Exploring the Effectiveness of Student Behavior in Prerequisite Relation Discovery for Concepts

Jifan Yu[1], Hanming Li[1], Gan Luo[2], Yankai Lin[3], Peng Li[1], Jianjun Xu[4], Lei Hou[1](✉), and Bin Xu[1]

[1] Tsinghua University, Haidian District, Beijing, China
{yujf21,lhm22}@mails.tsinghua.edu.cn,
{houlei,xubin}@tsinghua.edu.cn
[2] Meituan Inc., Beijing, China
[3] Renmin University of China, Beijing, China
[4] Beijing Caizhi Technology Co., Ltd., Beijing, China
xjj@czkj1010.com

Abstract. What knowledge should a student grasp before beginning a new MOOC course? To answer this question, it is essential to automatically discover prerequisite relations among course concepts. Although researchers have devoted intensive efforts to detecting such relations by analyzing various types of information, there are few explorations of utilizing student behaviors in this task. In this paper, we investigate the effectiveness of student behaviors in prerequisite relation discovery. Specifically, we first construct a novel dataset to support the study, and then formally define four typical student behavior patterns based on analysis of the dataset. Moreover, We explore how this behavior information can be utilized to enhance existing methods, including feature-based and graph-based, via extensive experiments. Experimental results demonstrate that proper modeling of student behaviors can significantly improve the performance of the methods on this task. We hope our study could call for more attention and efforts to explore student behavior for prerequisite relation discovery.

Keywords: Prerequisite relation discovery · Student behavior · MOOCs

1 Introduction

Since the first edition of Robert Gagne's *Principles of Instructional Design* [6] came out, many efforts from pedagogy have suggested that students should grasp prerequisite knowledge before moving forward to learn subsequent knowledge. Such prerequisite relations are described as the dependence among knowledge concepts and are crucial for students to learn, organize, and apply knowledge [19].

J. Yu and H. Li—Equal contribution.

X. Song et al. (Eds.): APWeb-WAIM 2023, LNCS 14334, pp. 359–374, 2024.
https://doi.org/10.1007/978-981-97-2421-5_24

Figure 1 shows an example of the prerequisite relations in Massive Open Online Courses (MOOCs). For a student who wants to learn the concept "Convolutional Neural Network" (CS224:video18), it is beneficial to previously grasp the prerequisite concepts ("Gradient Descent" and then "Back Propagation Algorithm") for better understanding.

Prerequisite relations play an essential role in intelligent educational applications such as curriculum planning [1], reading list generation [7], etc. Figure 1 illustrates an example where explicit prerequisite relations among concepts (represented in red) enable the recommendation of a coherent and reasonable learning sequence to the student (represented in green or blue).

However, as the quantity of educational resources grows rapidly in the era of MOOCs, it is expensive and ineffective to obtain fine-grained prerequisite relations by expert annotations as before [4]. Therefore, automatically discovering prerequisite relations becomes a rising topic in recent years.

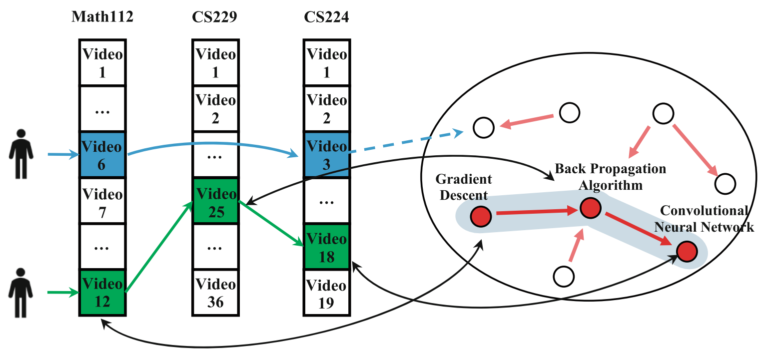

Fig. 1. An application example of prerequisite relations. A student who wants to study "Convolutional Neural Network" can be suggested to follow the prerequisite chain (green path) to learn "Gradient Descent" and "Back Propagation Algorithm" first. (Color figure online)

This task is previously defined as **Prerequisite Relation Discovery of Concepts in MOOCs**. Earlier works on this topic include extracting such relations from the content of MOOC videos [15,17] and the preset orders of MOOC resources [16,21]. Recently, graph-based approaches are introduced into the task, either relying solely on graph [9] or taking advantage of both content- and graph-based methods [27]. However, despite these attempts on this topic, the current methods are still inadequate for direct application in practical scenarios, primarily due to the following challenges

First, unlike the factual relations of general entities (e.g., *Bill Gates* has the relation `founder` with *Microsoft Inc.*), prerequisite relations are more cognitive than factual, which makes it rarely mentioned explicitly in texts and challenging to be captured from MOOC corpus. Second, existing MOOC resources considered prerequisite clues are noisy, e.g., the order of MOOC videos. As a video usually teaches several concepts, it is common that some of these concepts are not prerequisites to the ones in later videos. Therefore, it is crucial to explore more resources to help the discovery of prerequisite relations.

Inspired by the idea from educational psychology that students' learning behaviors are highly related to the cognitive structure of knowledge [3], we investigate leveraging the *video watching behaviors of students* in the task of discovering prerequisite concepts in MOOCs. For supporting the investigation, we collect student behavior data from real MOOC courses and organize expert annotations to construct a novel dataset of prerequisite concept relations. With careful analysis of the data, we then summarize four typical behavior patterns.

Furthermore, to explore the use of behavior patterns to enhance existing feature- and graph-based methods, we design features for each of these patterns, the effectiveness of which is examined through experiments. Results show that incorporating student behaviors effectively enhances existing representative models. We also provide several empirical suggestions for research on related topics.

Our contributions include:

- An investigation on how to leverage student behaviors to extract prerequisite relations of concepts;
- Validation of the effect that existing methods can be enhanced significantly combined with behaviors information;
- A novel benchmark dataset using real courses from MOOC websites[1].

2 Related Work

Our work mainly follows the efforts in discovering *prerequisite relations* among course concepts, aiming to detect the dependence of concepts from MOOC resources. The task of identifying prerequisite relations originates from educational data mining, which could help in automatic curriculum planning [19] and other educational applications [20]. In the area of education, early works discover general prerequisite structures from students' test performance [8,23,25], and these early efforts have mainly focused on discovering the dependence among courses or knowledge units. [24] and [14] further propose to learn more fine-grained prerequisite relations, i.e., the prerequisite relations among concepts. In recent years, detecting prerequisite concepts from courses (especially online courses) has become a rising research topic. Researchers explore various kinds of methods from matrix optimization [16], feature engineering [17], neural networks [21] to the recent trend of graph-based models [9,13,27] to consider the static information of MOOCs (course/video) as indispensable clues for discovering such relations.

There are also some attempts to extract prerequisite relations from other resources, e.g., paper citation networks [7] and textbooks' unit sequences and titles [11]. Recently, the user clickstream of Wikipedia pages [22] is also proven to

[1] https://github.com/HwHunter/Prerequisite-Relation-Discovery-in-MOOCs.

indicate concept dependence. This inspires us to improve prerequisite prediction by considering the user behaviors in MOOCs, which contain more behavioral details and are more relevant to the cognitive learning process.

3 Problem Formulation

In this section, we give fundamental definitions and present the problem formulation for discovering prerequisite relations among course concepts in MOOCs.

A MOOC corpus is composed of courses from MOOCs, denoted as $\mathcal{M} = \{\mathcal{C}_i\}_{i=1}^{|\mathcal{M}|}$, where $\mathcal{C}_i$ indicates the i-th course. Each course includes a sequence of videos, i.e., $\mathcal{C}_i = [v_{ij}]_{j=1}^{|\mathcal{C}_i|}$, where v_{ij} refers to a video with its subtitles from the course. And the **Student Behavior** that we use in this paper is the Video Watching Behaviors $\mathcal{S} = \{(u, v, t)\}$, where each behavior records student $u \in U$ started to watch the video v at time t, and U is the set of all students.

Course dependence is defined as a prerequisite relation between courses [16], denoted as $\mathcal{D} = \{(\mathcal{C}_i, \mathcal{C}_j) \,|\mathcal{C}_i, \mathcal{C}_j \in \mathcal{M}\}$, which indicates that course $\mathcal{C}_i$ is a prerequisite course of $\mathcal{C}_j$. This information is often provided by the teachers when setting up new courses.

Course Concepts refer to the subjects taught in a course (e.g., "LSTM" is a concept of the Deep Learning course). We respectively denote the concepts of a certain video, a course, and the whole MOOC corpus as $\mathcal{K}^v$, $\mathcal{K}^c$ and $\mathcal{K}$. The video concepts $\mathcal{K}_{ij}^v = \left\{c_1, ..., c_{|\mathcal{K}_{ij}^v|}\right\}$ is the concepts taught in course video v_{ij}. As a course consists of several videos, the course concept $\mathcal{K}_i^c = \mathcal{K}_{i1}^v \cup ... \cup \mathcal{K}_{i|\mathcal{C}_i|}^v$. And all the concepts of the MOOC corpus is $\mathcal{K} = \mathcal{K}_1^c \cup ... \cup \mathcal{K}_{|\mathcal{M}|}^c$.

Discovering prerequisite relation of course concepts in MOOCs is formulated as follows: Given the MOOC corpus $\mathcal{M}$, course dependence $\mathcal{D}$, student behaviors $\mathcal{S}$, and the corresponding course concepts $\mathcal{K}$, the goal is to learn a function $\mathcal{L} : \mathcal{K}^2 \rightarrow \{0, 1\}$ that maps a pair of concepts (c_a, c_b), where $c_a, c_b \in \mathcal{K}$, to a binary class indicating whether c_a is a prerequisite of c_b.

4 Data Collection and Annotation

Although there are a few datasets for mining prerequisite relations from online courses [12,17], the information on student behaviors is rarely studied due to the lack of accessible data. To support our investigation, we collect data on courses, videos, and student behaviors from XuetangX[2], a large MOOC website, and organize multi-level annotations to construct a novel prerequisite concept dataset in MOOCs.

Due to the sparsity of prerequisite relationships, it is challenging to ensure data quality while keeping low annotation cost [16]. Hence, we consider two issues in data construction: (1) the connectivity of course concepts, and (2) the effectiveness of annotations. In particular, with the protection of user privacy in mind, we collect and annotate the student behavior data in three stages after constructing the annotation team.

[2] https://www.xuetangx.com/.

Annotation Team: The group of annotators consists of 1 experienced CS professor, 2 CS Ph.D.s, and 8 Students who enrolled in these courses. We also employ the teachers of our selected courses as consultants to deal with disagreements in the annotation process.

Table 1. Statistics of our dataset. As course concepts and students may overlap in different courses, their total number is not a simple numerical addition.

		CS	PL	AI	ALL
#Course		4	4	4	12
#Video		312	222	233	767
#Concept		369	227	377	700
#Pair	*+pos*	672	673	267	1,612
	−neg	1,258	539	218	2,015
#Student		12,094	12,014	3,541	17,587
#Behavior		430,769	337,953	39,136	807,858
Kappa		0.765	0.737	0.769	0.754

Stage 1: MOOC Information Collection. We select 12 sample courses in three domains to collect information on MOOCs, including "Basic Knowledge of Computer Science" (**CS**), "Programming Languages" (**PL**), and "Artificial Intelligence" (**AI**). These courses are selected because their concepts are highly relevant. Then we collect course and student data in three steps: (1) downloading all course materials which include the video orders and subtitles; (2) obtaining the video watching logs of students who participated in these courses during 2017–2019 as user behavior data source, which could help us to infer a student's learning frequency, coverage, watching duration, and other information of a particular video; (3) annotating the dependence of courses. The annotation requires each course pair to be labeled by two students. If there exist conflicts, two Ph.D. students will further label the result. And the final results are confirmed by the professor and corresponding teacher.

Stage 2: Data Processing. Regarding all the subtitles of selected courses as the MOOC corpus, we employ a concept extraction method [18] that is widely used in MOOC-related tasks to obtain concept candidates. For each candidate, two annotators label it as "not course concept" or "course concept", and the disagreements are confirmed by the teacher. Each labeled concept's Wikipedia abstract is also dumped as side information for the reproduction of baseline methods.

Considering the protection of user privacy, we strictly abide by the agreement between the platform and the users, remove sensitive personal information, and anonymize the users into UserIDs. Especially, to alleviate the noises introduced by users' random operation, we filter out the student behavior data by following steps: (i) Remove students who have less than 2 elective courses and watch less than 10 videos. (ii) Delete behavior records that the student watched less than 30% of the video.

Stage 3: Prerequisite Relation Annotation. For all the labeled course concepts, we manually annotate the prerequisite relations among them. A critical challenge in the annotation is the giant quantity and sparsity. If the concept number is n, the candidate pair number is $n(n-1)/2$, which necessitates extensive human labeling efforts. Therefore, we present a two-step strategy to reduce the workload:

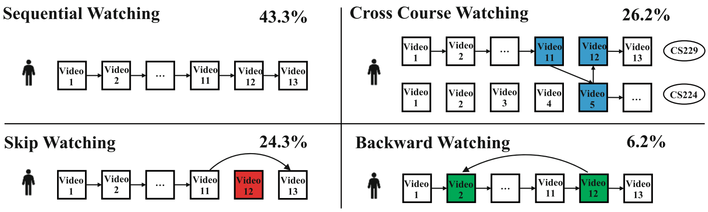

Fig. 2. Four typical patterns when a student watches course videos. The figure shows the proportion of video pairs that match each pattern in all behaviors.

• Step 1: The teacher of the corresponding course leads the annotators to cluster the concepts into several groups, which may maintain possible prerequisite relations. After this step, we get 28 clusters of 700 course concepts, where the largest contains 210 concepts and the smallest contains 14 concepts. We organize the following annotation within these concept clusters.

• Step 2: We generate the candidate concept pairs within the clusters, and sample a small scale of them as golden standard (300). Then we employ them to train existing baselines (i.e., MOOC-RF [18], GlobalF [15], PREREQ [21] and CPR-Recover) as candidate filters. To ensure the Recall, we only filter out the pair if none of the classifiers predict it to be a prerequisite, and preserve the remaining pairs in the annotation.

In the annotation process, two annotators from corresponding domains are asked to label whether concept A is a prerequisite of concept B, i.e. the annotator needs to answer the question of "whether A helps understand B". A pair is labeled as positive only when the two annotators agree. The statistics of our dataset are shown in Table 1, where *#Course*, *#Video*, *#Concept*, *#Pair*, *#Student* are the number of corresponding items and *#Behavior* is the number of video watching records. The *Kappa* statistics of the inter-annotator agreement is 0.754, showing the reliability of the annotation results[3].

5 Student Behavior Patterns and Prerequisite Features

To make full use of the student behavior information, a closer view of the characteristic of data is necessary. In this section, four typical patterns of student

[3] From our experience, such annotation is not easy to determine a perfect standard. E.g. "Stack" and "Queue": Grasping one of them indeed help the understanding of the other, but someone may think these two concepts are not prerequisite. We finally control the Kappa over 0.7 with the help of corresponding teachers, which indicates a good quality of the final presented dataset.

behavior during watching videos in MOOCs are summarized based on our observation and analysis of the dataset. Furthermore, we present several features to model student behaviors, which will be validated with thorough experiments in the following section. The design of our approach is based on the following cognitive learning hypothesis:

Hypothesis 1. *Students tend to follow the prerequisite cognitive structure to learn new knowledge.*

The hypothesis was proposed in educational psychology [3], and was widely applied in prerequisite-driven instructional design [19,20]. We extend this hypothesis by analyzing the clues of the prerequisite concepts implied in student learning orders. Surprisingly, although MOOCs preset the video order, our observation of student behavior data indicates that students often learn MOOC videos in their orders. As shown in Fig. 2, we summarize four typical behavior patterns, and the out-of-pre-order learning behaviors are even more than half of the total (56.72%). To leverage student behaviors in this task, we analyze the causes of these patterns and build several features to model prerequisite relations from student behaviors.

To model the video watch behavior, we first construct a sequence $\mathcal{S}_u$ for each user u from student behavior record $\mathcal{S}$, where the video watch behaviors are sorted in time order. Comparing the preset video order of each $\mathcal{C}_i$ with $\mathcal{S}_u$, we summarize four typical patterns of students' video-watching behaviors: *Sequential Watching*, *Cross Course Watching*, *Skip Watching* and *Backward Watching* as shown in Fig. 2. Before introducing the modeling details, we will begin by providing definitions for the behavior patterns as follows.

Definition 1 (Behavior Pattern). *A behavior pattern $\mathcal{P}$ is formed by one or more video pairs. A video pair (v_i, v_j) belongs to a pattern $\mathcal{P}$ when it matches the conditions of the pattern.*

As the student behavior patterns are at *video* level, we infer the prerequisite features of a concept pair $c_a \in \mathcal{K}_i^v$ and $c_b \in \mathcal{K}_j^v$ by considering videos as bags of course concepts, where $\mathcal{K}_i^v$, $\mathcal{K}_j^v$ correspond to the concepts taught in v_i, v_j, and a concept may be taught in more than one videos.

Over 72.5% of the students' behavior records contain all the four typical patterns in Fig. 2, indicating that they are not accidental. Therefore, we attempt to speculate the causes of these four patterns from the cognitive perspective, and build prerequisite features $f^{\mathcal{P}}$ to model them.

Sequential Watching Pattern. Sequential watching indicates that a student watches videos in the course's preset video order, which indicates that the concepts taught in these videos are in accordance with the prerequisite cognitive structure. To leverage this pattern, we assign prerequisite feature $f_1^{\mathcal{P}}$ for the concepts c_a and c_b as:

$$f_1^{\mathcal{P}}(c_a, c_b) = \sum_{u \in U} \sum_{v_i, v_j \in \mathcal{S}_u} \alpha^{|j-i|} \cdot \frac{\mathtt{Seq}(u, v_i, v_j)}{\max(|\mathcal{K}_i^v|, |\mathcal{K}_j^v|)}, \tag{1}$$

where function $\texttt{Seq}(u, v_i, v_j) = 1$ holds when 1) v_i, v_j are the i-th, j-th videos of a student's watching record $\mathcal{S}_u$ and are in the same course, $j > i$; 2) $c_a \in \mathcal{K}_i^v$ and $c_b \in \mathcal{K}_j^v$ (Otherwise $\texttt{Seq}(u, v_i, v_j) = 0$).

Considering there are multiple concepts taught in each video, we employ $\max(|\mathcal{K}_i^v|, |\mathcal{K}_j^v|)$ to normalize the feature of a certain concept pair. Furthermore, since the distance between watching videos corresponds to their relatedness, we employ an attenuation coefficient of $\alpha \in (0, 1)$ to capture distant dependence from long sequences in this pattern.

Cross Course Watching Pattern. Besides watching in one course, there is a phenomenon that some students choose to watch videos in other courses before continuing on the present study. The main reason is that the knowledge provided by other courses' videos is helpful to study this course. Hence, cross course watching behavior could reflect the dependence between concepts from different courses. The prerequisite feature $f_2^{\mathcal{P}}$ for c_a and c_b is calculated as:

$$f_2^{\mathcal{P}}(c_a, c_b) = \sum_{u \in U} \sum_{v_i, v_j \in \mathcal{S}_u} \alpha^{|j-i|} \cdot \frac{\texttt{Crs}(u, v_i, v_j)}{\max(|\mathcal{K}_i^v|, |\mathcal{K}_j^v|)}, \tag{2}$$

where function $\texttt{Crs}(u, v_i, v_j) = 1$ holds when 1) v_i, v_j are the i-th, j-th videos of a student's watching record $\mathcal{S}_u$ and are in the different course; 2) $c_a \in \mathcal{K}_i^v$ and $c_b \in \mathcal{K}_j^v$ (Otherwise $\texttt{Crs}(u, v_i, v_j) = 0$).

Skipping Watching Pattern. An abnormal student behavior is skipping some videos when learning a course, which drops a hint that the "skipped videos" are not so necessary for the latter videos' comprehension. Given a student behavior sequence $\mathcal{S}_u$ and course video orders $\mathcal{C} = [v_1..v_i...]$, we can detect the skipped video pairs and assign a negative $f_3^{\mathcal{P}}$ for the concept pair c_a and c_b as:

$$f_3^{\mathcal{P}}(c_a, c_b) = -\sum_{u \in U} \sum_{v_i, v_j \in \mathcal{S}_u} \alpha^{|j-i|} \cdot \frac{\texttt{Skp}(u, v_i, v_j)}{\max(|\mathcal{K}_i^v|, |\mathcal{K}_j^v|)}, \tag{3}$$

where function $\texttt{Skp}(u, v_i, v_j) = 1$ holds when 1) v_i, v_j are the i-th, j-th videos of a same course, and $i < j$; 2) v_j is watched by user u but v_i is not watched; 3) $c_a \in \mathcal{K}_i^v$ and $c_b \in \mathcal{K}_j^v$ (Otherwise $\texttt{Skp}(u, v_i, v_j) = 0$).

Backward Watching Pattern. This pattern means a student goes back to a video that he/she watched before. A possible explanation is he/she jumps back to a video to re-learn prerequisite knowledge of the current video. Based on this assumption, we adjust the equation for the feature $f_4^{\mathcal{P}}$ between c_a and c_b.

$$f_4^{\mathcal{P}}(c_a, c_b) = \sum_{u \in U} \sum_{v_i, v_j \in \mathcal{S}_u} \alpha^{|j-i|} \cdot \frac{\texttt{Bck}(u, v_i, v_j)}{\max(|\mathcal{K}_i^v|, |\mathcal{K}_j^v|)}, \tag{4}$$

where function $\texttt{Bck}(u, v_i, v_j) = 1$ holds when 1) v_i, v_j are the i-th, j-th videos of a student behavior record $\mathcal{S}_u$, and $i < j$; 2) v_i is watched again after v_j; 3) $c_a \in \mathcal{K}_i^v$ and $c_b \in \mathcal{K}_j^v$ (Otherwise $\texttt{Bck}(u, v_i, v_j) = 0$).

6 Explore Graph-Based Modeling of Student Behavior

Building *concept graphs* is a common idea in concept mining tasks, including concept extraction and expansion [18,26]. Since the prerequisite relations among concepts are transitive, i.e., if $a \to b, b \to c$ then $a \to c$, previous works also often employ a directed graph to describe the dependence on a set of concepts [5,7]. Moreover, in the past few years, works like [13,27] tend to treat prerequisite relations as edges in a graph. Thus the task is formalized as link prediction and can be tackled with graph-based methods. All this progress inspires us to leverage student behaviors better by building a *concept graph*, defined as:

Definition 2 (Concept Graph). *A concept graph* $\mathcal{G} = (\mathcal{K}, E)$ *is a weighted directed graph, whose nodes are course concepts* $\mathcal{K}$ *and each edge* $e = (c_a \to c_b) \in E$ *is associated with a weight* w_e.

Regarding the prerequisite relation learning as a link prediction problem in a graph, we can leverage the student behavior better by utilizing Graph Convolutional Networks (GCNs) [10] to model information propagation of the concepts. Meanwhile, as several types of MOOC information have been applied to detect prerequisite concepts in previous research, including course dependence [16,21], video order [17], we also design similar concept graphs for these resources. By comparing the model performance of different graphs, we can explore the role of student information more fairly (excluding the factors of the graphical modeling).

In the following part of this section, we introduce the construction of concept graphs and how to conduct prerequisite relation learning on them. We also present experimental results to give insights into the effectiveness of the student behavior in this task.

6.1 Concept Graph Construction

As shown in Fig. 3, we design a concept graph $\mathcal{G}^s$ based on student behaviors as while as $\mathcal{G}^c$ based on course dependence graph and $\mathcal{G}^v$ based on video order graph. As the nodes of these concept graphs are the same course concept $\mathcal{K}$, the only difference is the setting of their edges.

Our graph construction stage's main idea is to assign edge weight for each concept pair in these graphs. After calculating all edges' weights in a graph, we only preserve the edges with positive weights, for they are helpful for relation reasoning.

Concept Graph Based on Student Behavior. To build a concept graph from student behaviors, a straightforward idea is to model the prerequisite clues by combining the extracted features in Sect. 5. Hence, we assign the weight w_e^s for the edge $e = (c_a \to c_b)$ in this graph $\mathcal{G}^s$ as:

$$w_e^s = \sum_{i=1}^{4} f_i^{\mathcal{P}}(c_a, c_b) \times \frac{\texttt{log}(|U|)}{|U|}, \tag{5}$$

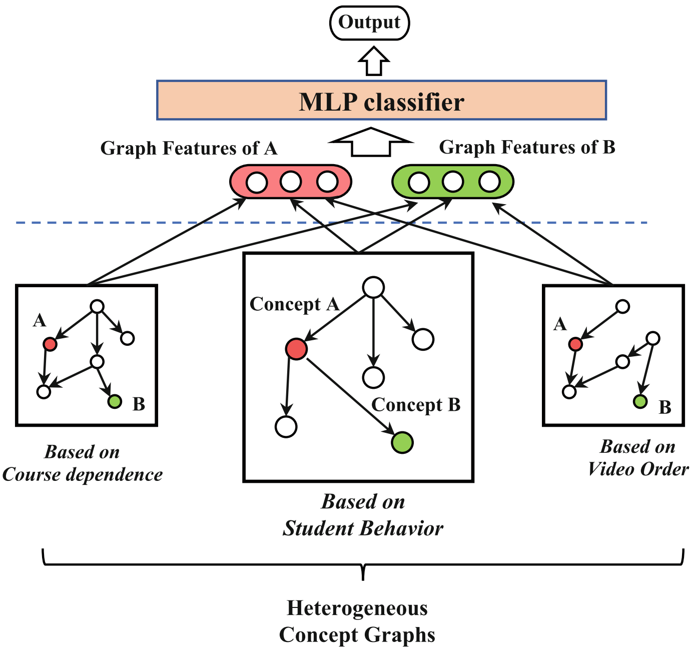

Fig. 3. The framework of our graph-based model. The nodes of each graph are course concepts. Features of concepts are the concatenation of the corresponding node embeddings learned by GCN.

where $f_i^{\mathcal{P}}(i = 1, 2, 3, 4)$ denotes the features of the concept c_a and c_b from the four behavior patterns. $log(|U|)/|U|$ is used to normalize the weight to combine with the other two user-independent graphs.

Except for student behaviors, we also build concept graphs for existing static MOOC prerequisite clues through similar methods, including the dependence among courses and the preset order of videos. By modeling this information, we can more fairly compare the contribution of these clues in graphs and explore whether they can be integrated to further enhance the model.

Concept Graph Based on Course Dependence. Course dependence is widely used in prerequisite learning. When a course is certain to be a prerequisite course of another one, there must be dependent relations between some of its concepts. So we build a concept graph $\mathcal{G}^c$ based on course dependency to exploit this information. Suppose c_a and c_b are respectively concepts of course $\mathcal{C}_i$ and $\mathcal{C}_j$, for an edge $e = (c_a \rightarrow c_b)$ of this concept graph, we can calculate its weight w_e^c as:

$$w_e^c = \sum_{\mathcal{C}_i, \mathcal{C}_j \in \mathcal{M}} \frac{\mathtt{CD}(\mathcal{C}_i, \mathcal{C}_j)}{\max(|\mathcal{K}_i^c|, |\mathcal{K}_j^c|)}, \tag{6}$$

where function $\mathtt{CD}(\mathcal{C}_i, \mathcal{C}_j) = 1$ only when pair $(\mathcal{C}_i, \mathcal{C}_j)$ is in course dependence set $\mathcal{D}$ (otherwise $\mathtt{CD}(\mathcal{C}_i, \mathcal{C}_j) = 0$). We also use $\max(|\mathcal{K}_i^c|, |\mathcal{K}_j^c|)$ to normalize such information to concept-level.

Concept Graph Based on Video Order. Video order indicates the dependence between videos. In general, the previous videos in a course are helpful for the latter ones [21] and such dependence is stronger when two videos are closer. Based on this assumption, when calculating the weight for the concept graph $\mathcal{G}^v$ based on video order, we also apply the attenuation coefficient α to obtain edge weight w_e^v for the edge e between concept c_a and c_b:

$$w_e^v = \sum_{u \in U} \sum_{v_i, v_j \in \mathcal{S}_u} \alpha^{|j-i|} \cdot \frac{\mathtt{VO}(u, v_i, v_j)}{\max(|\mathcal{K}_i^v|, |\mathcal{K}_j^v|)}, \tag{7}$$

where function $\mathtt{VO}(v_i, v_j) = 1$ only when 1) v_i, v_j are the i-th, j-th videos of a same course; 2) $c_a \in \mathcal{K}_i^v$ and $c_b \in \mathcal{K}_j^v$ (otherwise $\mathtt{VO}(v_i, v_j) = 0$).

6.2 Prerequisite Relation Learning

After building concept graphs $\mathcal{G}^c$, $\mathcal{G}^v$, and $\mathcal{G}^s$, we utilize GCNs to reason prerequisite relations in these graphs. We first train the embedding representation of all concept nodes via GCNs. And then we feed the graph embeddings of the concept pair (c_a, c_b) into an MLP classifier to predict whether concept c_a is a prerequisite concept of c_b.

In particular, we initialize the adjacency matrix A of the graph and the feature matrix X of the concept nodes for each graph. The adjacency matrix A, with a size of $|\mathcal{K}|^2$, can be derived from edge weights, e.g., for the adjacency matrix A^s of the student behavior graph $\mathcal{G}^s$, we have $A_{ij}^s = w_e^s$, where w_e^s is the weight of edge $e = (c_i \rightarrow c_j)$. And the $|\mathcal{K}| \times d$ sized feature matrix X of the concept nodes in all graphs is initialized by a pre-trained d-dimension language model, i.e., X_i is the word embedding of the text concept c_i.

The training of GCNs on our directed concept graphs follows the propagation rule shown below, which is an adapted version for directed graphs:

$$Z = \hat{D}^{-1} \hat{A} X \Theta, \tag{8}$$

where Θ is a matrix of filter parameters, Z is the convolved signal matrix, $Z_i = h_i$ is the graph embedding of concept c_i, $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$ and the Laplacian is $\hat{A} = I_N + A$. The other settings are the same with [10].

After the graph-training stage, we input the graph embeddings h_a,h_b of a concept pair (c_a, c_b) into a two-layer MLP followed with a sigmoid function to do classification:

$$\Pr(\mathcal{L}(c_a, c_b) = 1) = \sigma(\max(0, (h_a \oplus h_b)\mathbf{W_1})\mathbf{W_2}), \tag{9}$$

where Pr is the probability, $\sigma(\cdot)$ is the sigmoid function, $\mathbf{W_1} \in \mathcal{R}^{2d \times d}$ and $\mathbf{W_2} \in \mathcal{R}^{d \times 1}$ are trainable matrices, and $\oplus$ denotes vector concatenation.

6.3 Experiment

To have a clear assessment of the effectiveness of the graph-based modeling of student behavior, we conduct experiments on the newly presented dataset in

Table 2. Overall performance. P, R and $F1$ represent *precision*, *recall*, and *F1 score* respectively, and Δ represents the improvement of F1 score after adding student behavior features. $^{+sf}$: enhanced with student behavior features. For graph-based models, $^{+cv}$: $\mathcal{G}^c$ and $\mathcal{G}^v$ are used; $^{+s}$: only $\mathcal{G}^s$ is used; $^{+cvs}$: all $\mathcal{G}^c$, $\mathcal{G}^v$ and $\mathcal{G}^s$ are used.

	P	R	$F1$	Δ
MOOC-RF	0.749	0.584	0.656	–
MOOC-RF^{+sf}	0.755	0.639	0.691	**+3.5**
GlobalF	0.679	0.631	0.650	–
GlobalF^{+sf}	0.710	0.657	0.680	**+3.0**
PREREQ	0.468	0.792	0.567	–
PREREQ^{+sf}	0.511	0.712	0.595	**+2.8**
Seq2Seq	0.706	0.743	0.723	–
Seq2Seq^{+sf}	0.707	0.736	0.720	–0.3
GCN^{+cv}	0.789	0.792	0.790	–
GCN^{+s}	0.762	0.784	0.772	–
GCN^{+cvs}	**0.792**	**0.814**	**0.802**	**+1.2**

Sect. 4. Besides, supplemental experiments are carried out on other typical methods with different features and structures to further verify the general accordance of student behaviors and prerequisite structure. The models we select for the supplemental experiments are described as follows:

- **MOOC-RF**: A widely-used method [17], which extracts seven features from the video and subtitle corpus of MOOCs. We reproduce this method and select Random Forest as the classifier to match its claimed best performance.
- **GlobalF**: This method [15] extracts the graph-based features and text-based features for a certain concept pair. The graph-based features are based on Wikipedia Anchor Links, and the text features are based on the description of concepts. We reproduce this method as a typical baseline.
- **PREREQ**: This method [21] utilizes course dependence and video orders to find prerequisite relations through a siamese network.
- **Seq2Seq**: Several efforts have been put forth to utilize neural models to extract prerequisite relations from text, and we reproduce the sequence-to-sequence model in [2] to encode the concepts' texts as prerequisite features.

For enhancing the above models, we concatenate the student behavior features $f_k^{\mathcal{P}}(k = 1, 2, 3, 4)$ with original features and then utilize the same classifiers in the respective papers to obtain experimental results. During all the experiments, we apply 10-fold cross-validation and balance the training set by oversampling the positive instances. Table 2 presents a summary of the comparative outcomes obtained from different methods. We analyze the performance in the following aspects: (1) **Student behaviors are effective in prerequisite relation discovery**. MOOC-RF, GlobalF, and PREREQ gain significant improvement after adding the extracted student behavior features. It preliminarily proves that the student behaviors imply clues of prerequisite concepts and are useful to prerequisite relation discovery. (2) **The effectiveness of Student Behavior in Graph Modeling**. GCN^{+s} performs better than the Seq2Seq model and has a competitive performance among all baselines. Further, GCN^{+cvs} performs better than GCN^{+cv}, which indicates that except for the improvement of graph-based modeling, student behavior is still beneficial in advanced attempts of prerequisite relation discovery.

6.4 Analysis of Graph Modeling

We also analyze different components of the graph-based modeling, including:

Necessity of Edge Weights. We set all the edge weights to 1 to convert three concept graphs into unweighted ones and present the corresponding results in Table 3. The performance of all three GCN-based models declines severely, especially the most competitive GCN^{+cvs}, indicating the necessity of edge weights.

Table 3. Absolute performance drops without the edge weights of the concept graphs.

	P	R	F_1
GCN^{+cv}	–8.5	–11.1	–9.9
GCN^{+s}	–7.7	–11.8	–9.8
GCN^{+cvs}	**–10.5**	**–14.5**	**–12.6**

Analysis of Different Behavior Patterns. We further investigate the impact of four behavior patterns by only using some patterns when building graph $\mathcal{G}^s$, the performance changes are shown in Table 4, which provides some insights for understanding the student behaviors: (1) Sequential watching covers high-quality prerequisite concept pairs, resulting in a significant improvement of the precision (P). However, such a pattern is not so effective for discovering prerequisite relations that do not match with the preset order of courses, resulting in a relatively small improvement of recall (R); (2) The other three patterns, which do not follow the preset order of courses, improve recall significantly, indicating that they are effective for discovering prerequisite relations those not covered by sequential watching; (3) Therefore, the four behavior patterns are complementary, and all of them help to discover prerequisite concepts.

Table 4. Performance improvement of GCN^{+cvs} compared with GCN^{+cv} when only using one student behavior pattern while building the concept graph $\mathcal{G}^s$. Seq: sequential watching, Crs: cross course watching, Bck: backward watching, and Skp: skip watching.

Pattern	P	R	F_1
Seq.	**+0.6**	+0.5	+0.5
Crs.	+0.3	+2.1	+1.2
Skp.	+0.1	+2.3	+1.1
Bck.	+0.1	**+2.6**	**+1.3**

7 Conclusion and Future Work

In this work, we perform an investigation on employing the students' video-watching behaviors in prerequisite relation discovery of concepts in MOOCs. To support the study, we collect student behaviors and conduct data annotations to build a novel dataset. After analyzing the typical patterns, we design features for each of them and experimentally verify the student behaviors' effectiveness in enhancing existing models, both feature-based and graph-based.

Based on our investigation, we present several promising future directions, including 1) A more detailed analysis of the relationship between user behavior and prerequisite concepts, e.g., dividing the typical patterns into a finer-grained level for analysis. 2) Prerequisite-driven intelligent applications on real MOOCs. And it is also beneficial to develop interactive applications to collect more kinds of user behavior for prerequisite relation discovery.

Acknowledgement. This work is supported by the National Key R&D Program of China (2020AAA0105203). It also got partial support from the National Natural Science Foundation of China (No. 62277033), National Engineering Laboratory for Cyber-learning and Intelligent Technology, and Beijing Key Lab of Networked Multimedia.

References

1. Agrawal, R., Golshan, B., Papalexakis, E.: Datadriven synthesis of study plans. Data Insights Laboratories (2015)
2. Alzetta, C., et al.: Prerequisite or Not Prerequisite? That's the Problem! An NLP-Based Approach for Concept Prerequisites Learning (2019)
3. Ausubel: Educational Psychology: A Cognitive View (1968)
4. Bergan, J.R., Jeska, P.: An examination of prerequisite relations, positive transfer among learning tasks, and variations in instruction for a seriation hierarchy. Contemp. Educ. Psychol. **5**, 203–215 (1980)
5. Brunskill, E.: Estimating Prerequisite Structure from Noisy Data (2011)
6. Gagne, R.M., Briggs, L.J.: Principles of Instructional Design. Holt, Rinehart & Winston, Austin (1974)

7. Gordon, J., Zhu, L., Galstyan, A., Natarajan, P., Burns, G.: Modeling concept dependencies in a scientific corpus. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, vol. 1: Long Papers, pp. 866–875 (2016)
8. Huang, X., Yang, K., Lawrence, V.B.: An efficient data mining approach to concept map generation for adaptive learning. In: Industrial Conference on Data Mining (2015)
9. Jia, C., Shen, Y., Tang, Y., Sun, L., Lu, W.: Heterogeneous graph neural networks for concept prerequisite relation learning in educational data. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 2036–2047 (2021)
10. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR) (2017)
11. Labutov, I., Huang, Y., Brusilovsky, P., He, D.: Semi-supervised techniques for mining learning outcomes and prerequisites. In: SIGKDD. ACM (2017)
12. Li, I., Fabbri, A.R., Tung, R.R., Radev, D.R.: What should i learn first: introducing lecturebank for NLP education and prerequisite chain learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 6674–6681 (2019)
13. Li, I., Fabbri, A.R., Hingmire, S., Radev, D.: R-VGAE: relational-variational graph autoencoder for unsupervised prerequisite chain learning. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 1147–1157 (2020)
14. Liang, C., Wu, Z., Huang, W., Giles, C.L.: Measuring prerequisite relations among concepts. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 1668–1674 (2015)
15. Liang, C., Ye, J., Wang, S., Pursel, B., Giles, C.L.: Investigating active learning for concept prerequisite learning. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
16. Liang, C., Ye, J., Wu, Z., Pursel, B., Giles, C.L.: Recovering concept prerequisite relations from university course dependencies. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)
17. Pan, L., Li, C., Li, J., Tang, J.: Prerequisite relation learning for concepts in MOOCs. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, vol. 1: Long Papers, pp. 1447–1456 (2017)
18. Pan, L., Wang, X., Li, C., Li, J., Tang, J.: Course concept extraction in MOOCs via embedding-based graph propagation. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing, vol. 1: Long Papers, pp. 875–884 (2017)
19. Parkay, F.W., Hass, G.: Curriculum Planning: A Contemporary Approach. Allyn & Bacon Incorporated, Boston (1999)
20. Romero, C., Ventura, S.: Educational data mining: a survey from 1995 to 2005. Expert Syst. Appl. **33**, 135–146 (2007)
21. Roy, S., Madhyastha, M., Lawrence, S., Rajan, V.: Inferring concept prerequisite relations from online educational resources. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 9589–9594 (2019)
22. Sayyadiharikandeh, M., Gordon, J., Ambite, J.L., Lerman, K.: Finding prerequisite relations using the wikipedia clickstream. In: Companion Proceedings of the 2019 World Wide Web Conference, pp. 1240–1247 (2019)
23. Scheines, R., Silver, E., Goldin, I.M.: Discovering prerequisite relationships among knowledge components. In: EDM, pp. 355–356 (2014)

24. Talukdar, P., Cohen, W.: Crowdsourced comprehension: predicting prerequisite structure in wikipedia. In: Proceedings of the Seventh Workshop on Building Educational Applications Using NLP, pp. 307–315 (2012)
25. Vuong, A., Nixon, T., Towle, B.: A method for finding prerequisites within a curriculum. In: EDM (2011)
26. Yu, J., et al.: Course concept expansion in MOOCs with external knowledge and interactive game. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 4292–4302 (2019)
27. Zhu, Y., Zamani, H.: Predicting prerequisite relations for unseen concepts. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pp. 8542–8548 (2022)