [6] designed a reinforcement learning-based exercise recommendation system to recommend personalized exercises of appropriate difficulty and knowledge concepts for learners based on the data recorded by the system. Shrimali et al. [7] proposed a video recommendation model based on natural language processing to video based on the similarity between the video text and the query semantics for video recommendation. However, directly recommending courses or video resources cannot accurately reflect students' fine-grained learning status of knowledge points and ignore the differences in students' interests in the same learning resources. To solve the above problems, Ju et al. [8] proposed a knowledge concept recommendation model based on local subgraph embedding, which utilizes attentional graph convolution to fuse contextual information of different subgraphs and capture complex semantic relationships among entities. Wang et al. [9] proposed an end-to-end graph neural network-based approach for knowledge concept recommendation that uses graph convolution to learn the representation of entities under different meta-paths and adaptive fusion with the help of an attention mechanism to produce more coherent and personalized knowledge concept recommendation results.

The above work has achieved some results, however, there are still deficiencies in student modeling: 1) Existing recommendation models tend to ignore the temporal nature of students' learning behaviors and mine information by considering the correlation information between entity nodes such as students, knowledge concepts, and videos as a whole and mapping them into a high-dimensional space. In this paper, we refer to the information extracted from the modeling method that ignores the temporal order of associations among nodes as spatial information and the node representations learned in high-dimensional space using this method can effectively retain the integrity of heterogeneous information and mine the potential semantic associations among data, but lose the hidden learning preferences, interest biases, and other features of the students' temporal learning process; 2) the method of updating the representation of the student by aggregating information from neighboring nodes in the high-dimensional space does not take into account the differences in the students' perceptual ability, and the more the students are capable of perceiving the stronger the influence of the students by their neighboring nodes in the process of learning.

Based on the above analysis, this study proposes a knowledge concept recommendation model with spatio-temporal information (STKCR). Spatially, the HIN is constructed by considering the correlation information between all heterogeneous nodes in the MOOC platform as a whole, and then learning the representation of students and knowledge concepts in multiple meta-paths through TSA-GCN, and adaptive fusion in high-dimensional space; temporally focus on students' learning paths and use attention RNN to mine students' learning interests based on students' temporal behavioral data, and finally merge the learned student and concept representations into the extended matrix factorization model to achieve accurate and personalized knowledge concept recommendation. This paper mainly includes the following contributions:

First, it effectively combines GCN and RNN, while retaining the integrity of high-dimensional spatial information, it considers the perceptual ability, interest bias, and other characteristics of the student's learning process to achieve more accurate personalized knowledge concept recommendations.

Second, the TSA-GCN model is proposed to simulate the students' ability to perceive external information by constructing a trainable self-weighted adjacency matrix, and adaptively adjusting the influence of neighboring nodes on the students to make the representation of the students more accurate.

Third, a large number of experiments were carried out on a large public MOOC dataset, and the experimental results prove that the STKCR proposed in this paper has certain advantages over the state-of-the-art knowledge concept recommendation algorithms.

The rest of the paper is organized as follows: section II briefly describes the related work of this paper. Section III explains the definitions related to the algorithms of this paper. Section IV describes the proposed method in detail. Section V shows the experimental results and analyzes the algorithm's reliability in this paper. Section VI summarizes the work and gives an outlook.

II. RELATED WORK

In order to preserve the integrity of heterogeneous data and mine potential semantic associations between different nodes, graph structures have been widely used in MOOC resource recommendations. Yu et al. [10] proposed a knowledge concept recommendation algorithm based on adaptive augmented graph contrast learning, which introduces contrast learning to alleviate the interaction imbalance problem of the recommender system. Shou et al. [11] proposed a learning partner recommendation model based on weighted heterogeneous information networks in order to alleviate the loneliness of learners in the online learning process by automatically generating all meaningful meta-paths to extract more complete interaction information and reveal students' unique preferences. Wang et al. [12] proposed a multifaceted heterogeneous information network considering diverse relationships between learners and knowledge concepts and used the Gumbel-Softmax method to dynamically assign aspect context to each node to improve the accuracy

TABLE 1. Experimental environment.

Models	Paper Numbers	Advantages	Limitations
Recommendation models based on graph structures	[10-16]	Effectively preserves the integrity of heterogeneous data and mines its potential correlations in a high-dimensional space	Insufficient attention to temporal information and difficult to scale up
Recommendation models based on tensor decomposition	[17,18]	Suitable for the representation and extraction of spatial potential features of high-dimensional data	Higher memory consumption and less ability to capture sequential information
Reinforcement Learning Based Recommendation Modeling	[19-22]	Great at modeling the dynamic interactions between users and projects to capture long-term user interest	Requires large-scale training and difficult reward function design

of recommendation. Wang et al. [13] in order to explore the higher-order similarity characteristics of learners and knowledge concepts under different meta-paths, proposed a method for recommending knowledge concepts based on heterogeneous information networks and attentional graph convolution, which introduces the nodes and their node neighbors in meta-paths to construct self-supervised signaling. Alatrash et al. [14], in order to alleviate the limitations of graph neural networks in the field of knowledge concept recommendation while mainly related to complexity, semantics, and transparency, proposed an end-to-end framework combining KGs, Graph Convolutional Networks (GCNs), and pre-trained transformer language model encoders (SBERT) in an end-to-end framework aiming to provide users with personalized and transparent recommendations of knowledge concepts. However, recommendation models based on graph structures usually construct interactions between nodes as a whole and map them into a high-dimensional space in order to learn potential associations between nodes, and the above approach neglects the temporal information of interactions between nodes. Ling and Shan [15] proposed a knowledge concept recommendation model based on structurally augmented interacting graph neural networks, which learns the representations of students and knowledge concepts through graph convolution based on meta-path guidance and knowledge concept interaction sub-sequence, respectively, and utilizes the extended matrix factorization to recommended knowledge concepts. Klasnja-Milicevic and Milicevic [16] proposed the NCO-A model to recommend TOP-N knowledge concepts. The model alleviates the data sparsity problem and improves the model's generalization ability by implementing dynamic growth of HIN and selectively identifying implicit features. The above studies have somewhat compensated for the lack of temporal information in graph-structured models, yet the importance of accurately modeling learners from both spatial and temporal dimensions still cannot be ignored.

In recent years, methods such as tensor decomposition and reinforcement learning have been applied to the field of online learning resource recommendation. Liu et al. [17] proposed an incremental tensor-based correlation analysis and personalized recommendation algorithm, which realized the accurate recommendation of learning resources in different environments by multidimensional correlation analysis of educational data through incremental tensor decomposition. Shou et al. [18] proposed a knowledge concept recommendation model based on tensor decomposition and Transformer reordering, which utilizes tensor decomposition to retain the information integrity of the high-dimensional space and uses Transformer to fuse the knowledge concept learning order information. Recommendation models based on tensor decomposition ensure the integrity of heterogeneous information and facilitate the discovery of hidden structures and values from massive data through the form of high-dimensional data, but the process of tensor construction has a considerable complexity and a great demand on computer memory. Gong et al. [19] constructed a knowledge concept recommendation model based on a heterogeneous information network and reinforcement learning, which is capable of automatically identifying effective meta-paths and utilizing a reinforcement learning framework to capture students' long-term interests. Wu et al. [20] designed a deep knowledge preference-aware reinforcement learning network for knowledge concept recommendation, which uses a hierarchical propagation path construction method to explore further paths to reduce the training burden of the reinforcement learning model. Liang et al. [21] combined the strengths of graph convolutional networks and reinforcement learning for learning resource recommendation, where the model learns multipath embedding through graph convolution and uses reinforcement learning to make student-centered suggestions. Gong et al [22] formulated knowledge concept recommendation as a reinforcement learning problem better to model the dynamic interaction between students and knowledge concepts and introduced a heterogeneous information network (HIN) between students, courses, videos, and concepts to alleviate the data sparsity problem in the recommendation task. The recommendation model based on reinforcement learning pays more attention to the temporal information in the interaction between the user and the knowledge concept and improves the modeling of the learner from the dynamic interaction sequences. Still, the data sparsity problem it faces is more severe. The large number of hyperparameters involved in the reward mechanism in reinforcement learning

TABLE 2. Extracting meta-paths for student and knowledge concept.

Type	Meta-path
Student	$SCS: Student \xrightarrow{choose} Course \xrightarrow{chosen-by} Student$
	$SVS: Student \xrightarrow{watch} Video \xrightarrow{watched-by} Student$
	$SKS: Student \xrightarrow{learn} Knowledge\ Concept \xrightarrow{learned-by} Student$
	$SCTCS: Student \xrightarrow{choose} Course \xrightarrow{recorded-by} Teacher \xrightarrow{record} Course \xrightarrow{chosen-by} Student$
Knowledge Concept	$KSK: Knowledge\ Concept \xrightarrow{learned-by} Student \xrightarrow{learn} Knowledge\ Concept$
	$KVCVK: Knowledge\ Concept \xrightarrow{included-in} Video \xrightarrow{contained-in} Course \xrightarrow{contain} Video \xrightarrow{include} Knowledge\ Concept$

is also a major difficulty it faces. Based on the above studies, the summary of the relevant research models is shown in Table 1.

In summary, most of the existing knowledge concept recommendation models fail to comprehensively consider the potential correlation information among nodes and the characteristics of students' perceptual ability and interest bias during the learning process in terms of student modeling. At the same time, the methods based on tensor decomposition and reinforcement learning are also subjected to the constraints of memory and arithmetic power. Therefore, in this paper, based on TSA-GCN, we mine the correlation information among nodes in the high-dimensional space to adaptively update the node representations, and mine the learning interests of students according to their time-series learning behaviors to achieve accurate student modeling and precisely recommend suitable knowledge concepts.

III. RELEVANT DEFINITION

This section explains the relevant definitions and computational methods of the proposed algorithm in order to explain the method proposed in this paper more clearly.

A. HETEROGENEOUS INFORMATION NETWORKS

1) CONSTRUCTING HETEROGENEOUS INFORMATION NETWORKS

A heterogeneous information network [23] is defined as a directed graph $G = (V, E)$ with object type mappings $\phi : V \rightarrow A$ and relation type mappings $\psi : E \rightarrow R$, where the sum of the total number of object types $|A|$ and the total number of relation types $|R|$ is greater than 2. Fig. 1 illustrates a heterogeneous information network constructed based on the MOOC platform, which consists of five object types: student (S), teacher (T), course (C), video (V)and knowledge concept(K), and the 14 relations between them. where R_i denotes the correspondence between different types of objects (R_i^{-1} is denoted as the inverse of R_i), and R_1 $\left(R_1^{-1}\right)$, R_2 $\left(R_2^{-1}\right)$, R_3 $\left(R_3^{-1}\right)$, R_4 $\left(R_4^{-1}\right)$, R_5 $\left(R_5^{-1}\right)$, R_6 $\left(R_6^{-1}\right)$, R_7 $\left(R_7^{-1}\right)$ denotes respectively teach (taught-by), choose (chosen-by), learn (learned-by), record (recorded-by), contain (contained-in), include (included-in), and watch (watched-by).

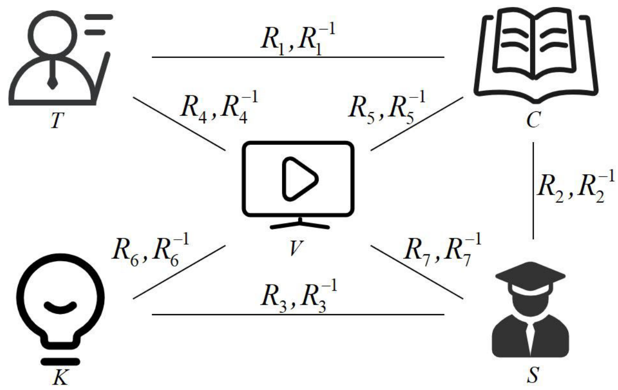


FIGURE 1. Heterogeneous information networks in MOOC platforms.

2) META-PATHS

Meta-paths are defined in heterogeneous information networks G of the form $A_1 \xrightarrow{R_1} A_2 \ldots \xrightarrow{R_l} A_{l+1}$, which can lead to richer and more effective semantics by combining the types of relationships in the network [24]. Table 2 demonstrates the six meta-paths $[MP]$ selected for students and knowledge concepts in this study, where $\{SCS, SVS, SKS, SCTCS\}$ are student meta-paths and $\{KUK, KVCVK\}$ are knowledge concept meta-paths.

B. GRAPH CONVOLUTIONAL NETWORK BASED ON TRAINABLE SELF-WEIGHTED ADJACENT MATRIX (TSA-GCN)

Graph Convolutional Networks (GCNs) learn node representations by aggregating information from neighboring nodes, but their superior performance usually relies on the homogeneity of the network [25]. Therefore, selecting a suitable meta-path using a restriction on the type of head and tail nodes can be used to mine potential associations between nodes of the same type in a heterogeneous information network, and thus learn the semantic representation of a node under that meta-path through GCN.

1) TRAINABLE SELF-WEIGHTED ADJACENT MATRIX

According to the constructed heterogeneous information network $G = (V, E)$, under each meta-path MP an adjacency matrix with Boolean matrix elements $A^{MP} \in R^{N \times N}$ can be obtained. Where N denotes the number of nodes, if $A_{ij}^{MP} = 1$ denotes that the node i can be linked to the node j via the meta-path MP.

The traditional approach to the adjacency matrix is to join the unit matrix ωI as a way of adding its own information during the update of the node representation. Where ω is a trainable parameter, when ω takes the value of 1, it means that the node's own features are as important as those of its neighboring nodes; when ω tends to 0, it means that the node's own features have almost no effect on itself; the larger ω is, the more the node's own features have a greater effect on itself.

However, this adjacency matrix ignores the differences in students' perceptual abilities. In the process of updating the student representation using GCN, the neighbor node information will have different degrees of influence on each student node according to the strength of the student's perceptual ability. In this paper, we construct the trainable self-weighted adjacency matrix as shown in (1):

$$\tilde{P}^{MP} = D^{-1}\left(A^{MP} + diag\left(B\right)\right) \tag{1}$$

where B is the trainable vector, which is diagonalized by the $diag\left(\bullet\right)$ function and added to the adjacency matrix A. D is the degree matrix of the matrix $A^{MP} + diag\left(B\right)$, and D^{-1} is multiplied with the matrix to achieve the normalization of the adjacency matrix. The trainable self-weighted adjacency matrix constructed using the trainable vector B can fully consider the degree of influence of neighboring nodes on this node when updating the node representation and learn a more accurate node representation.

2) NODE REPRESENTATION LEARNING

Given a heterogeneous information network $\mathrm{G} = (\mathrm{V}, \mathrm{E})$ and a meta-path MP, after obtaining their adjacency matrices, this paper learns the node representations using the layer-by-layer propagation rule, as shown in (2):

$$h^{(l+1)} = \mathrm{Re}lu\left(\tilde{P}^{MP} h^{l} \mathrm{W}^{l}\right) \tag{2}$$

where l denotes the number of layers, W^{l} denotes the trainable weight matrix shared by all nodes in layer l, and each layer is activated using the Relu function. In this study the initial features h^0 of students and knowledge concepts are randomly initialized and continuously trained by single layer TSA-GCN, the nodes are represented after single layer TSA-GCN as $e^{MP} = h_{MP}^{1}$. Trained by (2), h^0 can be passed to any node.

In the heterogeneous information network G, $S = \{s_1, s_2, \cdots, s_i, \cdots\}$ denotes the set of students and $|S|$ is the number of students. An example of student representation learning based on TSA-GCN is shown in Fig. 2.

C. ATTENTION MECHANISMS

In deep learning, the introduction of an attention mechanism enables neural networks to automatically learn and select

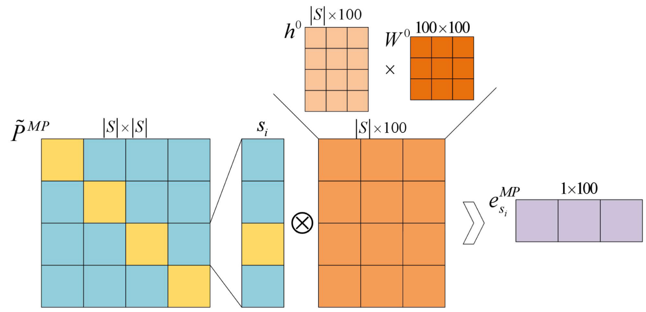


FIGURE 2. Example of student representation learning.

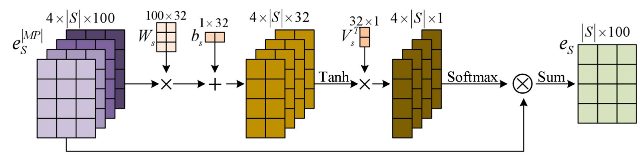


FIGURE 3. Visualization of attention mechanisms incorporating student representations.

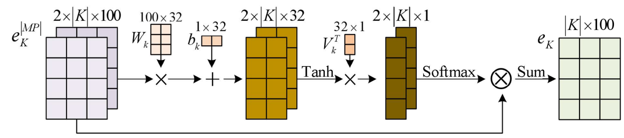


FIGURE 4. Visualization of the conceptual representation of attention mechanisms integrating knowledge.

the important information in the input, improving the performance and generalization of the model [26]. In this paper, the node representations learned from different meta-paths have different importance to the nodes, thus the attention mechanism is used to fuse the information under multiple meta-paths to generate the final node representations.

Taking the fusion of student representations under four student meta-paths as an example, the sequence of student representations $\left\{e_s^{SCS}, e_s^{SVS}, e_s^{SKS}, e_s^{SCTCS}\right\}$ output from TSA-GCN is taken as input, and the formula for calculating the attentional weight of each meta-path is:

$$\alpha_s^{MP_i} = \frac{\exp\left(\mathrm{V}_s^T \sigma\left(\mathrm{W}_s e_s^{MP_i} + b_s\right)\right)}{\sum_{j\in[MP]} \exp\left(\mathrm{V}_s^T \sigma\left(\mathrm{W}_s e_s^{MP_j} + b_s\right)\right)} \tag{3}$$

where V_s^T, W_s, b_s is the trainable matrix, $\sigma\left(\bullet\right)$ is the activation function tanh, and the output $\alpha_s^{MP_i}$ represents the weights of the student representations $e_s^{MP_i}$ under the meta-path MP_i. Based on the obtained weights, the student representations under multiple meta-paths are fused:

$$e_u = \sum_{j\in[MP]} \alpha_s^{MP_j} e_s^{MP_j} \tag{4}$$

where e_s is the final student representation, and the visualization of the attention mechanism incorporating the student representation is shown in Figure 3.

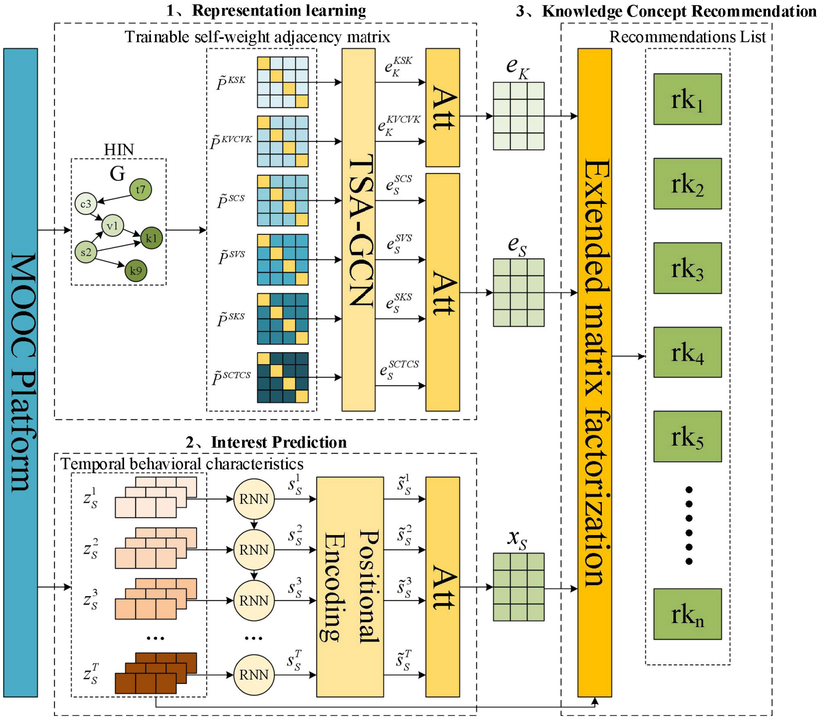


FIGURE 5. STKCR model framework diagram.

Similarly, the final knowledge concept representation e_k is obtained by fusing the knowledge concept representations under different meta-paths through the above steps, visualized as shown in Fig. 4, where $|K|$ is the number of knowledge concepts.

IV. KNOWLEDGE CONCEPT RECOMMENDATION MODEL WITH SPATIO-TEMPORAL INFORMATION

The STKCR model architecture is shown in Fig. 5, which is divided into three main parts: a module for learning student and knowledge concept representations based on TSA-GCN; a module for predicting students' learning interests based on Attention RNN; and a module for recommending Top-N knowledge concepts based on Extended Matrix Decomposition.

Based on the MOOC platform, the representation learning module firstly constructs a HIN containing five kinds of objects: students, teachers, courses, videos, and knowledge concepts and the corresponding relationships among them; secondly, it selects the student and knowledge concept meta-paths to generate the corresponding trainable self-weighted adjacency matrices, and then learns the representation of students and knowledge concepts under the meta-paths through the TSA-GCN model; finally, it utilizes the attention mechanism to fuse the multi-meta-path information; learning interest prediction The module captures the dependencies in the students' time-series behavioral data through RNN, and inputs them into the attention network after adding the positional encoding to predict the students' current learning interests; the above information is combined to generate the preference matrix using the extended matrix factorization, and a personalized knowledge concept recommendation list is generated for each student.

A. TSA-GCN BASED LEARNING OF STUDENT, KNOWLEDGE CONCEPT REPRESENTATION

As shown in the first part of Figure 5, representation learning of students and knowledge concepts can be done synchronously. The steps are as follows:

First, in the constructed heterogeneous information network G, the corresponding adjacency matrices under the meta-paths can be obtained according to the selected set of meta-paths $[MP]$, and the set of trainable self-weighted adjacency matrices $\left[\tilde{P}\right]$ is calculated according to (1), the upper two in the figure are knowledge concept adjacency matrices $\tilde{P}^K \in R^{|K|\times|K|}$, and the lower four are the students' adjacency matrices $\tilde{P}^U \in R^{|S|\times|S|}$. Second, initialize the features of students and knowledge concepts h^0 and the adjacency matrices $\tilde{P}^{MP}$ obtained in different meta-paths MP are fed into a single layer TSA-GCN, and learn the representations e^{MP} of students and knowledge concepts under MP according to (2). Finally, save the learned knowledge concept representations $\{e_K^{KSK}, e_K^{KVCVK}\}$ and student representations $\{e_S^{SCS}, e_S^{SVS}, e_S^{SKS}, e_S^{SCTCS}\}$ under different meta-paths in the form of sequences, and learn the weights of different meta-paths by using the attention mechanism and weight the fusion to get the final knowledge concept representations e_K and student representations e_S.

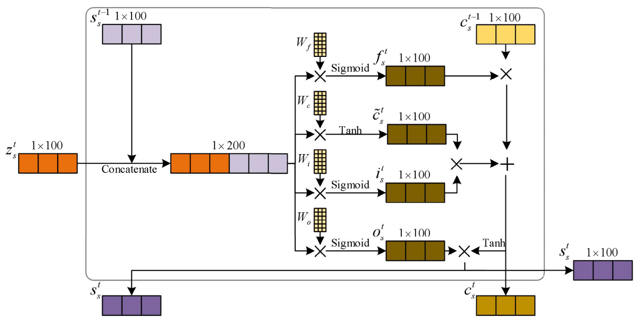


FIGURE 6. LSTM structure.

B. ATTENTION RNN-BASED PREDICTION OF STUDENT INTEREST FEATURES

Recurrent Neural Network (RNN) is mainly used to model sequence data, which can effectively mine the temporal and semantic information in the data [27]. LSTM is a variant of RNN, which can better capture long temporal sequence dependencies. In this paper, LSTM will be used to capture the long dependencies in the student's first T temporal sequential behaviors $\{k_s^1, k_s^2, \cdots, k_s^t, \cdots, k_s^T\}$, where k_s^t denotes the knowledge concept k that the student s learns at the moment t; and then adaptively assign weights to the sequence elements through the attention mechanism. The structure of LSTM is shown in Figure 6 below.

The feature matrix of the knowledge concept is $z_K \in R^{|K|\times d}$, where $d = 100$ is the feature dimension, and the student's temporal behavioral features $\{z_s^1, z_s^2, \cdots, z_s^t, \cdots, z_s^T\}$ are represented by the features of the current moment of learning the knowledge concept k_s^t. The LSTM realizes the functions of selectively forgetting the information of the previous moment, selectively remembering the information of the current moment, and selecting the information as the output of the current moment, respectively, through the three gating units of the forgetting gate f_s^t, the input gate i_s^t, and the output gate o_s^t. The formula is as follows:

$$\begin{aligned} i_s^t &= sigmoid\left(\mathrm{W}_i \cdot \left[z_s^t || s_s^{t-1}\right] + b_i\right), \\ f_s^t &= sigmoid\left(\mathrm{W}_f \cdot \left[z_s^t || s_s^{t-1}\right] + b_f\right), \\ o_s^t &= sigmoid\left(\mathrm{W}_o \cdot \left[z_s^t || s_s^{t-1}\right] + b_o\right), \end{aligned} \tag{5}$$

where W_i, W_f, W_o and b_i, b_f, b_o are trainable parameters and $||$ denotes the serial operation. After passing through three gates, the memory cell vector c^t and the state vector s^t are computed as shown in (6):

$$\begin{aligned} \tilde{c}_s^t &= \tanh\left(\mathrm{W}_c \cdot \left[z_s^t || s_s^{t-1}\right] + b_c\right), \\ c_s^t &= f_s^t \cdot c_s^{t-1} + i_s^t \cdot \tilde{c}_s^t, \\ s_s^t &= o_s^t \cdot \tanh\left(c_s^t\right) \end{aligned} \tag{6}$$

The state sequence $\{s_s^1, s_s^2, \cdots, s_s^t, \cdots, s_s^T\}$ obtained from the students' temporal behavioral features after LSTM is entered into the attention network, which is different from

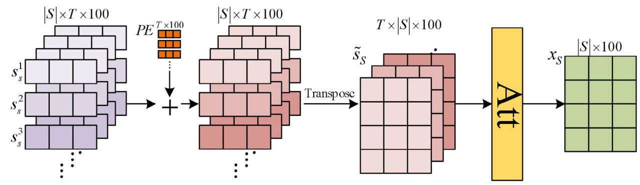


FIGURE 7. Positional encoding and prediction of student interest.

the node representation sequences under different meta-paths in that the state sequences are temporal in nature. Therefore, before entering the attention network, adding the position encoding PE to the state sequence, whose dimension d is consistent with the dimension of the knowledge concept feature, which is calculated using (7).

$$\begin{aligned} PE^{(t,2i)} &= \sin(t/10000^{2i/d}) \\ PE^{(t,2i+1)} &= \cos(t/10000^{2i/d}) \end{aligned} \tag{7}$$

where $2i$ denotes the even dimension and $2i + 1$ denotes the odd dimension. The new state vector $\tilde{s}_s^t$ is obtained by adding the positional encoding PE and the state vector s_s^t of the corresponding position.

$$s_u^t + PE^t = \tilde{s}_u^t \tag{8}$$

The new state sequence is fed into the attention network to predict the interest features x_s of student s. The positional encoding and student interest prediction process is shown in Fig. 7.

C. EXTENDED MATRIX FACTORIZATION BASED ON FUSION OF SPATIAL AND TEMPORAL INFORMATION

Matrix factorization [1], [28], is widely used in recommender systems to effectively integrate information learned from network representations for the final joint optimization prediction task. The basic idea is to map the user-item matrix into the product of two low-dimensional latent factor matrices $P \in R^{|S|\times d'}$ and $Q \in R^{d'\times|K|}$, where d' denotes the dimensionality of the latent space, and use the similarity between the user and item latent factors to simulate the user's preference for items. This method can effectively alleviate the data sparsity problem and explore the hidden features of users and items in the same low-rank space. In this paper, we adopt the extended matrix factorization method to integrate the student representation e_s, the knowledge concept representation e_k, the knowledge concept features z_k, and the student interest features x_s. The formulas are as follows:

$$y_{sk} = p_s^T q_k + \beta_1 \cdot e_s^T \mathrm{M}_1 e_k + \beta_2 \cdot x_s^T \mathrm{M}_2 z_k + b_k \tag{9}$$

where y_{sk} reflects the final preference of student s for knowledge concept k, β_1, β_2 is a trainable parameter that weighs spatial and temporal features, M_1, M_2 is a trainable matrix allowing e_s, e_k and x_s, y_k to be in the same space, and b_k is a bias term making the prediction more accurate.

The visualization of the extended matrix factorization is shown in Figure 8.

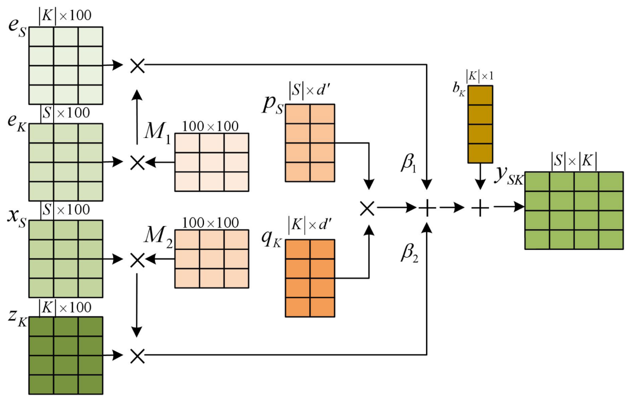


FIGURE 8. Visualization of the extended matrix factorization.

In this paper, the loss function is constructed based on Bayesian personalized ranking [29], and the basic idea is that students' preference ratings for learned knowledge concepts are higher than preference ratings for unlearned knowledge concepts, and the specific loss function is shown in (10):

$$L = \sum_{s \in S, (i,j) \in K} - \ln\left(sigmoid\left(y_{si} - y_{sj}\right)\right) + \lambda \left\|\Theta\right\|^2 \quad (10)$$

where i, j is the interacted and un-interacted knowledge concepts of student s, respectively, and $y_{si} - y_{sj}$ is used to compute the preference difference between the two knowledge concepts, which is increased by the training loss function. In addition, the L2 regularization term is added, λ is the regularization parameter, and Θ denotes all trainable parameters.

Algorithm 1 illustrates the basic steps of STKCR.

Algorithm 1 STKCR Algorithm

Input: G = (V, E): network schema of heterogeneous information networks in MOOC Platforms
S : set of students
K : set of knowledge concepts
$\left\{k_S^1, k_S^2, \cdots, k_S^t, \cdots, k_S^T\right\}$: the first T sequential actions of the students
z_K : characteristics of knowledge concepts
Output: y_{SK}: students-knowledge concepts preference matrix
1: According to G = (V, E), select the appropriate meta-path set $[MP]$ of the students and knowledge concepts
2: **for** each $MP \in [MP]$ **do**
3: Calculate $\tilde{P}$ using formula (1)
4: Initialize the initial features of the students or knowledge concept h^0
5: Calculate e^{MP} according to the definition 3.2.2
6: **end for**
7: According to $e^{[MP]}$, combined with the definition 3.3, the node representation of students e_S and knowledge concepts e_K is generated.
8: **for** each $s \in S$ **do**
9: According to $\left\{k_s^1, k_s^2, \cdots, k_s^t, \cdots, k_s^T\right\}$ and z_K, the characteristics of student temporal behavior $\left\{z_s^1, z_s^2, \cdots, z_s^t, \cdots, z_s^T\right\}$ are obtained
10: The state sequence $\left\{s_s^1, s_s^2, \cdots, s_s^t, \cdots, s_s^T\right\}$ is obtained by LSTM
11: The interest features of student x_s are predicted after adding the positional encoding PE
12: **end for**
13: The interest features of all students x_S are obtained
14: According to e_S, e_K, x_S, z_K, the final students-knowledge concepts preference matrix y_{SK} is calculated using the extended matrix factorization of definition 4.3

V. EXPERIMENTAL

A. DATASETS

The STKCR algorithm proposed in this paper is validated on the MOOCCube dataset (available online at http://moocdata.cn/data/MOOCCube). MOOCCube is an open data warehouse for large-scale online education, which mainly collects three dimensions around courses, knowledge concepts, and student behaviors in the XuetangX platform data, containing massive information such as 706 MOOC courses, 38,181 videos, 114,563 concepts, and 199199 real MOOC students [30]. In this paper, we use the data from 2017 to 2019 to start the study, in which the student behaviors between 2017-01 and 2019-10 are used as the training set, and the student behaviors from 2019-11 to 2019-12 are used as the validation set. In order to verify the effectiveness of knowledge concept recommendation, students who did not learn new knowledge concepts (knowledge concepts not learned in the training set) during the validation set were removed. After screening, 2005 students and their associated learning behavior data were selected.

B. BASELINE MODEL AND ASSESSMENT INDICATORS

To evaluate the model performance, in this paper, each student-interacted concept in the validation set is combined with 99 non-interacted concepts as a group, and HR@K, NDCG@K, MRR, and AUC metrics are computed based on the student's preference y_{sk}. Where K is set to 5 and 10.

In order to validate the model effect, STKCR is compared with five baseline models in the MOOCCube dataset, and the baseline models are introduced as shown in Table 3.

C. EXPERIMENTAL ENVIRONMENT AND EXPERIMENTAL HYPERPARAMETER SETTINGS

The experimental environment of this paper is shown in Table 4.

In this paper, the learning rate of the proposed STKCR model is set to 0.01, the regularization parameter $\lambda = 1e-8$, the embedding dimensions of students and knowledge concepts in both the representation learning model and the interest feature prediction model are 100, the dimension of the hidden layer of the attention mechanism is set to 32, the length of the sequence of student behaviors input to the LSTM is 20, and the potential embedding dimensions of the students' matrix P and the knowledge concepts' matrix Q in the extended matrix factorization module are is set to 32, the batch size for training the model is 512, and gradient descent optimization is performed using the Adam optimizer.

TABLE 3. Baseline model.

Baseline	Introduce
MFBPR [29]	A Top-N sorting algorithm that can be applied to various existing recommendation algorithm frameworks. Utilizing implicit feedback data to solve the problem of ranking among recommended items.
metapath2vec [31]	A representation learning model in a heterogeneous network that uses random wandering to construct node neighborhoods and then learns node representations using a heterogeneous skip-gram.
ACKRec[9]	A graph convolutional neural network model with an attention mechanism that learns entity embeddings under different meta-paths through the adjacency matrix and uses the attention mechanism to learn the final embedding.
MOOCIR[32]	A knowledge concept recommendation method based on heterogeneous information networks and graph convolution, by selecting suitable meta-paths to construct neighbourhood matrices and adaptively learning the network representation of entities.
TTRKRec[18]	A knowledge concept recommendation method based on tensor decomposition and transform reordering, generating a list of knowledge concept recommendations in a high-dimensional space and reordering them using transforms.

TABLE 4. Experimental environment.

Experimental environment	Environment configuration
Operating systems	Linux
CPU	Intel(R) Xeon(R) Gold 6330H
Video Cards	GeForce RTX 3090
RAM	32GB
ROM	1T SSD
Programming Languages	Python 3.6
Framework	Tensorflow

D. EXPERIMENTS AND ANALYSIS OF RESULTS

The ROC curves for the STKCR model and the other baseline models in the MOOCCube dataset are shown in Figure 9. Comparison of the area under the ROC curve shows that the STKCR model has the optimal AUC value.

Table 5 shows the performance metrics of the STKCR model on the MOOCCube dataset versus other baseline models, as analyzed below:

(1)The method proposed in this paper outperforms other models in all evaluation metrics, which demonstrates the importance of student modeling for knowledge concept recommendation and the effectiveness of the TSA-GCN and Attention RNN-based model proposed in this paper.

(2)STKCR shows a significant improvement compared to the matrix factorization-based MFBPR. This demonstrates the importance of mining potential associations between nodes and learning entity representations based on heterogeneous information networks.

(3) The metapath2vec model mines potential semantics among nodes under meta-paths in heterogeneous networks

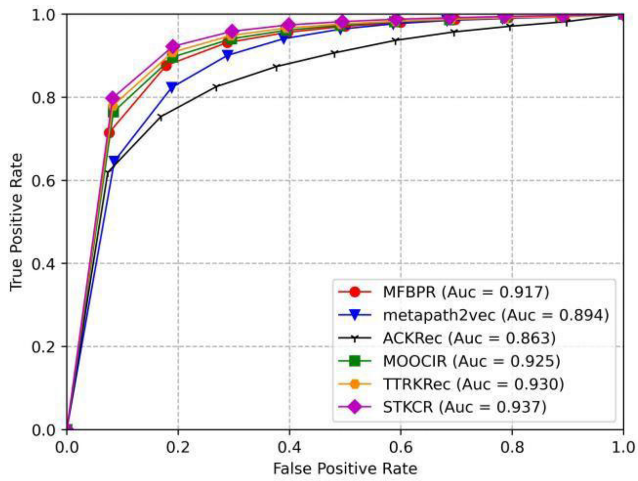


FIGURE 9. Comparison of the ROC curves of all models on MOOCCube datasets.

based on random wandering and skip-gram models, and the STKCR model learns node representations under multiple meta paths through graph convolution models and attention mechanisms represents better than metapath2vec, demonstrating that attentional graph convolution better balances the effects of different meta-paths on nodes.

(4) Both STKCR and ACKRec models use graph convolution and attention mechanisms to learn node representations, and STKCR works better because the differences in nodes' ability to perceive the outside world are taken into account in the process of aggregating information about neighboring nodes using TSA-GCN, which proves the effect of a trainable self-weighted adjacent matrix.

(5) STKCR has better performance than the MOOCIR model that does not consider students' sequential learning behaviors, and from the perspective of student modeling, the combined consideration of potential relationships between nodes and student interest offset features can help the model learn a more accurate representation of students.

(6) Comparison in the experimental process found that the TTRKRec method based on tensor decomposition has a high complexity in the tensor construction process, and will occupy a large amount of memory in the pre-training period. the STKCR model can ensure the integrity of information in a smaller memory environment with the help of a heterogeneous information network.

E. ABLATION EXPERIMENT

1) EFFECTIVENESS OF TSA-GCN AND ATTENTION RNN

The model proposed in this paper is based on TSA-GCN and Attention RNN to realize the extraction of spatial and temporal information from the data, and in order to verify the effect of the two modules on the STKCR, the ablation study is carried out on the MOOCCube dataset. The experimental results are shown in Table 6, where STKCR-og and STKCR-or are the methods to disable the TSA-GCN and Attention RNN modules, respectively, which is to evaluate the effect

TABLE 5. Comparison results of the stkcr model with the baseline model on the mooccube dataset.

Model	HR@5	HR@10	NDCG@5	NDCG@10	MRR	AUC
MFBPR	0.6357	0.7856	0.4493	0.4981	0.4185	0.9166
METAPATH2VEC	0.6543	0.7995	0.4697	0.5170	0.4384	0.8943
ACKRec	0.5936	0.7227	0.4404	0.4824	0.4187	0.8633
MOOCIR	0.7096	0.8415	0.5220	0.5650	0.4861	0.9246
TTRKRec	0.7044	0.8374	0.5188	0.5621	0.4837	0.9297
STKCR	**0.7150**	**0.8494**	**0.5268**	**0.5706**	**0.4906**	**0.9369**

TABLE 6. The results of stkcr without and with TSA-GCN and attention RNN on the mooccube dataset.

Model	HR@5	HR@10	NDCG@5	NDCG@10	MRR	AUC
MFBPR	0.6357	0.7856	0.4493	0.4981	0.4185	0.9166
STKCR-og	0.6423	0.7959	0.4533	0.5034	0.4217	0.9235
STKCR-or	0.7109	0.8468	0.5235	0.5677	0.4877	0.9205
STKCR	**0.7150**	**0.8494**	**0.5268**	**0.5706**	**0.4906**	**0.9369**

TABLE 7. Experimental results in order to verify that the trainable adaptive adjacency matrix and LSTM.

Model	HR@5	HR@10	NDCG@5	NDCG@10	MRR	AUC
STKCR-gcn	0.7103	0.8427	0.5210	0.5641	0.4841	0.9338
STKCR-gru	0.7074	0.8422	0.5210	0.5650	0.4858	**0.9370**
STKCR	**0.7150**	**0.8494**	**0.5268**	**0.5706**	**0.4906**	0.9369

TABLE 8. The results of STKCR without and with positional encoding on the mooccube dataset.

Model	HR@5	HR@10	NDCG@5	NDCG@10	MRR	AUC
STKCR-op	0.7114	0.8458	0.5227	0.5665	0.4864	0.9340
STKCR	**0.7150**	**0.8494**	**0.5268**	**0.5706**	**0.4906**	**0.9369**

of spatial and temporal information on the final recommendation effect. MFBPR is a recommendation model in which STKCR eliminates the modules TSA-GCN and Attention RNN to retain only the matrix decomposition, which is to demonstrate the validity of spatial and temporal information.

From Table 6, it can be found that STKCR-og and STKCR-or outperform MFBPR on the MOOCCube dataset, which proves the effectiveness of the TSA-GCN module and the Attention RNN module, and the adequate extraction of both the temporal and spatial information can improve the recommendation performance of the model. In addition, STKCR outperforms STKCR-og and STKCR-or in all metrics, which fully demonstrates the influence of spatio-temporal information on the final recommendation results, and modeling students from both spatial and temporal dimensions can effectively ameliorate the limitations of single-dimension modeling.

2) EFFECTIVENESS OF TRAINABLE SELF-WEIGHTED ADJACENCY MATRIX AND LSTM

Trainable self-weighted adjacency matrix is the improvement method for traditional GCN in this paper, in order to prove the effectiveness of trainable self-weighted adjacency matrix, the recommended performance of STKCR is compared in the dataset with the model variant STKCR-gcn (which replaces the trainable self-weighted adjacency matrixwith a conventional adjacency matrix). The experimental results are shown in Table 7. Meanwhile, in order to verify the effect of LSTM, the model variant STKCR-gru generated by replacing LSTM with GRU is also added together in Table 7 for demonstration.

From Table 7, it can be found that all the indicators of STKCR are better than the STKCR-gcn variant, which proves that the trainable self-weighted adjacency matrix proposed in this paper can fully take into account the differences in the user's perceptual ability to achieve accurate modeling of

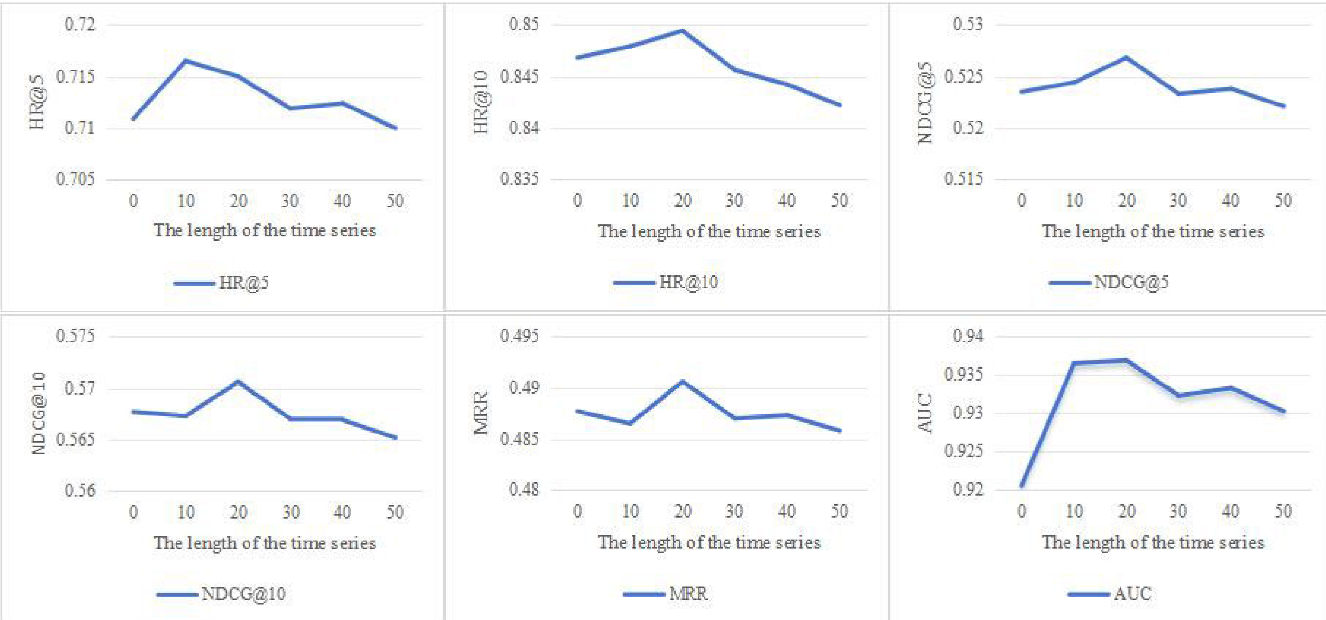


FIGURE 10. Effect of different sequence lengths on recommended performance.

TABLE 9. Recommendation lists and actual learning records of student U_270210 under different recommendation models.

Real Learning Records	Recommended sequence		
	STKCR	TTRKRec	MOOCIR
K_场地分析_建筑学(K_Site Analysis_Architecture)	K_多发伤_医学(K_Multiple Injuries_Medicine)	K_资产支持证券_管理科学技术(K_Asset-Backed Securities_Management Science and Technology)	K_多发伤_医学(K_Multiple Injuries_Medicine)
K_儿童游戏场_建筑学(K_Children's Playground_Architecture)	**K_大气候_地理学(K_Great Climate_Geography)**	K_静电位_电气工程(K_Static potential_Electrical engineering)	K_领导决策_管理科学技术(K_Leadership Decision-Making_Management Science and Technology)
K_土地用途_建筑学(K_Land Use_Architecture)	**K_场地分析_建筑学(K_Site Analysis_Architecture)**	K_多发伤_医学(K_Multiple Injuries_Medicine)	**K_大气候_地理学(K_Great Climate_Geography)**
K_地形设计_建筑学(K_Topography_Architecture)	**K_儿童游戏场_建筑学(K_Children's Playground_Architecture)**	**K_大气候_地理学(K_Great Climate_Geography)**	**K_儿童游戏场_建筑学(K_Children's Playground_Architecture)**
K_大气候_地理学(K_Great Climate_Geography)	**K_土地用途_建筑学(K_Land Use_Architecture)**	**K_场地分析_建筑学(K_Site Analysis_Architecture)**	K_项目章程_管理科学技术(K_Project Charter_Managing Science and Technology)
K_土壤调查_地理学(K_Soil Survey_Geography)	**K_地形设计_建筑学(K_Topography_Architecture)**	**K_儿童游戏场_建筑学(K_Children's Playground_Architecture)**	**K_场地分析_建筑学(K_Site Analysis_Architecture)**
K_地理纬度_地理学(K_Geographic Latitude_Geography)	K_噻唑_化学(K_Thiazole_Chemistry)	K_韭葱_农学(K_Leek_Agronomy)	**K_土地用途_建筑学(K_Land Use_Architecture)**
K_冬季风_地理学(K_Winter Winds_Geography)	K_寺庙园林_建筑学(K_Temple Garden_Architecture)	**K_地形设计_建筑学(K_Topography_Architecture)**	K_互调_电子学(K_Intermodulation_Electronics)
K_土壤资源_地理学(K_Soil Resources_Geography)	K_网络操作系统_计算机科学技术(K_Network Operating System_Computer Science and Technology)	**K_土地用途_建筑学(K_Land Use_Architecture)**	K_IgA 肾病_医学(K_IgA Nephropathy_Medical)
K_太阳高度角_建筑学(K_Solar Altitude Angle_Architecture)	K_静电位_电气工程(K_Static potential_Electrical engineering)	K_噻唑_化学(K_Thiazole_Chemistry)	K_红树林沼泽_地理学(K_Mangrove Swamp_Geography)

the user and improve the accuracy of the recommendation. Furthermore, STKCR outperformed the STKCR-gcn variant overall, which demonstrates that LSTM is better able to capture long-term dependencies in students' temporal behaviors and predict students' interest profiles from temporal information compared to GRU.

3) EFFECTIVENESS OF POSITIONAL ENCODING

Considering that positional encoding is not necessary for the attention mechanism, experimental validation is performed in this section in order to verify the impact of positional encoding in the Attention RNN module. Table 8 shows the experimental results of the model variant STKCR-op without positional encoding versus STKCR. The results show that positional encoding enhances the temporal information in the student behavior data and helps the model extract the temporal information more adequately, thus improving the model's recommendation.

4) INFLUENCE OF THE LENGTH OF STUDENTS' TIME-SERIES BEHAVIORAL SEQUENCES

To verify the effect of the length of students' temporal behavioral sequences on students' interest profiles, ablation experiments were conducted on the MOOCCube dataset. The experimental results are shown in Fig. 10, which shows the effect of using different lengths of sequences on the recommendation performance in the student interest feature prediction module based on Attention RNN, where the sequence length grows from 0 to 50 in steps of 10.

Figure 10 illustrates the variation of the recommendation metrics for the STKCR model in the MOOCCube dataset with different sequence lengths.

From Fig. 10, it can be found that the model performance tends to increase with the increment of sequence length, which fully proves the importance of student modeling and the effectiveness of using attention RNN to predict students' learning interests.

The model reaches the best effect when the sequence length increases to 20, and the model performance starts to decline when the sequence length increases again, which proves that the sequence length is not as long as better. Students learning interests are highly time-sensitive, and too long sequences will cause information interference and increase the burden of model training, leading to the disappearance of the gradient affecting the recommendation performance.

F. COMPARATIVE ANALYSIS OF RECOMMENDED SEQUENCES OF KNOWLEDGE CONCEPTS AND ACTUAL LEARNING RECORDS

In order to demonstrate the effectiveness of this paper's algorithm STKCR, this section randomly selects a student (id: U_270210) from the MOOCCube dataset and demonstrates the student's actual learning records as well as the Top-10 knowledge concepts recommendation lists generated under different recommendation models, as shown in Table 9.

The concepts of recommending correct knowledge are bolded in the table, and as can be seen in Table 9, the STKCR model can achieve better accuracy in the pre-recommendation period compared to the other two models because it takes into account the student's interest bias characteristics during the learning process. From the comprehensive Top-10 recommendation list, the recommendation list generated by the model in this paper is more in line with the actual learning situation of students in terms of recommendation order and accuracy, which verifies the validity of the STKCR algorithm proposed in this paper, and also proves that modeling students from both spatial and temporal dimensions can improve the accuracy of the model's recommendation.

VI. CONCLUSION

In this paper, the proposed STKCR model that integrates spatial and temporal information is used for knowledge concept recommendation on the MOOC platform. The model comprehensively considers the potential correlation information between nodes in the heterogeneous information network and the student's perceptual ability, interest bias, and other features in the learning process to model students, which not only achieves good recommendation performance in the MOOCCube dataset but also verifies the effectiveness of the algorithm in this paper by comparing it with the baseline model. In addition, this paper conducts a large number of ablation experiments to verify the importance of spatial and temporal information as well as the impact of different modules on the recommendation performance and tests the recommendation performance of the model under different lengths of student behavior sequences.

This study focuses on the problem of personalized learning resource recommendation for MOOC platforms, using heterogeneous information from online platforms and students' time-series learning behavior data to construct a knowledge concept recommendation model with student modeling as the core. As an auxiliary teaching module for online learning, this model can stimulate students' interest in online learning and overcome problems such as knowledge disorientation in the learning process; at the same time, effective modeling of students is also conducive to teachers' understanding of the student's learning status and improving the design of the teaching process. The idea of improving recommendation performance in this paper is also applicable to other recommendation domains: focusing on the temporal sequence of user behavior and fully considering the user's perceptual ability, interest bias, and other characteristics can more accurately portray the user's image and provide more personalized recommendation services. In future research, the implementation of joint recommendations of different learning resources is considered to correlate and analyze the impact between different learning resources based on sufficiently accurate user profiles, giving students richer and more effective personalized choices while maintaining the recommendation cost.

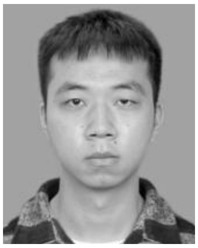
REFERENCES

[1] I. Uddin, A. S. Imran, K. Muhammad, N. Fayyaz, and M. Sajjad, "A systematic mapping review on MOOC recommender systems," *IEEE Access*, vol. 9, pp. 118379–118405, 2021, doi: 10.1109/ACCESS.2021.3101039.

[2] A. Faroughi and P. Moradi, "MOOCs recommender system with Siamese neural network," in *Proc. 9th Int. 15th Nat. Conf. E-Learn. E-Teach. (ICeLeT)*, Sanandaj, Iran, Mar. 2022, pp. 1–6, doi: 10.1109/ICeLeT55619.2022.9765439.

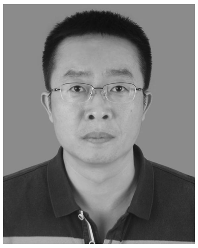
[3] A. Picado, C. Finamore, A. M. Santos, and C. Antunes, ''Students temporal profiling and e-Learning resources recommendation based on NLP's terms extraction,'' in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Orlando, FL, USA, Nov. 2022, pp. 264–273, doi: 10.1109/ICDMW58026.2022.00044.
[4] X. Diao, Q. Zeng, L. Li, H. Duan, H. Zhao, and Z. Song, ''Personalized learning path recommendation based on weak concept mining,'' *Mobile Inf. Syst.*, vol. 2022, pp. 1–17, May 2022, doi: 10.1155/2022/2944268.
[5] X. Tian and F. Liu, ''Capacity tracing-enhanced course recommendation in MOOCs,'' *IEEE Trans. Learn. Technol.*, vol. 14, no. 3, pp. 313–321, Jun. 2021, doi: 10.1109/TLT.2021.3083180.
[6] A.-C. Chuang, N. -F. Huang, J.-W. Tzeng, C.-A. Lee, Y.-X. Huang, and H.-H. Huang, ''MOOCERS: Exercise recommender system in MOOCs based on reinforcement learning algorithm,'' in *Proc. 8th Int. Conf. Soft Comput. Mach. Intell. (ISCMI)*, Cario, Egypt, 2021, pp. 186–190, doi: 10.1109/ISCMI53840.2021.9654805.
[7] H. Shrimali, R. Saxena, and Kavita, ''Content based video recommendation system,'' in *Proc. 3rd Int. Conf. Intell. Commun. Comput. Techn. (ICCT)*, Jaipur, India, Jan. 2023, pp. 1–3. [Online]. Available: https://ieeexplore.ieee.org/document/10075906, doi: 10.1109/ICCT56969.2023.10075906.
[8] C. Ju, Y. Zhu, and C. Wang, ''Knowledge concept recommendation model for MOOCs with local sub-graph embedding,'' in *Proc. Int. Conf. Autom., Robot. Comput. Eng. (ICARCE)*, Wuhan, China, Dec. 2022, pp. 1–8, doi: 10.1109/ICARCE55724.2022.10046444.
[9] S. Wang, J. Gong, J. Wang, W. Feng, H. Peng, J. Tang, and P. S. Yu, ''Attentional graph convolutional networks for knowledge concept recommendation in MOOCs in a heterogeneous view,'' in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.* New York, NY, USA: Association for Computing Machinery, 2020, pp. 79–88, doi: 10.1145/3397271.3401057.
[10] M. Yu, Z. Ding, J. Yu, W. Zhang, M. Yang, and M. Zhao, ''Graph contrastive learning with adaptive augmentation for knowledge concept recommendation,'' in *Proc. 26th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, Rio de Janeiro, Brazil, 2023, pp. 1281–1286, doi: 10.1109/CSCWD57460.2023.10152806.
[11] Z. Shou, Z. Shi, H. Wen, J. Liu, and H. Zhang, ''Learning peer recommendation based on weighted heterogeneous information networks on online learning platforms,'' *Electronics*, vol. 12, no. 9, p. 2051, Apr. 2023, doi: 10.3390/electronics12092051.
[12] X. Wang, L. Jia, L. Guo, and F. Liu, ''Multi-aspect heterogeneous information network for MOOC knowledge concept recommendation,'' *Appl. Intell.*, vol. 53, no. 10, pp. 11951–11965, May 2023, doi: 10.1007/s10489-022-04025-x.
[13] W. Wang, L. Wei, Y. Li, Y. Zhu, B. Li, and L. Zhang, ''Attentional meta-path contrastive graph convolutional networks for knowledge concept recommendation,'' in *Proc. 10th Int. Conf. Adv. Cloud Big Data (CBD)*, Guilin, China, Nov. 2022, pp. 200–205, doi: 10.1109/CBD58033.2022.00043.
[14] R. Alatrash, M. A. Chatti, Q. U. Ain, Y. Fang, S. Joarder, and C. Siepmann, ''ConceptGCN: Knowledge concept recommendation in MOOCs based on knowledge graph convolutional networks and SBERT,'' *Comput. Educ., Artif. Intell.*, vol. 6, Jun. 2024, Art. no. 100193, doi: 10.1016/j.caeai.2023.100193.
[15] Y. Ling and Z. Shan, ''Knowledge concept recommender based on structure enhanced interaction graph neural network,'' in *Knowledge Science, Engineering and Management*. Cham, Switzerland: Springer, 2022, pp. 173–186, doi: 10.1007/978-3-031-10983-6_14.
[16] A. Klasnja-Milicevic and D. Milicevic, ''Top-N knowledge concept recommendations in MOOCs using a neural co-attention model,'' *IEEE Access*, vol. 11, pp. 51214–51228, 2023, doi: 10.1109/ACCESS.2023.3278609.
[17] H. Liu, J. Ding, L. T. Yang, Y. Guo, X. Wang, and A. Deng, ''Multi-dimensional correlative recommendation and adaptive clustering via incremental tensor decomposition for sustainable smart education,'' *IEEE Trans. Sustain. Comput.*, vol. 5, no. 3, pp. 389–402, Jul. 2020, doi: 10.1109/TSUSC.2019.2954456.
[18] Z. Shou, Y. Chen, H. Wen, J. Liu, J. Mo, and H. Zhang, ''A knowledge concept recommendation model based on tensor decomposition and transformer reordering,'' *Electronics*, vol. 12, no. 7, p. 1593, Mar. 2023, doi: 10.3390/electronics12071593.
[19] J. Gong, C. Wang, Z. Zhao, and X. Zhang, ''Automatic generation of meta-path graph for concept recommendation in MOOCs,'' *Electronics*, vol. 10, no. 14, p. 1671, Jul. 2021, doi: 10.3390/electronics10141671.
[20] D. Wu, M. Tang, S. Zhang, A. You, and W. Gao, ''KPRLN: Deep knowledge preference-aware reinforcement learning network for recommendation,'' *Complex Intell. Syst.*, vol. 9, no. 6, pp. 6645–6659, Dec. 2023, doi: 10.1007/s40747-023-01083-7.
[21] Z. Liang, L. Mu, J. Chen, and Q. Xie, ''Graph path fusion and reinforcement reasoning for recommendation in MOOCs,'' *Educ. Inf. Technol.*, vol. 28, no. 1, pp. 525–545, Jan. 2023, doi: 10.1007/s10639-022-11178-2.
[22] J. Gong, Y. Wan, Y. Liu, X. Li, Y. Zhao, C. Wang, Y. Lin, X. Fang, W. Feng, J. Zhang, and J. Tang, ''Reinforced MOOCs concept recommendation in heterogeneous information networks,'' *ACM Trans. Web*, vol. 17, no. 3, pp. 1–27, Aug. 2023, doi: 10.1145/3580510.
[23] Y. Yin and W. Zheng, ''An efficient recommendation algorithm based on heterogeneous information network,'' *Complexity*, vol. 2021, pp. 6689323:1–6689323:18, Mar. 2021, doi: 10.1155/2021/6689323.
[24] T. Li, X. Su, W. Liu, W. Liang, M.-Y. Hsieh, Z. Chen, X. Liu, and H. Zhang, ''Memory-augmented meta-learning on meta-path for fast adaptation cold-start recommendation,'' *Connection Sci.*, vol. 34, no. 1, pp. 301–318, Dec. 2022, doi: 10.1080/09540091.2021.1996537.
[25] P. G. Duran, A. Karatzoglou, J. Vitrià, X. Xin, and I. Arapakis, ''Graph convolutional embeddings for recommender systems,'' *IEEE Access*, vol. 9, pp. 100173–100184, 2021, doi: 10.1109/ACCESS.2021.3096609.
[26] S. Sun, Y. Tang, Z. Dai, and F. Zhou, ''Self-attention network for session-based recommendation with streaming data input,'' *IEEE Access*, vol. 7, pp. 110499–110509, 2019, doi: 10.1109/ACCESS.2019.2931945.
[27] H. Xue, L. Yang, W. Jiang, Y. Wei, Y. Hu, and Y. Lin, ''Modeling dynamic heterogeneous network for link prediction using hierarchical attention with temporal RNN,'' in *Machine Learning and Knowledge Discovery in Databases*. Cham, Switzerland: Springer, 2021, pp. 282–298, doi: 10.1007/978-3-030-67658-2_17.
[28] R. Salakhutdinov and A. Mnih, ''Probabilistic matrix factorization,'' in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 20, Vancouver, BC, Canada, 2007, pp. 1257–1264.
[29] R. Steffen, F. Christoph, G. Zeno, and S. Lars, ''BPR: Bayesian personalized ranking from implicit feedback,'' in *Proc. 25th Conf. Uncertainty Artif. Intell.*, Montreal, QC, Canada, 2009, pp. 452–461, doi: 10.5555/1795114.1795167.
[30] J. Yu, G. Luo, T. Xiao, Q. Zhong, Y. Wang, W. Feng, J. Luo, C. Wang, L. Hou, J. Li, Z. Liu, and J. Tang, ''MOOCCube: A large-scale data repository for NLP applications in MOOCs,'' in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3135–3142, doi: 10.18653/v1/2020.acl-main.285.
[31] Y. Dong, N. V. Chawla, and A. Swami, ''metapath2vec: Scalable representation learning for heterogeneous networks,'' in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Halifax, NS, Canada, Aug. 2017, pp. 135–144, doi: 10.1145/3097983.3098036.
[32] G. Piao, ''Recommending knowledge concepts on MOOC platforms with meta-path-based representation learning,'' in *Proc. 14th Int. Conf. Educ. Data Mining (EDM)*, Jun./Jul. 2021.

ZHAOYU SHOU received the B.S., M.S., and Ph.D. degrees in computer science and technology from Guilin University of Electronic Technology, in 1999, 2004, and 2019, respectively. He presided over and mainly participated in more than 30 projects, including projects of the National Natural Science Fund, the National Development and Reform Commission, Guangxi Natural Science Fund, and Guangxi Scientific Research and Technological Development Project. His research interests include digital image fusion, computer information management, and education big data. He is currently paying more attention to big data in education.



YIXIN CHEN received the B.S. degree from Jiangxi University of Science and Technology, China, in 2021. He is currently pursuing the M.S. degree in electronic and communication engineering with Guilin University of Electronic Technology, China.

His current research interests include data mining and deep learning.



HUIBING ZHANG received the B.S. and M.S. degrees in computer science and technology from Guilin University of Electronic Technology, Guilin, China, in 2000 and 2005, respectively, and the Ph.D. degree in computer science and technology from Beijing University of Technology, Beijing, China, in 2012. He is currently a Professor with Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology. He presided over and participated in a number of NSFC projects. His research interests include educational big data, the Internet of Things, and social computing.

JIANWEN MO received the B.S. degree from the Department of Electronic Engineering, North China Electric Power University, in 1994, the M.S. degree from the School of Mathematics and Computer Science, Guangxi Normal University, in 2002, and the Ph.D. degree from the School of Electronic Engineering, Xidian University, in 2011. He is currently an Assistant Professor with Guangxi Key Laboratory of Cognitive Radio and Information Processing, Ministry of Education, Guilin University of Electronic Technology. He has presided over and mainly participated in numerous projects, including projects for the National Natural Science Fund and Guangxi Natural Science Fund. His research interests include intelligent information processing, computer vision, image processing, machine learning, and artificial intelligence.