

## Article

# Research on Joint Recommendation Algorithm for Knowledge Concepts and Learning Partners Based on Improved Multi-Gate Mixture-of-Experts

Zhaoyu Shou <sup>1,2</sup> , Yixin Chen <sup>1</sup>, Hui Wen <sup>1,\*</sup>, Jinghua Liu <sup>1</sup>, Jianwen Mo <sup>1</sup>  and Huibing Zhang <sup>3</sup>

- <sup>1</sup> School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China; guilinshou@guet.edu.cn (Z.S.); 21022201005@mails.guet.edu.cn (Y.C.); ddy2003@guet.edu.cn (J.L.); jwmo@guet.edu.cn (J.M.)
- <sup>2</sup> Guangxi Wireless Broadband Communication and Signal Processing Key Laboratory, Guilin University of Electronic Technology, Guilin 541004, China
- <sup>3</sup> School of Computer and Information Security, Guilin University of Electronic Technology, Guilin 541004, China; zhanghuibing@guet.edu.cn
- \* Correspondence: huiwen@guet.edu.cn

**Abstract:** The rise of Massive Open Online Courses (MOOCs) has increased the large audience for higher education. Different learners face different learning difficulties in the process of online learning. In order to ensure the quality of teaching, online learning resource recommendation services should be more personalised and have more choices. In this paper, we propose a joint recommendation algorithm for knowledge concepts and learning partners based on improved MMoE (Multi-gate Mixture-of-Experts). Firstly, the heterogeneous information network (HIN) is constructed based on the MOOC platform and appropriate meta-paths are selected in order to extract the human–computer interaction information and student–student interaction information generated during the learners’ online learning processes more completely. Secondly, the temporal behavioural characteristics of students are obtained based on their learning paths as well as their knowledge of conceptual characteristics, and LSTM (Long Short-Term Memory) is used to mine students’ current learning interests. Finally, the gating network in MMoE is changed into an attention mechanism network, and for different tasks, multiple attention mechanism networks are used to fuse the learner’s human–computer interaction information, student–student interaction information, and interest characteristics to generate learner representations that are more in line with the respective task and to complete the tasks of knowledge conception and learning partner recommendation. Experiments on publicly available MOOC datasets show that the method proposed in this paper provides more accurate and varied personalization services to online learners compared to the latest proposed methods.

**Keywords:** multi-objective recommendation; online learning; learner modelling; heterogeneous information networks; LSTM



**Citation:** Shou, Z.; Chen, Y.; Wen, H.; Liu, J.; Mo, J.; Zhang, H. Research on Joint Recommendation Algorithm for Knowledge Concepts and Learning Partners Based on Improved Multi-Gate Mixture-of-Experts. *Electronics* **2024**, *13*, 1272. <https://doi.org/10.3390/electronics13071272>

Academic Editor: Arkaitz Zubiaga

Received: 18 March 2024

Revised: 26 March 2024

Accepted: 27 March 2024

Published: 29 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The spread of higher education has led to a gradual imbalance between the quantity and quality of education, and the rise of online education platforms has increased the large educational audience and brought to the forefront the problem of information overload within them [1]. In order to effectively improve the problems of knowledge disorientation and information overload during users’ online learning, a large number of researchers have focused their attention on the recommendation of online learning resources [2]. Tian et al. [3] integrated the extension of Multi-dimensional Item Response Theory (MIRT) as a competency tracking model into the work of course recommendation for MOOCs to improve the validity and interpretability of MOOCs. Harshal et al. [4] have proposed a video

recommendation model based on natural language processing to recommend videos based on the similarity of video text and query semantics. Chuang et al. [5] developed a reinforcement learning-based exercise recommendation system that recommends personalised exercises suitable for learners concerning their difficulty and knowledge concepts based on the data recorded by the system. In order to capture students' more fine-grained learning states, knowledge concept recommendation has become the mainstream of current online learning resource recommendation. Wang et al. [6] considered the diverse relationships between learners and knowledge concepts to propose a multifaceted heterogeneous information network and used the Gumbel-Softmax method to dynamically assign aspectual contexts to each node to improve the accuracy of knowledge concept recommendations. Gong et al. [7] proposed an end-to-end graph neural network-based knowledge concept recommendation model that uses an attention mechanism to adaptively fuse the representations of entities learnt from graph convolutional networks (GCN) under different meta-paths in order to produce better results for knowledge concept recommendation.

The recommendation of knowledge concepts can effectively alleviate problems such as knowledge disorientation of online learners facing massive learning resources; however, the loneliness and helplessness of users due to the separation of time and space in the process of online learning also need to be emphasized by researchers. Miao et al. [8] showed that online interactions will help to increase users' sense of social presence and further influence learning engagement. Shao et al. [9] proposed a friend recommendation method based on fine-grained interest feature labels, which can leverage the labelling system of a learning community for learning partner recommendation. Hu et al. [10] considered the dynamic interaction between students and learning content to propose a learning partner recommendation framework based on a convolutional neural network and dynamic interaction tripartite graphs.

All of the above studies can effectively improve the learning effect of students in the process of online learning, but they still have some limitations: Firstly, most existing recommendation models only consider recommendations for individual learning resources, such as course recommendations, knowledge concept recommendations, and student partner recommendations, etc. However, learners face different dilemmas in the online learning process. In order to provide more diverse and personalised services to learners while saving computational costs, it is important to study the joint recommendation methods for multiple learning content. Secondly, existing learning resource recommendation models tend to be limited to the nature of the task in terms of learner modelling, e.g., the knowledge concept recommendation model tends to model the learner from the learner's human-computer interaction data, and the learning partner recommendation model tends to look for similar learning partners based on the student's student-student interaction information and some of the interest characteristics. This learner modelling approach cannot realistically portray online learners and is prone to overfitting.

In summary, this paper proposes a joint recommendation algorithm for knowledge concepts and learning partners based on improved MMoE (KLJRec). Firstly, the HIN based on the MOOC platform is constructed and the learner's human-computer interaction information and student-student interaction information are extracted in suitable meta-paths using GCN; secondly, the hidden trend of interest evolution in the learner's temporal behavioural features is mined using LSTM to obtain the current interest features; lastly, in the multi-task scenarios of knowledge concept recommendation and learning partner recommendation, the learner representations that meet the task requirements are generated using multiple attention networks based on the learner's multi-dimensional information features, respectively. The main contributions of this study are as follows:

1. Based on the MMoE framework, this paper effectively integrates the knowledge concept and the learning partner recommendation task. The gating network in the MMoE framework is replaced by an attention network with a stronger ability to capture important information, and by adding a lightweight attention network to generate learner representations that meet the requirements of different tasks, we

- generate a list of knowledge concepts and learning partner recommendations that meet the learner's preferences while saving computing costs.
2. In this paper, the algorithm mixes GCN and LSTM networks in the shared hidden layer MoE at the bottom of the model, which can better extract the human-computer interaction information, student-student interaction information, and learning timing information generated during the user learning process, and portray a more accurate and perfect learner portrait from multiple dimensions, thus reducing the risk of overfitting for the task.
  3. A large number of experiments have been conducted on the MOOCCubeX dataset, and the experimental results demonstrate that KLJRec proposed in this paper is better to consider the coupling between the two recommendation tasks than the state-of-the-art single knowledge concept or learning partner recommendation algorithms, which further improves the recommendation accuracy.

The rest of the paper is organized as follows: Section 2 provides a brief overview of the related work in this paper. Section 3 explains the definitions related to this paper's algorithms. Section 4 gives a detailed description of the proposed method. Section 5 shows the experimental results and analyses the reliability of the algorithm of this paper. Section 6 concludes the work with an outlook.

## 2. Related Work

Existing knowledge concept recommendation models can be mainly classified into three categories: knowledge concept recommendation models based on graph structure, knowledge concept recommendation models based on tensor decomposition, and knowledge concept recommendation models based on reinforcement learning. Ju et al. [11] proposed a model for recommending knowledge concepts based on local subgraph embedding, which uses attention graph convolution to fuse contextual information from different subgraphs to capture complex semantic relationships between entities. In order to increase the interpretability of graph neural networks in the field of knowledge concept recommendation, Alatrash et al. [12] proposed an end-to-end framework combining graph convolutional networks and a pre-trained language model encoder (SBERT) to provide users with personalised lists of recommendations for knowledge concepts with enhanced interpretability. The model based on graph structure can effectively consider the association between heterogeneous information and mine the potential semantic associations from different nodes; however, it is unable to effectively extract the temporal features in the user's behaviour and is difficult to extend in the time dimension. Liu et al. [13] proposed a personalised recommendation algorithm based on an incremental tensor, which performs multi-dimensional correlation analysis of educational data through incremental tensor decomposition to achieve accurate recommendations of learning resources in different environments. The knowledge concept recommendation model based on tensor decomposition can better ensure the integrity of the data and help to discover the hidden structure and value from the massive data, but the tensor construction requires high device memory and computing power. Reinforcement learning-based knowledge concept recommendation models are gradually emerging. Wu et al. [14] designed a reinforcement learning network for knowledge concept recommendation which uses a hierarchical propagation path construction method to help explore further paths and capture students' deep knowledge preferences. Gong et al. [15] formulated knowledge concept recommendation as a reinforcement learning problem to help better model the dynamic interactions between students and knowledge concepts, in addition to introducing a heterogeneous information network between students, courses, videos, and concepts to alleviate the data sparsity problem in the recommendation task. Reinforcement learning-based recommendation methods can improve the modelling of learners from the dynamic interaction between students and knowledge concepts, and can make up for the problem of insufficient extraction of temporal information in graphical structure and tensor decomposition models, but the data

sparsity problem faced by these methods is more serious and the design of their reward mechanism is also a major difficulty.

Learning partner recommendations can effectively enhance student interaction as a way to mitigate the problem of high dropout rates in online courses due to the lack of social connections [16]. Kang et al. [17] proposed an Evaluation Latent Delicacy Allocation (Evaluation-LDA) algorithm to cluster learners with similar learning interests based on constructing learner document datasets, calculating learner similarity, and modelling friend topics as a way to help students in online education recommend suitable learning partners. Shao et al. [18] proposed a learning partner recommendation algorithm that is based on the evolution of learning interests and recommends suitable learning partners for students through interest similarity. The above study calculates the similarity between students through partial interaction information and interest information to recommend appropriate learning partners without considering the integrity of heterogeneous data and ignoring the importance of student–student interaction information. Liu et al. [13] proposed an adaptive clustering and community recommendation algorithm based on incremental tensors, which uses tensor modelling to preserve the integrity of the data, and in this way recommends appropriate learning partners to students in various contexts. In order to alleviate learners' loneliness during online learning, Shou et al. [19] proposed a learning partner recommendation model based on a weighted heterogeneous information network, which extracts more complete interaction information by automatically generating all meaningful meta-paths to reveal students' unique preferences. The above studies have compensated for the lack of information completeness in the learning partner recommendation model to a certain extent; however, the importance of accurately modelling learners from multiple dimensions of human–computer interaction information, student–student interaction information, and students' interest characteristics cannot be ignored. Based on the above studies, the summary of related research models is shown in Table 1.

**Table 1.** Summary of relevant research models.

Model	Paper Numbers	Advantages	Limitations
Recommendation model for knowledge concepts based on graph structures	[6,7,11,12]	Ability to effectively model heterogeneous data and mine potential correlations in heterogeneous data	Difficulty in extracting time series information on user behaviour
Tensor decomposition-based recommendation model for knowledge concepts	[13]	Strong extraction of potential features in high-dimensional data	High memory requirements for the device and difficulty in capturing sequential information
Reinforcement learning-based recommendation model for knowledge concepts	[14,15]	Great for capturing long-term user interest by simulating dynamic user interactions with a project	Reward function design is hard and data sparsity is a serious problem
Learning partner recommendation model based on learners' interest characteristics	[9,17,18]	Suggests similar learners by tagging user interests to make recommendations more interpretable	Vulnerable to loss of information integrity
Learning partner recommendation model based on high-dimensional web space modelling	[13,19]	More complete data information can be stored to reveal user preferences	Insufficient consideration of users' temporal information and difficult to extend in the time dimension

In summary, existing recommendation models only consider recommendations for individual learning content. The knowledge concept recommendation model uses human–computer interaction information and behavioural timing information to model the learner and recommend appropriate knowledge concepts; the learning partner recommendation model relies on partial student behavioural data to compute student similarity

and thus match learning partners with similar interests. None of the above studies provide a complete and accurate characterisation of online learners due to task characteristics, and the risk of overfitting due to feature loss in a single recommendation task cannot be ignored. Therefore, based on the MMoE framework, this paper mixes GCN and LSTM in the shared hidden layer at the bottom of the model to model the learner completely and accurately based on the human–computer interaction information, student–student interaction information, and behavioural time sequence information and uses the attention mechanism for different training tasks to obtain the representation of the learner for the task so as to recommend the appropriate knowledge concepts and learning partners.

### 3. Relevant Definitions

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.

#### 3.1. Heterogeneous Information Network

##### 3.1.1. Building Heterogeneous Information Networks

A heterogeneous information network [20] is defined as a directed graph  $G = (V, E)$  with object type mappings  $\varphi : V \rightarrow A$  and relation type mappings  $\psi : E \rightarrow R$ , where the sum of the total number of object types  $|A|$  and the total number of relationship types  $|R|$  is greater than 2. Figure 1 illustrates a heterogeneous information network constructed for a particular course based on an MOOC platform, consisting of three object types: students ( $S$ ), videos ( $V$ ), and knowledge concepts ( $K$ ), and 10 relationships between them, where  $R_i$  denotes the correspondence between different types of objects ( $R_i^{-1}$  denotes the inverse of  $R_i$ ),  $R_1(R_1^{-1})$ ,  $R_2(R_2^{-1})$ ,  $R_3(R_3^{-1})$ ,  $R_4(R_4^{-1})$ , and  $R_5(R_5^{-1})$  denote learn (learnt by), include (included in), watch (watched by), reply (replied by), and comment (commented by), respectively.

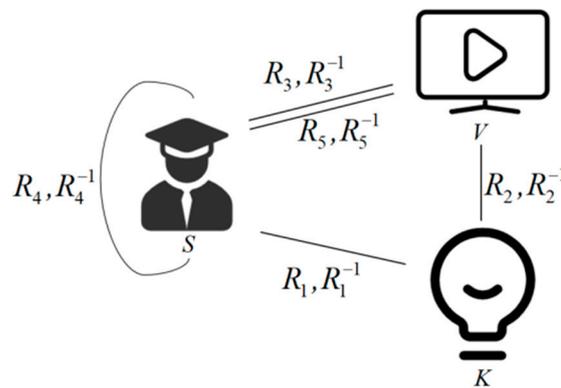


Figure 1. Heterogeneous Information Networks Based on MOOC Courses.

##### 3.1.2. Meta-Path

Meta-paths are defined in heterogeneous information networks where combining the types of relationships in the network through meta-paths leads to richer and more effective semantics [21]. Meta-paths are of the form  $A_1 \xrightarrow{R_1} A_2 \dots \xrightarrow{R_l} A_{l+1}$ . Table 2 demonstrates the six meta-paths [MP] selected for students and knowledge concepts in this study. In the table,  $\{SrS, ScVcS\}$  indicates that the student replied to another student’s statement and that the student commented on the same video, respectively, from which student–student interaction information can be extracted;  $\{SwVwS, SIKIS\}$  represents that the student watched the same video and that the student learnt the same knowledge concept, respectively, as a way of obtaining information about the student’s human–computer interaction. In addition,  $\{KiViK, KISIK\}$  is the knowledge concept meta-path, representing knowledge concepts being included in the same video and both knowledge concepts being learnt by the same user, respectively. It is used to learn the exact knowledge concept representation.

**Table 2.** Meta-paths of Student–Student and Knowledge Concept–Knowledge Concept.

Type	Meta-Path
Student	$SrS : \text{Student} \xrightarrow{\text{reply}} \text{Student}$
	$ScVcS : \text{Student} \xrightarrow{\text{comment}} \text{Video} \xrightarrow{\text{commentedby}} \text{Student}$
	$SwVwS : \text{Student} \xrightarrow{\text{watch}} \text{Video} \xrightarrow{\text{watchedby}} \text{Student}$
	$SIKIS : \text{Student} \xrightarrow{\text{learn}} \text{Knowledge Concept} \xrightarrow{\text{learnedby}} \text{Student}$
Knowledge Concept	$KiViK : \text{Knowledge Concept} \xrightarrow{\text{includedin}} \text{Video} \xrightarrow{\text{include}} \text{Knowledge Concept}$
	$KISIK : \text{Knowledge Concept} \xrightarrow{\text{learnedby}} \text{Student} \xrightarrow{\text{learn}} \text{Knowledge Concept}$

### 3.2. Graph Convolutional Networks (GCN)

Graph convolutional networks learn node representations by aggregating information from neighbouring nodes, but their superior performance usually relies on the homogeneity of the network [22]. Therefore, restricting the head and tail node types to select appropriate meta-paths can be used to mine potential associations between nodes of the same type in heterogeneous information networks so as to learn the semantic representations of nodes under that meta-path with the help of GCN.

#### 3.2.1. Adjacency Matrix

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

With the large number of applications of GCNs, the construction of adjacency matrices has become more and more sophisticated. In order to include information about itself in the process of updating the node representation, a unit matrix  $I$  is often added to the adjacency matrix  $A$ . Also, the adjacency matrix is multiplied by  $D^{-1}$  for normalisation and  $D$  is the degree matrix of matrix  $A + I$ . In this paper, the adjacency matrix is constructed as shown in Equation (1):

$$\tilde{P}^{MP} = D^{-1}(A^{MP} + I) \tag{1}$$

#### 3.2.2. Node Representation Learning

Given a heterogeneous information network  $G = (V, E)$ , in this paper we use layer-by-layer propagation rules to learn the representation of a node under a meta-path  $MP$ , as shown in Equation (2):

$$h^{(l+1)} = Re\ lu(\tilde{P}^{MP}h^lW^l) \tag{2}$$

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 initialised and continuously trained by single-layer GCN, and the node is represented as  $e^{MP} = h_{MP}^1$  after single-layer GCN. Based on Equation (2), for many iterations of training, the initial features  $h^0$  of students and knowledge concepts can be passed to any node.

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

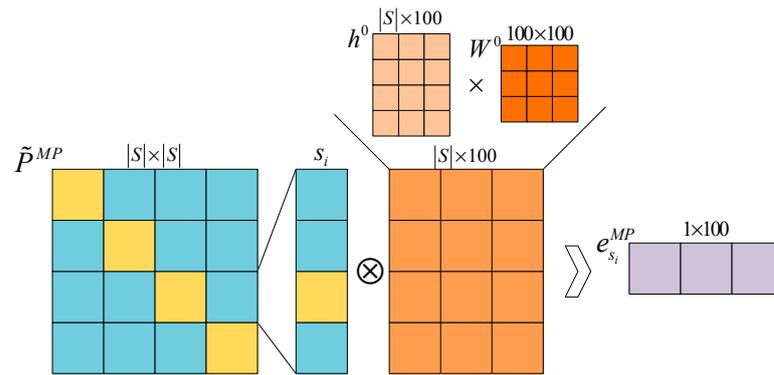


Figure 2. Example of student representation learning.

### 3.3. Attention Mechanism

In deep learning, the introduction of an attention mechanism enables neural networks to automatically learn and select important information in the input, improving the performance and generalisation of the model [23]. The node representations learnt under different meta-paths represent the feature of the node in different contexts, which has different importance in multi-task scenarios as the task changes. In this paper, we use an attentional mechanism to fuse the multi-dimensional feature information in order to generate a final node representation that is more adapted to the task.

Taking the fusion of knowledge concept representations of two knowledge concept meta-paths as an example, the sequence  $\{e_k^{K1S1K}, e_k^{KiViK}\}$  of knowledge concept representations output from the GCN is taken as input, and the formula for calculating the attentional weight of each meta-path is

$$\alpha_k^{MP_i} = \frac{\exp\left(V_k^T \sigma\left(W_k e_k^{MP_i} + b_k\right)\right)}{\sum_{j \in [MP]} \exp\left(V_k^T \sigma\left(W_k e_k^{MP_j} + b_k\right)\right)} \quad (3)$$

where  $V_k^T$ ,  $W_k$ , and  $b_k$  are trainable matrices,  $\sigma(\bullet)$  is the *tanh* activation function, and the output  $\alpha_k^{MP_i}$  denotes the weights of the knowledge concept representation  $e_k^{MP_i}$  under the meta-path  $MP_i$ . Based on the obtained weights, the knowledge concept representations under multiple meta-paths are fused:

$$e_k = \sum_{j \in [MP]} \alpha_k^{MP_j} e_k^{MP_j} \quad (4)$$

where  $e_k$  is the final knowledge concept representation. The visualisation of the fusion of knowledge concept representations by the attention mechanism is shown in Figure 3, where  $|K|$  is the number of knowledge concepts.

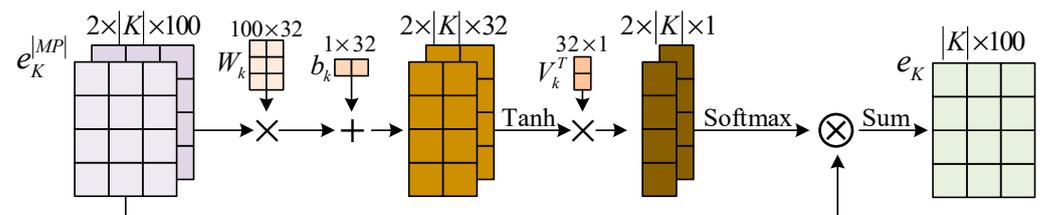


Figure 3. Visualisation of Attentional Mechanisms Integrating Knowledge Concept Representations.

#### 4. Joint Recommendation Model for Knowledge Concepts and Learning Partners Based on Improved MMoE

The KLJRec model architecture is shown in Figure 4, which is divided into three main parts: a GCN-based module for learning student and knowledge concept representations; an LSTM-based module for predicting students' learning interests; and an improved MMoE-based module for joint recommendation of knowledge concepts and learning partners.

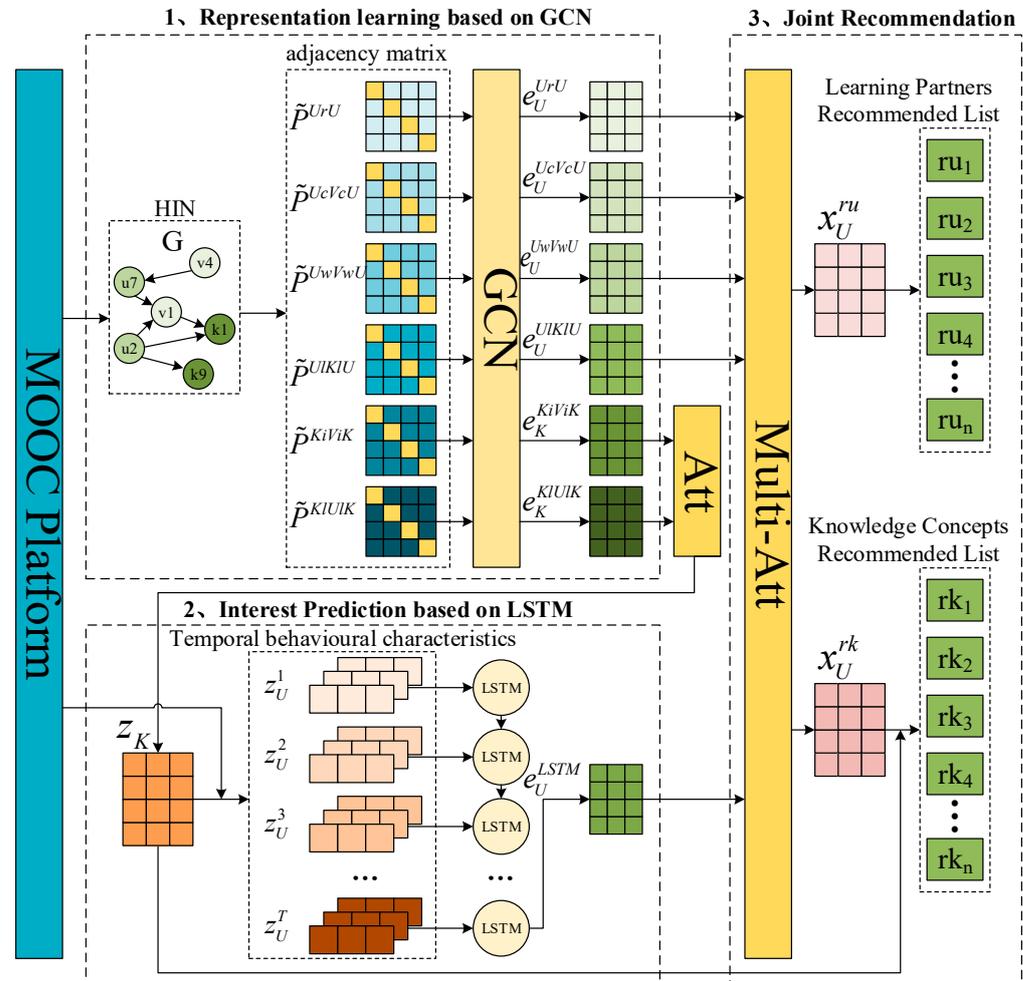


Figure 4. KLJRec model framework diagram.

Based on the MOOC platform, the representation learning module constructs HINs containing three kinds of objects and their correspondences, namely, students, videos, and knowledge concepts, and selects student and knowledge concept meta-paths [MP] to generate the corresponding adjacency matrices  $[\tilde{P}^{MP}]$ , and learns the student representations  $[e_U^{MP}]$  and knowledge concept representations  $[e_K^{MP}]$  under different meta-paths through the GCN model, so as to be prepared for the subsequent fusion of student and knowledge concept representations by using the attention mechanism. The knowledge concept representation  $[e_K^{MP}]$  is used as input to the attention network to obtain the final knowledge concept feature  $z_K$ . The Learning Interest Prediction Module obtains the student's temporal behavioural features  $z_K$  based on the student's temporal learning behaviours and knowledge conceptual features, and captures the dependencies in the student's temporal behavioural data through LSTM to predict the student's current learning interest  $e_U^{LSTM}$ . Based on the MMoE framework, multiple attention networks are used in the joint recommendation module to fuse the students' representations under different meta-paths as well as the students' interest features in order to learn the students' representations  $x_U^{rk}$ .

and  $x_U^r$  that satisfy the two tasks of knowledge concept recommendation and learning partner recommendation, respectively, and to generate a list of the students' personalised knowledge concepts and learning partner recommendations.

#### 4.1. Students and Knowledge Concepts Representation Learning Based on GCNs

As shown in the first part of Figure 4, the representation learning of students and knowledge concepts under the meta-path can be synchronised. The steps are as follows: Firstly, in the constructed heterogeneous information network  $G$ , according to the selected set of meta-paths  $[MP]$ , the set of adjacency matrices under the meta-paths  $[\tilde{P}]$  can be calculated by Formula (1). In the figure, the upper four are the student adjacency matrices  $\tilde{P}^U \in R^{|U| \times |U|}$ , and the lower two are the knowledge concepts adjacency matrices  $\tilde{P}^K \in R^{|K| \times |K|}$ . Secondly, the features  $h^0$  of the initialised student and knowledge concepts are combined with the adjacency matrix  $\tilde{P}^{MP}$  obtained from the different meta-paths  $MP$  and are input into a single-layer GCN, and the representations  $e^{MP}$  of the student and knowledge concepts under  $MP$  are learnt according to Equation (2). Finally, the learnt knowledge concept representation  $\{e_K^{KiViK}, e_K^{KIUIK}\}$  and student representation  $\{e_U^{UrU}, e_U^{UcVcU}, e_U^{UwVwU}, e_U^{UIKIU}\}$  under different meta-paths are saved in the form of sequences, and the weights of different knowledge concept meta-paths are learnt by using the attention mechanism and weighted and merged to obtain the final knowledge concept representation  $Z$ . Meanwhile, the list of student representations is retained in order to prepare for the subsequent generation of the final student representations.

#### 4.2. LSTM-Based Prediction of Student Interest Features

Recurrent Neural Networks (RNN) are mainly used to model sequential data and can effectively mine the temporal and semantic information in the data [24]. LSTM is a variant of RNN that better captures long time sequence dependencies. In this paper, LSTM is used to capture the long dependencies in the first  $T$  temporal behaviours  $\{k_u^1, k_u^2, \dots, k_u^t, \dots, k_u^T\}$  of the student, where  $k_u^t$  denotes the knowledge concept  $k$  that student  $u$  learns at moment  $t$ . The structure of LSTM is shown in Figure 5 below.

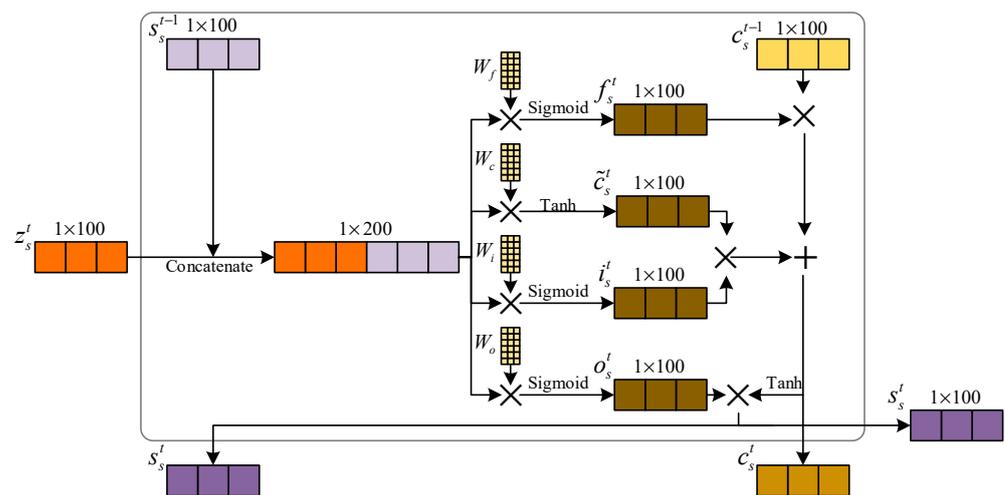


Figure 5. LSTM structure diagram.

Based on the knowledge concept representation  $z_K$  and the student's temporal behaviour sequence  $\{k_u^1, k_u^2, \dots, k_u^t, \dots, k_u^T\}$ , one can obtain the student's temporal behaviour feature as  $\{z_u^1, z_u^2, \dots, z_u^t, \dots, z_u^T\}$ , where  $z_u^t$  is the feature of the knowledge concept  $k_u^t$  learnt by the student at moment  $t$ . LSTM implements 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 through

three gating units, namely, the forgetting gate  $f_u^t$ , the input gate  $i_u^t$ , and the output gate  $o_u^t$ , respectively. The formula is as follows:

$$\begin{aligned} i_u^t &= \text{sigmoid}(W_i \cdot [z_u^t || s_u^{t-1}] + b_i), \\ f_u^t &= \text{sigmoid}(W_f \cdot [z_u^t || s_u^{t-1}] + b_f), \\ o_u^t &= \text{sigmoid}(W_o \cdot [z_u^t || s_u^{t-1}] + b_o), \end{aligned} \tag{5}$$

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 the three gates, the memory cell vector  $H$  and the state vector  $I$  are computed as shown in Equation (6):

$$\begin{aligned} \tilde{c}_u^t &= \text{tanh}(W_c \cdot [z_u^t || s_u^{t-1}] + b_c), \\ c_u^t &= f_u^t \cdot c_u^{t-1} + i_u^t \cdot \tilde{c}_u^t, \\ s_u^t &= o_u^t \cdot \text{tanh}(c_u^t) \end{aligned} \tag{6}$$

The student temporal behavioural features are subjected to LSTM to obtain the state sequence  $\{s_u^1, s_u^2, \dots, s_u^t, \dots, s_u^T\}$ , where the state feature  $s_u^T$  at the last moment is represented as the current interest feature  $e_u^{LSTM}$  of student  $u$ , and it is added to the list of student representations to obtain  $\{e_U^{UrU}, e_U^{UcVcU}, e_U^{UwVwU}, e_U^{UkIU}, e_U^{LSTM}\}$ .

#### 4.3. Joint Recommendation Based on Improved MMoE

There are often multiple learning tasks in an application scenario, and modelling each task individually will incur significant computation and maintenance costs. To reduce the costs due to the increase in the number of models and to take into account the correlation between multiple tasks, multi-task learning is widely used.

MMoE is a multi-task learning framework consisting of a bottom-shared hybrid expert network and multiple gated networks [25]. The input is  $x \in R^{bitsize \times d}$ , where  $d$  is the feature dimension, and the output of the expert network is  $f_i(x), i = 1, 2, \dots, n$ , where  $n$  denotes the number of expert networks. A gated network  $g^j$  is assigned to each task  $j$ . Combining multiple expert network outputs using the gated network yields the desired features  $f^j(x)$  for that task:

$$\begin{aligned} f^j(x) &= \sum_{i=0}^n g^j(x)_i f_i(x) \\ \text{where } g^j(x) &= \text{softmax}(W_{gk}x) \end{aligned} \tag{7}$$

where  $W_{gk} \in R^{n \times d}$  is the trainable matrix and  $g^j(x)_i$  denotes the weight of the expert network  $i$  in task  $j$ .

Based on the MMoE framework, this paper considers the list of student representations  $\{e_U^{UrU}, e_U^{UcVcU}, e_U^{UwVwU}, e_U^{UkIU}, e_U^{LSTM}\}$  as the output of multiple expert networks and replaces the gated network with an attentional mechanism that is more capable of capturing important information. Using two attention networks, the weights of different student representations are calculated separately according to Equation (3), and the student representations are weighted and fused using Equation (4) to generate the final student representation  $x_U^{rk}$  for knowledge concept recommendation and student representation  $x_U^{ru}$  for learning partner recommendation.

Based on the student representations  $x_U^{rk}, x_U^{ru}$  and the knowledge concept representation  $z_K$ , the final student's preference  $y_{UK}$  for the knowledge concept and student's preference  $y_{UU}$  for the student are generated with the following formulas:

$$\begin{aligned} y_{UK} &= x_U^{rk} M_1 z_U + b_K \\ y_{UU} &= x_U^{ruT} M_2 x_U^{ru} + b_U \end{aligned} \tag{8}$$

where  $M_1, M_2$  is the trainable matrix so that  $x_U^k, z_U$  and  $x_U^u$  can be in the same space, and  $b_K, b_U$  are two bias terms to make the prediction more accurate.

In this paper, the loss function is constructed based on Bayesian personalised ranking [26], and the basic idea is to make the ratings of the nodes that students have interacted with higher than those of the nodes that have not interacted with them. The specific loss function is shown in Equation (9):

$$L = \sum_{(u,a,b) \in U, (i,j) \in K} -\ln(\text{sigmoid}(y_{ua} - y_{ub})) - \ln(\text{sigmoid}(y_{ui} - y_{uj})) + \lambda \|\Theta\|^2 \quad (9)$$

where  $a, i$  is the students and knowledge concepts that Student  $u$  has interacted with, and  $b, j$  is the students and knowledge concepts that Student  $u$  has not interacted with, respectively.  $y_{ua} - y_{ub}$  and  $y_{ui} - y_{uj}$  are used to calculate the preference difference between the two nodes and increase it by training the loss function. In addition, the L2 regularisation term is added, where  $\lambda$  is the regularisation parameter and  $\Theta$  denotes all trainable parameters.

Algorithm 1 illustrates the basic steps of KLJRec.

---

#### Algorithm 1 KLJRec Algorithm

---

**Input:**  $G = (V, E)$ : network schema of heterogeneous information networks in MOOC platforms;  $S$ : set of students;  $K$ : set of knowledge concepts;  $\{k_U^1, k_U^2, \dots, k_U^t, \dots, k_U^T\}$ : the first  $T$  sequential actions of the students

**Output:**  $y_{UK}$ : students–knowledge concepts preference matrix;  $y_{UU}$ : students–students preference matrix

- 1: According to  $G = (V, E)$ , select the appropriate meta-path set  $[MP]$  of the students and knowledge concepts;
  - 2: **for** each  $MP \in [MP]$  perform
  - 3:     Calculate  $\bar{P}$  using Formula (1);
  - 4:     Initialise the initial features of the students or knowledge concept  $h^0$ ;
  - 5:     Calculate  $e_K^{MP}$  or  $e_U^{MP}$  according to the definition 3.2.2;
  - 6: **end**
  - 7: According to  $[e_K]$ , combined with definition 3.3, the node representation of knowledge concepts  $z_K$  is generated.
  - 8: **for** each  $u \in U$  perform
  - 9:     According to  $\{k_u^1, k_u^2, \dots, k_u^t, \dots, k_u^T\}$  and  $z_K$ , the characteristics of student temporal behaviour  $\{z_u^1, z_u^2, \dots, z_u^t, \dots, z_u^T\}$  are obtained;
  - 10:     The state sequence  $\{s_u^1, s_u^2, \dots, s_u^t, \dots, s_u^T\}$  is obtained by LSTM, and use the latest state  $s_u^T$  as the interest features of student  $e_u^{LSTM}$ ;
  - 11: **end**
  - 12:     Adding student interest features  $e_U^{LSTM}$  to the list of student representations yields  $[e_U]$ ;
  - 13:     Based on the MMoE framework, according to the list of student representations  $[e_U]$  and definition 3.3, the student representations  $x_U^k, x_U^u$  used for knowledge concept recommendation and learning partner recommendation are generated;
  - 14:     According to  $x_U^k, x_U^u, z_K$ , the final students–knowledge concepts preference matrix  $y_{UK}$  and the student-to-student preference matrix  $y_{UU}$  are calculated using Equation (9).
- 

## 5. Experimental Section

### 5.1. Datasets

The KLJRec algorithm proposed in this paper was validated using the MOOCCubeX dataset (available online at <https://github.com/THU-KEG/MOOCubeX>, accessed on 18 March 2024). MOOCCubeX is an open data warehouse for large-scale online education that collects knowledge-centric data on the XuetangX platform, consisting of 4216 MOOC courses, 230,263 videos, 358,265 exercises, 637,572 fine-grained concepts, and more than 2,960,000 behavioural data from 330,294 students [27]. In order to facilitate the tracking of students' learning processes, students and their learning behaviour data in the course ID 'C\_697791' were selected as the preprocessing dataset, and students who watched fewer videos and did not generate student–student interaction data were deleted, so that

206 students and their related learning behaviour data were finally filtered out. In order to verify the effectiveness of the knowledge concept and learning partner recommendations, the interaction data of the last video of each student were used as the test set and the rest of the interaction data were used as the training set according to the students' video viewing orders.

### 5.2. Baseline Models and Assessment Indicators

In order to evaluate the model knowledge concept recommendation performance, in this paper, each student interaction concept in the validation set was combined with 99 non-interaction concepts as a group, and HR@k, NDCG@k, and MRR metrics were computed based on the student preference  $y_{uk}$ , where k was set to 5 and 10. In the MOOCubeX dataset, KLJRec is compared with four benchmark knowledge concept recommendation models, and the benchmark models are introduced in Table 3.

**Table 3.** Knowledge concepts recommended baseline model.

Baseline Model	Model Introduction
MFBR [26]	The method assumes that users have higher preferences for interacted items than for items they have not interacted with, and uses implicit feedback data to solve the problem of ranking among recommended items
metapath2vec [28]	The method uses random wandering in a heterogeneous network to construct a neighbourhood of nodes, which is then trained using a skip-gram model to increase the similarity of nodes between domains
ACKRec [7]	A graph convolutional neural network model with an attention mechanism to learn node embeddings under different meta-paths in a heterogeneous network species and obtains final node representations by fusion with an attention mechanism
MOOCIR [29]	A knowledge concept recommendation model based on heterogeneous information networks that uses graph convolution and attention mechanisms to adaptively learn a network representation of entities and train the model using BPR

To validate the learning partner recommendation effect, the learning partner recommendation performance is measured using the precision and recall rates. In the MOOC-CubeX dataset, KLJRec is compared with three benchmark learning recommendation partner models, which are described in Table 4.

**Table 4.** Baseline learning partner recommendation model.

Baseline Model	Model Introduction
MF [30]	The method maps user-item interactions into a low-dimensional potential space and uses the inner product of the user and the item in space to model user interactions
metapath2vec [28]	A representation learning model based on random walks and skip-grams to generate paths in heterogeneous networks using random walks and learns node representations using skip-grams
LPRWHIN [19]	The model proposes a method for automatically extracting and identifying meaningful meta-paths and uses the BPR optimisation framework to learn the importance of different meta-paths of students in order to recommend learning partners

### 5.3. Experimental Environment and Hyperparameter Settings

The experimental environment of this paper is shown in Table 5.

The learning rate of the KLJRec model proposed in this paper was set to 0.01; the regularisation parameter was set to  $\lambda = 1 \times 10^{-8}$ ; the embedding dimensions of the student and knowledge concepts in both the representation learning module and the interest feature prediction module were 100; the hidden layer dimension of the attention mechanism was set to 32; the length of the sequence of the student behaviours input to the LSTM was 10; the batch size for training the model was 1024; and the Adam optimiser was used for gradient descent optimisation.

**Table 5.** Experimental environment.

Experimental Environment	Environment Configuration
Operating systems	Windows 11
CPU	AMD Ryzen 5 5600 H with Radeon Graphics
Video Cards	GeForce RTX 3050
RAM	16 GB
ROM	512 GB
Programming Languages	Python 3.6
Framework	Tensorflow

#### 5.4. Evaluation of Model Recommendation Performance and Analysis of Results in Multitasking Scenarios

Tables 6 and 7 show the performance metrics of the KLJRec model with the baseline model on the MOOCCubeX dataset for the knowledge concept recommendation task and the learning partner recommendation task, as analysed below:

**Table 6.** Results of the comparison of the KLJRec model with the baseline knowledge conceptual model on the MOOCCubeX dataset. The best performing scores are shown in bold.

Model	HR@5	HR@10	NDCG@5	NDCG@10	MRR
MFBPR	0.2922	0.4566	0.1914	0.2442	0.2019
metapath2vec	0.3414	0.4779	0.2389	0.2830	0.2421
ACKRec	0.3213	0.4686	0.2127	0.2604	0.2164
MOOCIR	0.3830	<b>0.5529</b>	0.2448	<b>0.3001</b>	0.2403
KLJRec	<b>0.3910</b>	0.5361	<b>0.2523</b>	0.2995	<b>0.2438</b>

**Table 7.** Results of the comparison of the KLJRec model with the benchmark learning partner recommendation model on the MOOCCubeX dataset. The best performing scores are shown in bold.

Model	Precision									
	P@1	P@2	P@3	P@4	P@5	P@6	P@7	P@8	P@9	P@10
MF	0.022	0.022	0.015	0.028	0.031	0.033	<b>0.035</b>	<b>0.036</b>	<b>0.032</b>	0.029
metapath2vec	0.156	<b>0.100</b>	<b>0.067</b>	<b>0.050</b>	<b>0.040</b>	0.033	0.032	0.028	0.030	<b>0.031</b>
LPRWHIN	<b>0.178</b>	<b>0.100</b>	<b>0.067</b>	<b>0.050</b>	<b>0.040</b>	<b>0.037</b>	<b>0.035</b>	0.031	0.027	0.024
KLJRec	0.156	<b>0.100</b>	<b>0.067</b>	<b>0.050</b>	<b>0.040</b>	0.033	0.032	0.028	0.030	<b>0.031</b>
Model	Recall									
	R@1	R@2	R@3	R@4	R@5	R@6	R@7	R@8	R@9	R@10
MF	0.013	0.026	0.026	0.065	0.091	0.117	<b>0.143</b>	<b>0.169</b>	<b>0.169</b>	0.169
metapath2vec	0.091	<b>0.117</b>	<b>0.117</b>	<b>0.117</b>	<b>0.117</b>	0.117	0.130	0.130	0.156	<b>0.182</b>
LPRWHIN	<b>0.104</b>	<b>0.117</b>	<b>0.117</b>	<b>0.117</b>	<b>0.117</b>	<b>0.130</b>	<b>0.143</b>	0.143	0.143	0.143
KLJRec	0.091	<b>0.117</b>	<b>0.117</b>	<b>0.117</b>	<b>0.117</b>	0.117	0.130	0.130	0.156	<b>0.182</b>

1. KLJRec works better in the knowledge concept recommendation task compared to MFBPR based on matrix decomposition. This demonstrates the importance of mining potential associations between nodes and learning entity representations based on heterogeneous information networks.
2. The metapath2vec model is based on random wandering and skip-gram mining nodes' node representations under multiple meta-paths, and KLJRec generates node representations by fusing nodes' multi-dimensional semantic information through the attention mechanism, which is better overall than metapath2vec in the two recommending tasks, proving that the attention mechanism is better able to balance the influence of different meta-paths on nodes.
3. Both KLJRec and ACKRec models use graph convolution and attention mechanisms to learn node representations. KLJRec works better because before using the attention

mechanism, this paper also learns the student’s current interest features, which proves the role of the student’s interest features in modelling the learner.

4. The method proposed in this paper outperforms the MOOCIR model in HR@5 and NDCG@5, but declines relative to the MOOCIR model in the HR@10 and NDCG@10 metrics, which indicates that the model in this paper is capable of strong upfront prediction, which is a result of the model taking into account the user’s current interest characteristics.
5. Compared to the MF model, the model in this paper uses student chronological behavioural data to take into account the student’s current interest characteristics in the learning partner recommendation task, and thus focuses more on the pre-partner recommendation task, resulting in a decrease in the model’s performance in the top7–top9 learning partner recommendation.
6. The performance of KLJRec in the learning partner recommendation task is lower than that of the LPRWHIN model, which is mainly because the LPRWHIN model learns for a single task and can better optimise the representation of learners in the learning partner recommendation task. Meanwhile, the student–student interaction data are more sparse compared to the human–computer interaction data, and modelling the learner in this paper with multi-dimensional information in mind leads to a reduction in the weight of the student–student interaction data, which interferes with the final results.
7. Most current recommendation models make recommendations for a single learning resource and are difficult to scale in multi-tasking scenarios. The method proposed in this paper outperforms the baseline model overall in both tasks, which demonstrates that the model in this paper can be effectively applied in multi-tasking scenarios and proves the importance of accurate multi-dimensional modelling of learners for multi-tasking optimisation.

### 5.5. Ablation Experiment

#### 5.5.1. Effects of Human–Computer Interaction, Student–Student Interaction, and Interest Characteristics on Learner Modelling

The model proposed in this paper models learners from three dimensions: human–computer interaction, student–student interaction, and student interest features. In order to validate the effectiveness of modelling learners from the three dimensions, an ablation study was conducted on the MOOCCubeX dataset. The experimental results are shown in Table 8, in which the three models, KLJRec-dc, KLJRec-ds, and KLJRec-di, ignored the student human–computer interaction information, the student–student interaction information, and the learning interest characteristics, respectively, in the learner modelling stage, and their comparison with KLJRec can effectively validate the influence of multi-dimensional information on the final recommendation results.

**Table 8.** Results of KLJRec’s knowledge concept recommendation on the MOOCCubeX dataset with and without HCI information, student–student interaction information, and learning interest features. The best performing scores are shown in bold.

Model	HR@5	HR@10	NDCG@5	NDCG@10	MRR
KLJRec-dc	0.3446	0.5121	0.2187	0.2730	0.2185
KLJRec-ds	0.3840	0.5229	0.2446	0.2898	0.2358
KLJRec-di	0.3578	0.5234	0.2272	0.2809	0.2253
KLJRec	<b>0.3910</b>	<b>0.5361</b>	<b>0.2523</b>	<b>0.2995</b>	<b>0.2438</b>

Table 8 shows that KLJRec-dc is less effective than KLJRec-ds and KLJRec-di in the knowledge concept recommendation task, which suggests that the human–computer interaction information is more important than the student–student interaction information as well as the student interest information for the knowledge concept recommendation task. In addition, Table 9 shows that the KLJRec-ds model, which does not take into account

student–student interactions, performs much lower than the other three models in the learning partner recommendation task, which suggests that the student–student interaction information plays a decisive role in the learning partner recommendation task.

**Table 9.** Results of KLJRec’s learning partner recommendation on the MOOCCubeX dataset with and without HCI information, student–student interaction information, and learning interest features. The best performing scores are shown in bold.

Model	Precision									
	P@1	P@2	P@3	P@4	P@5	P@6	P@7	P@8	P@9	P@10
KLJRec-dc	<b>0.156</b>	<b>0.100</b>	<b>0.074</b>	<b>0.056</b>	<b>0.044</b>	<b>0.037</b>	<b>0.035</b>	<b>0.031</b>	<b>0.032</b>	<b>0.033</b>
KLJRec-ds	0.000	0.011	0.007	<b>0.006</b>	0.004	0.004	0.003	0.003	0.007	0.007
KLJRec-di	0.156	<b>0.100</b>	0.067	<b>0.056</b>	<b>0.044</b>	<b>0.037</b>	<b>0.035</b>	<b>0.031</b>	0.030	0.031
KLJRec	<b>0.156</b>	<b>0.100</b>	0.067	0.050	0.040	0.033	0.032	0.028	0.030	0.031
Model	Recall									
	R@1	R@2	R@3	R@4	R@5	R@6	R@7	R@8	R@9	R@10
KLJRec-dc	<b>0.091</b>	<b>0.117</b>	<b>0.130</b>	<b>0.130</b>	<b>0.130</b>	<b>0.130</b>	<b>0.143</b>	<b>0.143</b>	<b>0.169</b>	<b>0.195</b>
KLJRec-ds	0.000	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.039	0.039
KLJRec-di	<b>0.091</b>	<b>0.117</b>	0.117	<b>0.130</b>	<b>0.130</b>	<b>0.130</b>	<b>0.143</b>	<b>0.143</b>	0.156	0.182
KLJRec	<b>0.091</b>	<b>0.117</b>	0.117	0.117	0.117	0.117	0.130	0.130	0.156	0.182

Comprehensively, Tables 8 and 9 show that in the learning partner recommendation task, due to the inclusion of human–computer interaction information and student interest features, the weight of student–student interaction data is reduced to a certain extent, which leads to a slight decrease in the recommendation effect of the KLJRec model, but the indicators of the KLJRec model are all better than those of the KLJRec-dc, the KLJRec-ds, and the KLJRec-di models in the knowledge concept recommendation task. Thus, the overall performance of KLJRec in the multitasking scenario was higher than that of the KLJRec-dc, KLJRec-ds, and KLJRec-di models, which demonstrates the effectiveness of modelling learners from multiple dimensions.

### 5.5.2. Comparison of Gate Functions and Attention Mechanisms

The model in this paper is improved based on the multitasking framework MMoE, for example, replacing the gated network in MMoE with an attention network, which is better at capturing important information compared to the gated network. In order to verify the effect of the attention network, the attention mechanism in the model of this paper is replaced by the gated network to obtain the variant model KLJRec-g, and it is compared with the KLJRec model. The experimental results are shown in Tables 10 and 11. The student features in KLJRec-g1 are used as inputs to the gated network and are set as randomly initialised trainable matrices, and the student features in the KLJRec-g2 model are the user’s interaction behaviour matrices.

**Table 10.** Results of KLJRec’s Knowledge Concept Recommendation using Gated Networks or using Attention Mechanisms on the MOOCCubeX Dataset. The best performing scores are shown in bold.

Model	HR@5	HR@10	NDCG@5	NDCG@10	MRR
KLJRec-g <sub>1</sub>	0.2675	0.4018	0.1892	0.2323	0.2008
KLJRec-g <sub>2</sub>	0.2852	0.4211	0.1932	0.2371	0.2011
KLJRec	<b>0.3910</b>	<b>0.5361</b>	<b>0.2523</b>	<b>0.2995</b>	<b>0.2438</b>

**Table 11.** Results of KLJRec’s Learning Partner Recommendations using Gated Networks or using Attention Mechanisms on the MOOCCubeX Dataset. The best performing scores are shown in bold.

Model	Precision									
	P@1	P@2	P@3	P@4	P@5	P@6	P@7	P@8	P@9	P@10
KLJRec-g <sub>1</sub>	0.044	0.033	0.022	0.017	0.013	0.011	0.010	0.008	0.007	0.007
KLJRec-g <sub>2</sub>	0.000	0.000	0.015	0.017	0.013	0.011	0.010	0.008	0.007	0.007
KLJRec	<b>0.156</b>	<b>0.100</b>	<b>0.067</b>	<b>0.050</b>	<b>0.040</b>	<b>0.033</b>	<b>0.032</b>	<b>0.028</b>	<b>0.030</b>	<b>0.031</b>
Model	Recall									
	R@1	R@2	R@3	R@4	R@5	R@6	R@7	R@8	R@9	R@10
KLJRec-g <sub>1</sub>	0.026	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039
KLJRec-g <sub>2</sub>	0.000	0.000	0.026	0.039	0.039	0.039	0.039	0.039	0.039	0.039
KLJRec	<b>0.091</b>	<b>0.117</b>	<b>0.117</b>	<b>0.117</b>	<b>0.117</b>	<b>0.117</b>	<b>0.130</b>	<b>0.130</b>	<b>0.156</b>	<b>0.182</b>

From Tables 10 and 11, it can be found that KLJRec outperforms KLJRec-g in both tasks, which fully demonstrates the ability of the attention network to capture important information. In a multi-task model, treating user features with different weights for each task can effectively improve the accuracy of user modelling for that task, thus achieving parallel, high-quality completion of multiple tasks.

### 5.5.3. Comparison of Single-Tasking and Multi-Tasking

In order to verify the advantages of multi-task optimisation over single-task optimisation, this paper compares the KLJRec model with the KRec and LRec models, where the KRec and LRec models denote the variants of the model under a single knowledge concept recommender task and a single learning partner recommender task, respectively. The experimental results are shown in Tables 12 and 13.

**Table 12.** Comparison results between the KLJRec model and the KRec model on the MOOCCubeX dataset for the knowledge concept recommendation task. The best performing scores are shown in bold.

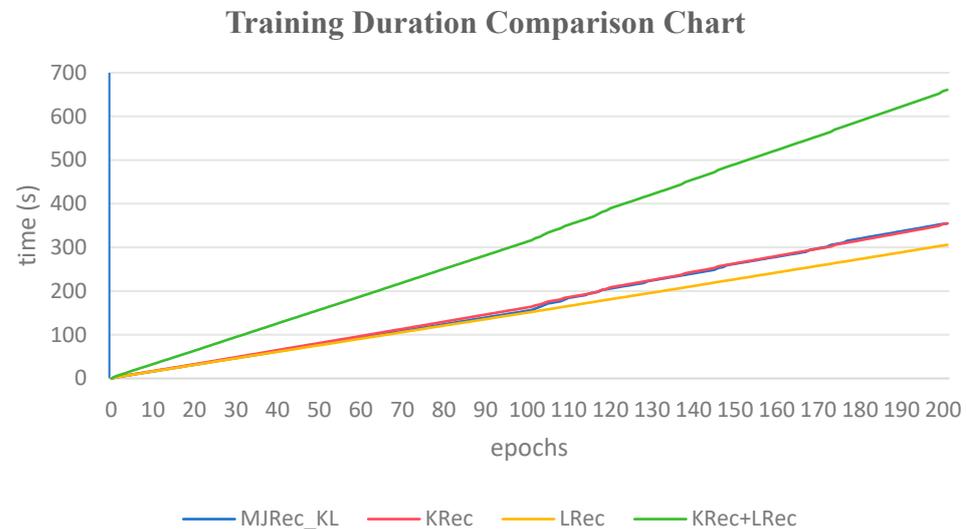
Model	HR@5	HR@10	NDCG@5	NDCG@10	MRR
KRec	0.3473	0.5092	0.2250	0.2772	0.2246
KLJRec	<b>0.3910</b>	<b>0.5361</b>	<b>0.2523</b>	<b>0.2995</b>	<b>0.2438</b>

**Table 13.** Comparison results between the KLJRec model and the LRec model on the MOOCCubeX dataset for the learning partner recommendation task. The best performing scores are shown in bold.

Model	Precision									
	P@1	P@2	P@3	P@4	P@5	P@6	P@7	P@8	P@9	P@10
LRec	0.067	0.056	0.044	0.039	0.031	0.030	0.029	0.025	0.025	0.024
KLJRec	<b>0.156</b>	<b>0.100</b>	<b>0.067</b>	<b>0.050</b>	<b>0.040</b>	<b>0.033</b>	<b>0.032</b>	<b>0.028</b>	<b>0.030</b>	<b>0.031</b>
Model	Recall									
	R@1	R@2	R@3	R@4	R@5	R@6	R@7	R@8	R@9	R@10
LRec	0.039	0.065	0.078	0.091	0.091	0.104	0.117	0.117	0.130	0.143
KLJRec	<b>0.091</b>	<b>0.117</b>	<b>0.117</b>	<b>0.117</b>	<b>0.117</b>	<b>0.117</b>	<b>0.130</b>	<b>0.130</b>	<b>0.156</b>	<b>0.182</b>

As can be seen in Tables 12 and 13, the performance of the multi-task optimisation model is improved in both the knowledge concept recommendation task and the learning partner recommendation task compared to single-task optimisation. This suggests that a multi-task optimisation model can reduce the risk of overfitting in a single task based on the association between multiple tasks, enabling more accurate learner modelling. In addition,

the training duration of single-task optimisation and multi-task optimisation was compared in the same experimental environment, which fully demonstrates the computational cost advantage of the multi-task optimisation model over the single-task optimisation model. A comparison of the training durations for single-task and multi-task is shown in Figure 6.



**Figure 6.** Comparison of single-task and multi-task training duration graph.

## 6. Conclusions

The algorithm proposed in this paper has been extensively experimented on the MOOCubeX dataset, and the experimental results comparing the state-of-the-art knowledge concept recommendation algorithms and learning partner recommendation algorithms show that KLJRec has the best overall performance in the knowledge concept recommendation task, and only performs lower than LPRWHIN in the learning partner recommendation task. This demonstrates the effectiveness of this paper's algorithm in modelling learners using the three dimensions of human–computer interaction, student–student interaction, and interest characteristics, and shows the reliability of generating learner representations that meet the needs of different tasks with the help of multi-attention mechanisms. In addition, the model was subjected to a large number of ablation studies, which not only verified the effects of human–computer interaction information, student–student interaction information, and students' interest characteristics on the model performance, but also verified the advantages of multi-task optimization over single-task optimization in terms of recommendation performance and computational cost. The algorithm in this paper can be well applied in multi-tasking scenarios to give learners richer personalised recommendation services while saving computational costs; however, since this algorithm models learners using multiple dimensions, the algorithm requires high data integrity, and the data imbalance problem will have an impact on the recommendation effect of the algorithm in this paper.

This study focuses on the problem of personalised learning resource recommendation for MOOC platforms. KLJRec depicts a more accurate user profile in terms of students' human–computer interactions, student–student interactions, and interest profiles, which can improve recommendation accuracy and reduce the risk of task overfitting. Meanwhile, based on the shared expert network, the attention network is used to train the learning of different tasks individually, which reduces the overhead arithmetic while achieving the joint recommendation of knowledge concepts and learning partners for students. Knowledge concept recommendation can effectively alleviate the problem of knowledge disorientation faced by students in the process of online learning, meanwhile, taking into account the loneliness that students are prone to in the process of online learning, the recommendation of learning partners can compensate for the social deficits in the process of students' online

learning. In future research, the concept of knowledge and joint recommendation of learning partners can be considered for application in hybrid smart classroom platforms to model students in more complex environments and provide more personalised learning resource recommendation services, thus assisting teachers to improve course design and enhance teaching effectiveness.

**Author Contributions:** Conceptualisation, Z.S. and Y.C.; methodology, Y.C.; software, Y.C.; validation, H.W., J.M. and H.Z.; formal analysis, H.W.; investigation, J.L.; resources, Z.S.; data curation, Y.C.; writing—original draft preparation, Y.C.; writing—review and editing, Z.S.; visualisation, Y.C.; supervision, Z.S.; project administration, Z.S.; funding acquisition, Z.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (62177012, 61967005, 62267003) and was supported by the Project of Guangxi Wireless Broadband Communication and Signal Processing Key Laboratory (GXKL06240107).

**Data Availability Statement:** MOOCCubeX data used can be accessed at <https://github.com/THU-KEG/MOOCcubeX> (accessed on 1 March 2024).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Faroughi, A.; Moradi, P. MOOCs Recommender System with Siamese Neural Network. In Proceedings of the 9th International and the 15th National Conference on E-Learning and E-Teaching (ICeLeT), Sanandaj, Iran, 9–11 March 2022; pp. 1–6.
- Picado, A.; Finamore, C.; Santos, A.M.; Antunes, C. Students Temporal Profiling and e-Learning Resources Recommendation Based on NLP's Terms Extraction. In Proceedings of the 2022 IEEE International Conference on Data Mining Workshops (ICDMW), Orlando, FL, USA, 28 November–1 December 2022; pp. 264–273.
- Tian, X.; Liu, F. Capacity Tracing-Enhanced Course Recommendation in MOOCs. *IEEE Trans. Learn. Technol.* **2021**, *14*, 313–321. [[CrossRef](#)]
- Shrimali, H.; Saxena, R. Kavita Content based Video Recommendation System. In Proceedings of the 2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT), Jaipur, India, 19–20 January 2023; pp. 1–3.
- Chuang, A.-C.; Huang, N.-F.; Tzeng, J.-W.; Lee, C.-A.; Huang, Y.-X.; Huang, H.-H. MOOCERS: Exercise Recommender System in MOOCs Based on Reinforcement Learning Algorithm. In Proceedings of the 2021 8th International Conference on Soft Computing & Machine Intelligence (ISCMI), Cairo, Egypt, 26–27 November 2021; pp. 186–190.
- Wang, X.; Jia, L.; Guo, L.; Liu, F. Multi-aspect heterogeneous information network for MOOC knowledge concept recommendation. *Appl. Intell.* **2022**, *53*, 11951–11965. [[CrossRef](#)]
- Gong, J.; Wang, S.; Wang, J.; Feng, W.; Peng, H.; Tang, J.; Yu, P.S. Attentional Graph Convolutional Networks for Knowledge Concept Recommendation in MOOCs in a Heterogeneous View. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '20, Virtual Event, 11–15 July 2021.
- Miao, J.; Ma, L. Students' online interaction, self-regulation, and learning engagement in higher education: The importance of social presence to online learning. *Front. Psychol.* **2022**, *13*, 815220. [[CrossRef](#)] [[PubMed](#)]
- Shao, M.; Jiang, W.; Zhang, L. FRFP: A friend recommendation method based on fine-grained preference. In Proceedings of the Smart City and Informatization: 7th International Conference, iSCI 2019, Guangzhou, China, 12–15 November 2019; Springer: Singapore, 2019; pp. 35–48.
- Hu, Q.; Han, Z.; Lin, X.; Huang, Q.; Zhang, X. Learning peer recommendation using attention-driven CNN with interaction tripartite graph. *Inf. Sci.* **2019**, *479*, 231–249. [[CrossRef](#)]
- Ju, C.; Zhu, Y.; Wang, C. Knowledge Concept Recommendation Model for MOOCs with Local Sub-graph Embedding. In Proceedings of the 2022 International Conference on Automation, Robotics and Computer Engineering (ICARCE), Wuhan, China, 16–17 December 2022; pp. 1–8.
- Alatrash, R.; Chatti, M.A.; Ain, Q.U.; Fang, Y.; Joarder, S.; Siepmann, C. ConceptGCN: Knowledge concept recommendation in MOOCs based on knowledge graph convolutional networks and SBERT. *Comput. Educ. Artif. Intell.* **2024**, *6*, 100193. [[CrossRef](#)]
- Liu, H.; Ding, J.; Yang, L.T.; Guo, Y.; Wang, X.; Deng, A. Multi-Dimensional Correlative Recommendation and Adaptive Clustering via Incremental Tensor Decomposition for Sustainable Smart Education. *IEEE Trans. Sustain. Comput.* **2019**, *5*, 389–402. [[CrossRef](#)]
- Wu, D.; Tang, M.; Zhang, S.; You, A.; Gao, W. KPRLN: Deep knowledge preference-aware reinforcement learning network for recommendation. *Complex Intell. Syst.* **2023**, *9*, 6645–6659. [[CrossRef](#)]
- Gong, J.; Wan, Y.; Liu, Y.; Li, X.; Zhao, Y.; Wang, C.; Lin, Y.; Fang, X.; Feng, W.; Zhang, J.; et al. Reinforced MOOCs Concept Recommendation in Heterogeneous Information Networks. *ACM Trans. Web* **2023**, *17*, 1–27. [[CrossRef](#)]
- Bouchet, F.; Labarthe, H.; Yacef, K.; Bachelet, R. Comparing peer recommendation strategies in a MOOC. In Proceedings of the Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, Bratislava, Slovakia, 9–12 July 2017; pp. 129–134.

17. Kang, J.; Zhang, J.; Song, W.; Yang, X. Friend relationships recommendation algorithm in online education platform. In Proceedings of the Web Information Systems and Applications: 18th International Conference, WISA 2021, Kaifeng, China, 24–26 September 2021; Springer International Publishing: Berlin, Germany; pp. 592–604.
18. Shao, M.M.; Jiang, W.J.; Wu, J.; Shi, Y.Q.; Yum, T.; Zhang, J. Improving Friend Recommendation for Online Learning with Fine-Grained Evolving Interest. *J. Comput. Sci. Technol.* **2022**, *37*, 1444–1463. [[CrossRef](#)]
19. Shou, Z.; Shi, Z.; Wen, H.; Liu, J.; Zhang, H. Learning Peer Recommendation Based on Weighted Heterogeneous Information Networks on Online Learning Platforms. *Electronics* **2023**, *12*, 2051. [[CrossRef](#)]
20. Yin, Y.; Zheng, W. An Efficient Recommendation Algorithm Based on Heterogeneous Information Network. *Complexity* **2021**, *2021*, 6689323. [[CrossRef](#)]
21. Li, T.; Su, X.; Liu, W.; Liang, W.; Hsieh, M.-Y.; Chen, Z.; Liu, X.; Zhang, H. Memory-augmented meta-learning on meta-path for fast adaptation cold-start recommendation. *Connect. Sci.* **2021**, *34*, 301–318. [[CrossRef](#)]
22. Duran, P.G.; Karatzoglou, A.; Vitria, J.; Xin, X.; Arapakis, I. Graph Convolutional Embeddings for Recommender Systems. *IEEE Access* **2021**, *9*, 100173–100184. [[CrossRef](#)]
23. Sun, S.; Tang, Y.; Dai, Z.; Zhou, F. Self-Attention Network for Session-Based Recommendation with Streaming Data Input. *IEEE Access* **2019**, *7*, 110499–110509. [[CrossRef](#)]
24. Xue, H.; Yang, L.; Jiang, W.; Wei, Y.; Hu, Y.; Lin, Y. Modeling Dynamic Heterogeneous Network for Link Prediction using Hierarchical Attention with Temporal RNN. *arXiv* **2020**, arXiv:2004.01024.
25. Ma, J.; Zhao, Z.; Yi, X.; Chen, J.; Hong, L.; Chi, E.H. Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '18, London, UK, 19–23 August 2018; pp. 1930–1939.
26. Rendle, S.; Freudenthaler, C.; Gantner, Z.; Schmidt-Thieme, L. BPR: Bayesian personalized ranking from implicit feedback. *arXiv* **2009**, arXiv:1205.2618.
27. Yu, J.; Wang, Y.; Zhong, Q.; Luo, G.; Mao, Y.; Sun, K.; Feng, W.; Xu, W.; Cao, S.; Zeng, K. MOOCCubeX: A Large Knowledge-centered Repository for Adaptive Learning in MOOCs. In Proceedings of the 30th ACM International Conference on Information and Knowledge Management, CIKM '21, Virtual Event, 1–5 November 2021.
28. Dong, Y.; Chawla, N.V.; Swami, A. metapath2vec: Scalable representation learning for heterogeneous networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 135–144.
29. Piao, G. Recommending Knowledge Concepts on MOOC Platforms with Meta-path-based Representation Learning. In Proceedings of the International Educational Data Mining Society, Online, 30 June 2021.
30. Koren, Y.; Bell, R.; Volinsky, C. Matrix factorization techniques for recommender systems. *Computer* **2009**, *42*, 30–37. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.