LECDPR:LLM Enhancement and Concept-Document Interactive Modeling for Prerequisite Relation Prediction

Kui Xiao[1,2,4] , **Lele Zheng**[1,4] , **Xiaoxue He**[1,4] , **Miao Zhang**[1,2,3*] , **Zhifang Huang**[1,4] , **Yan Zhang**[1,2,4]
[1]School of Computer Science, Hubei University, Hubei 430062, China
[2]Key Laboratory of Intelligent Sensing System and Security, Ministry of Education, Hubei 430062, China
[3]Hubei Provincial Key Laboratory of Artificial Intelligence and Smart Learning, Central China Normal University, Hubei 430079, China
[4]Hubei Key Laboratory of Big Data Intelligent Analysis and Application, China
xiaokui@hubu.edu.cn, {zhenglele, hexiaoxue}@stu.hubu.edu.cn, {zhangmiao, 20160006, zhangyan}@hubu.edu.cn

Abstract

Accurate prediction of prerequisite relations among concepts is important for course planning and intelligent tutoring systems. Existing text-based methods are frequently contaminated by noise such as redundant phrasing, ambiguous sentences, and domain-specific colloquialisms. Moreover, previous methods based on graph structures may have failed to capture the complex interactions between concepts and documents. Therefore, we propose an **L**LM **E**nhancement and **C**oncept-**D**ocument Interactive Modeling for **P**rerequisite **R**elation Prediction (LECDPR) model. Concretely, LECDPR elicits an LLM via two-stage prompting to distill concept descriptions and contextual evidence from documents, resulting in semantic space embeddings. Then, the complex interactive relations between concepts and documents are modeled to produce relational space embeddings, aligns two embeddings through random masking and contrastive learning, and adaptively fuses the embeddings. Finally, the fused concept representations are used for more accurate prerequisite relation prediction. Extensive experiments on the public UCD, LectureBank, and MOOC datasets show that LECDPR outperforms ten baselines on ACC, AUC, and F1. Our code is available at https://github.com/lecdpr/LECDPR.

1 Introduction

In the fields of education and knowledge engineering [Agrawal *et al.*, 2016; Allègre *et al.*, 2023], how to systematically organize the concepts and plan reasonable learning paths [Xianghan Christine and Stern, 2022; Yu *et al.*, 2024] is one of the core issues for improving learning effectiveness. Concept prerequisite relation learning [Bai *et al.*, 2025], as a key natural language processing task, aims to automatically

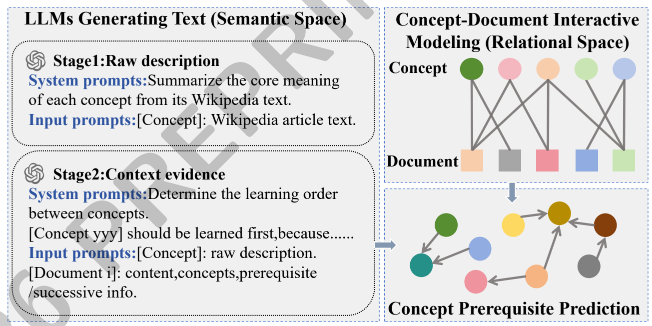

Figure 1: Combination of the semantic space and relational space.

identify the dependency relations between concepts, and determine whether mastering one concept is a prerequisite for learning another. For example, understanding "addition" is a prerequisite for learning "multiplication", which is of vital importance for building adaptive learning systems [Pecháč *et al.*, 2024], designing intelligent course outlines [Wang and Liu, 2016], and implementing personalized learning path recommendations [Dong *et al.*, 2024].

Currently, the mainstream methods for concept prerequisite relation learning mainly develop along two technical routes. The first methods uses the relations between concepts and documents to build a knowledge network [Jia *et al.*, 2021], employing GNN methods to learn the features of concept pairs [Sun *et al.*, 2022; Manrique *et al.*, 2018]. This approach leverages the powerful neighborhood information aggregation and propagation capabilities of graph neural networks, for example [Mazumder *et al.*, 2023; Sun *et al.*, 2024; Zhang *et al.*, 2025a]. The second is based on text information, which directly delves into the text to extract clues about concept relations. This method focuses on extracting descriptive evidence related to concepts from text content (such as textbooks, Wikipedia) and using it as the basis for judging the relations between concepts [Chen *et al.*, 2018]. For example, by analyzing the definitions or contextual information of concepts, identifying indicative phrases like "is a prerequisite", "needs to be mastered first", or through se-

*Corresponding author.

mantic analysis to assess the hierarchical and dependency strength between concepts, for example [Zhang *et al.*, 2024; Wang *et al.*, 2024a].

However, existing research still faces two significant challenges. Traditional text extraction method is often corrupted by redundant wording, ambiguous sentences, and other noise, causing downstream models to deviate. Therefore, as shown in Figure 1, we employs LLMs for a "second-stage" refinement: via prompt-based learning, the models map raw text into clean, structured latent representations and then distill key concepts, or confidence scores, markedly suppressing noise and enhancing signal purity. Moreover, due to inherently relations between documents and concepts, these interactions may respect the learning order of both documents and concepts. Modeling such constrained interactions is a complex task. Current graph-based methods may have failed to capture complex interactions, and graph-based methods and text-based methods are often studied in isolation, failing to effectively combine their advantages, which limits further improvements in downstream task prediction performance.

To address these challenges, we propose a **L**LM **E**nhancement and **C**oncept-**D**ocument Interactive Modeling for **P**rerequisite **R**elation Prediction (LECDPR) model. Our idea involves three steps: Firstly, innovatively employing a two-stage LLM prompt to generate high-quality semantic descriptions of concepts from documents, which are then encoded into semantic space embeddings; Secondly, the complex interactive relations between concepts and documents are modeled through three propagation mechanisms to generate realtional embeddings, implementing dual-space alignment and achieving fusion of two concept embedding spaces via adaptive fusion; Finally, we use the fused embeddings to predict prerequisite relations more accurately.

The main contributions of this study are as follows:

- We introduce LLMs into the concept prerequisite prediction task, using a two-stage prompt to generate high-quality concept texts, and then employing a word vector generator to obtain semantic embeddings of concepts, significantly improving the quality of concept semantic representations.
- We model complex interaction relations between concepts and documents to obtain relational space embeddings, then achieve dual-space alignment and dynamically fuse semantic and relational information.
- We compare our method with existing prerequisite relation prediction models on three public datasets, and the experimental results verify the effectiveness and superiority of our model.

2 Related Work

The core goal of concept prerequisite relation learning is to automatically identify "learning before/learning after" dependencies between concepts from educational resources [Chanaa and Faddouli, 2024], providing a foundation for personalized learning paths and knowledge graph construction [Wang *et al.*, 2021]. Current research on prerequisite relations mainly follows two technical routes: graph-based methods and text-based methods.

2.1 Graph-based Concept Prerequisite Relation Prediction

Graph-based methods treat concepts as nodes and relations as directed edges, using Graph Neural Networks (GNNs) to iteratively propagate neighbor information in the topological space, thereby learning concept embeddings and predicting prerequisite relations between concepts. For example, HGAPNet [Mazumder *et al.*, 2023] constructs a "document-concept" heterogeneous graph, explicitly characterizing different semantic associations with three types of weighted edges. LCPRE [Sun *et al.*, 2024] uses TF-IDF to obtain sparse RC-graphs, employing multi-heads self-attention to aggregate single-path semantics, then fusing multi-path signals through inter-path attention. DGCPL [Zhang *et al.*, 2025a] combines the "concept-resource hypergraph" and "learning behavior directed graph" into a dual-view, using gated knowledge distillation to fuse node representations from both graphs for prerequisite prediction. GKROM [Zhang *et al.*, 2025b] optimizes three types of edges (concept-concept, document-concept, and document-document) on heterogeneous graphs, training multi-task GNNs to improve the accuracy of prerequisite prediction.

2.2 Text-based Concept Prerequisite Relation Prediction

Prerequisite relation prediction methods based on text convert relation judgment into text matching or entailment reasoning tasks, extracting concept co-occurrences, contextual statements, or teaching hints from free texts such as course documents and Wikipedia entries, then encoding semantic evidence with pre-trained language models or word vector encoders. For example, EBCPL [Zhang *et al.*, 2024] classifies documents by topic, then uses three heuristic rules to extract evidence sentences that best illustrate the order of concepts, followed by encoding the paths with BiLSTM, and taking the maximum prediction probability across all evidence for each concept pair as the final output. MTN [Wang *et al.*, 2024b] encodes three features ("resource view", "relation view", and "text view") with three parallel transformers, and fuses these features through an attention mechanism to achieve end-to-end text-driven precedence relation prediction.

We integrate text and graph-structured approaches: LLMs [Becerra *et al.*, 2024] generate high-quality concept descriptions to neutralize the noise plaguing traditional text-based methods, while structured relational knowledge and deep semantic information are jointly fused to bridge the information gap between the two paradigms, yielding markedly more accurate concept prerequisite predictions.

3 Problem Definition

In this section, we introduce the key terminology and research definitions. Let $C = \{c_1, c_2, \cdots, c_n\}$ denote the set of knowledge concepts, where a knowledge concept refers to a core knowledge unit in a specific domain. $D = \{d_1, d_2, \cdots, d_m\}$ represents the set of learning documents, with each document containing one or more concepts.

Our goal is to predict binary labels $y_{ij} \in \{0, 1\}$ for any ordered pair of concepts (c_i, c_j) :

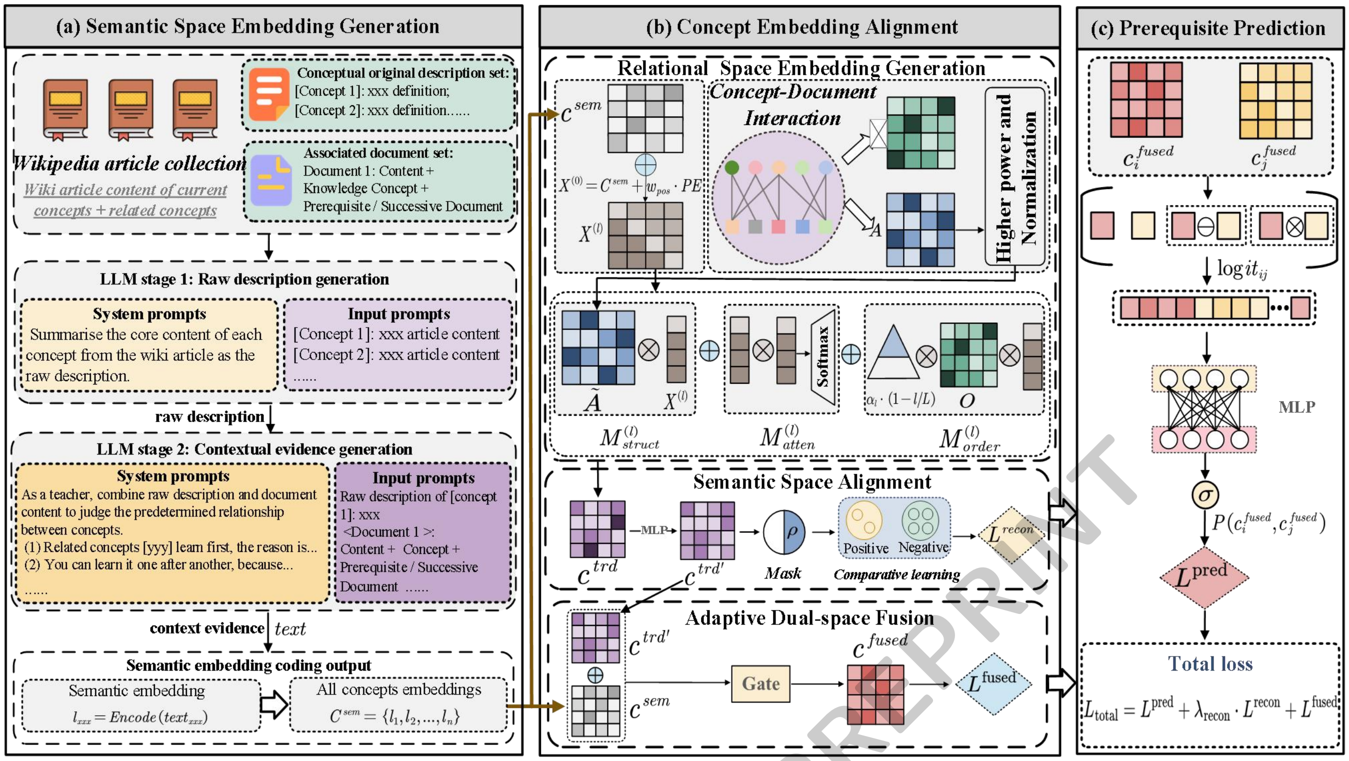

Figure 2: Overall architecture of LECDPR.

$$y_{ij} = \begin{cases} 1, & c_i \text{ is a prerequisite for } c_j, \\ 0, & \text{otherwise, } c_j \text{ is a prerequisite for } c_i. \end{cases}$$

The model needs to learn a mapping from $f : C \times C \to [0, 1]$ so that $f(c_i, c_j) = y_{ij}$, providing a reliable basis for personalized learning path planning.

4 Methodology

Our proposed method LECDPR includes the following three modules(Semantic Space Embedding Generation, Concept Embedding Alignment, and Concept Prerequisite Relation Prediction). Firstly, we leverage LLMs to produce concept embeddings that serve as the semantic space representations. Secondly, built upon these semantic embeddings, we model the intricate concept–document interactions and employ graph neural networks to generate relational space concept embeddings. We then align the two space embeddings via masking and contrastive learning, and adaptively fuse the two kinds of embeddings into unified representations. Finally, the fused concept embeddings are used to predict prerequisite relations among concepts. Overall architecture of LECDPR is as shown in Figure 2.

4.1 Semantic Space Embedding Generation

If two concepts have a prerequisite relation, they are likely to appear in the same document. Therefore, we first attempt to extract useful concept pair features from the document content.

For a concept, we focus on analyzing related documents of this concept. A document may contain one or more primary knowledge concepts. We consider all primary knowledge concepts in the related documents as related concepts of the current concept. Next, we generate effective text content for the current concept through a two-stage prompt learning with LLMs [Yang *et al.*, 2023], thereby creating the semantic space embedding of the concept.

In the first stage, our goal is to use LLM prompt learning to summarize the raw description content of the current concept and all its related concepts from the corresponding Wikipedia article [Talukdar and Cohen, 2012]. These descriptions may include a definition of the concept and other important content. The LLM prompt consists of two parts: system prompt and input prompt [Wei *et al.*, 2023]. The system prompt provides guiding instructions on what content the LLM needs to generate, and the input prompt is the raw data to be input into the LLM. First-stage prompt for concept description denaration is as shown in Figure 3 below:

System prompts:
Summarise the core content of each concept from the wiki article as the original description.
Input prompts:
[Concept 1]: xxx article content ;
[Concept 2]: xxx article content ;
......

Figure 3: First-stage prompt for concept description generation.

Both parts of the prompts should be input into the LLM simultaneously. If a concept does not have a corresponding Wikipedia article, we will select an article with content closest to the concept as a substitute.

In the second stage, our goal is to obtain the context evidence of the current concept. Similarly, the LLM prompt includes two parts: system prompt and input prompt. Second-stage prompt for contextual evidence extraction is as shown in Figure 4 below:

System prompts:
As a teacher, combine raw description and document content to judge the predetermined relationship between concepts.
(1) Related concepts [yyy] Learn first, the reason is...
(2) You can learn it one after another, because...
(3) Uncertain
....
Input prompts:
Raw description of [concept 1]: xxx;
<Document 1 >:
Content + Knowledge Concept + Prerequisite / Successive Document;
......

Figure 4: Second-stage prompt for context evidence extraction.

After two stage prompts, the LLM will further generate prerequisite relation text related to the current concept, labeled as text_{xxx} .

Then we encode this content using a word vector encoder. Specifically, we obtain the word vectors for all words in the text and perform average pooling to generate the semantic space embedding of the current concept $\mathbf{c}_i^{\text{sem}} \in R^{300}$:

$$\mathbf{c}_i^{\text{sem}} = \text{Encode}\left(\text{text}_{xxx}\right). \tag{1}$$

Using the same method, we can obtain the set of semantic space embeddings for all concepts, denoted as $\mathbf{C}^{sem} \in R^{N\times 300}$, N is the concept set size.

4.2 Concept Embedding Alignment

Relational Space Embedding Generation

Based on the above semantic space embeddings, we further use graph neural network methods to generate the relational space embeddings for each concept, denoted as $\mathbf{C}^{trd}$.

If two concepts appear in the same document, we define these two concepts as related concepts. Based on whether there is a relation between concepts, we construct an adjacency matrix $\mathbf{A} \in \{0,1\}^{N\times N}$; if other concepts exist in the prerequisite or successor documents of a concept's document, we consider the two concepts having an order, and based on the sequence order between concepts, we construct an order matrix $\mathbf{O} \in \{0,1\}^{N\times N}$. To capture higher-order neighbor information, we compute higher powers of the adjacency matrix and perform weighted aggregation [Tang *et al.*, 2020].

$$\mathbf{A}^{\text{multi}} = \mathbf{A} + \mathbf{w}_k \cdot I\left(\mathbf{A}^k > 0\right), \tag{2}$$

where $I\left(\mathbf{A}^k > 0\right)$ is an indicator function (value is 1 if the condition is true, 0 otherwise); $\mathbf{w}$ is the higher-order adjacency weight, k is the order, $\mathbf{A}^{\text{multi}} \in R^{N\times N}$.

Then perform normalization:

$$\widetilde{\mathbf{A}} = \mathbf{D}^{-1/2} \cdot \mathbf{A}^{\text{multi}} \cdot \mathbf{D}^{-1/2}, \tag{3}$$

where $\mathbf{D}$ is the degree matrix of the adjacency matrix A.

We initialize the concept features as semantic space embeddings $\mathbf{C}^{\text{sem}}$, encode the concept node indices in the adjacency matrix $\mathbf{A}$ with order i as positional information, and add positional encoding on top of the semantic space embeddings:

$$\mathbf{X}^{(0)} = \mathbf{C}^{\text{sem}} + \mathbf{W}_{\text{pos}} \cdot \text{PE}, \tag{4}$$

where $\mathbf{X}^{(0)} \in R^{N\times 300}$, positional encoding $\text{PE}_{i,2j} = \sin\left(i/10000^{2j/d}\right)$, $\text{PE}_{i,2j+1} = \cos\left(i/10000^{2j/d}\right)$.

Due to the diverse interactive relations that exist between concepts and documents, we then model the complex interactions between concepts and documents. Then we perform multi-layer graph convolution propagation [Kipf, 2016; Wei *et al.*, 2020], where each layer aggregates features through three mechanisms: structural propagation, attention propagation, and sequence-aware propagation. (i). If a document contains multiple concepts, the structural propagation mechanism [Wu *et al.*, 2018] is used to propagate features on the graph through a normalized adjacency matrix, allowing each concept to gather information from its neighbors. (ii). If the same concept appears in different documents, the attention propagation mechanism [Zhang *et al.*, 2022a] dynamically adjusts node importance via self-attention, learning which neighbors are more important for the current concept. (iii). If two conceptually prerequisite concepts occur respectively in two documents that bear a dependency relation, the sequence-aware propagation mechanism uses a sequence matrix to capture directional dependencies. The three aggregated features at the lth layer are:

$$\mathbf{M}_{\text{struct}}^{(l)} = \widetilde{\mathbf{A}}\mathbf{X}^{(l)}, \tag{5}$$

$$\mathbf{M}_{\text{atten}}^{(l)} = \text{Softmax}\left(\frac{\mathbf{X}^{(l)}\left(\mathbf{X}^{(l)}\right)^{\top}}{\sqrt{d}}\right)\mathbf{X}^{(l)}, \tag{6}$$

$$\mathbf{M}_{\text{order}}^{(l)} = \alpha_l \cdot (1 - l/L) \cdot \left(\mathbf{O} \cdot \mathbf{X}^{(l)}\right), \tag{7}$$

where α_l is a tunable decay coefficient.

Then the features at the $l+1$th layer can be updated as:

$$\begin{aligned}\mathbf{X}^{(l+1)} = \text{GELU}\Big(&\alpha_{\text{struct}} \cdot \mathbf{M}_{\text{struct}}^{(l)} + \alpha_{\text{atten}} \cdot \mathbf{M}_{\text{atten}}^{(l)} \\ &+\alpha_{\text{order}} \cdot \mathbf{M}_{\text{order}}^{(l)} + \alpha_{\text{res}} \cdot \mathbf{X}^{(l)}\Big).\end{aligned} \tag{8}$$

We define the concept features after L layer propagation as relational space embeddings $\mathbf{C}^{trd} \in R^{N\times 300}$.

Semantic Space Alignment

Now we face a problem: although embeddings in the semantic space and the relational space both represent the same concept, they are generated in completely different ways. The semantic space embeddings come from text encoding, while the relational space embeddings come from graph convolution. Inspired by [He *et al.*, 2022], we align the two space embeddings through random masking and contrastive learning. We train a mapping network to "translate" embeddings from the relational space into a form understandable by the semantic space. Specifically, the mapped embeddings should be able to reconstruct the original semantic embeddings. During this process, a masking operation is introduced to deliberately hide part of the information, forcing the model to learn more fundamental correspondences. This reliable mapping between the two space embeddings is established, laying the foundation for subsequent fusion.

First, we map the relational space embeddings to the semantic space using a MLP:

$$\mathbf{c}_i^{trd'} = \text{MLP}\left(\mathbf{c}_i^{trd}\right). \tag{9}$$

Then, the mapped embeddings are randomly masked proportionally by ρ, and the resulting embedding vectors are only used within the alignment module to compute contrastive learning loss, not as the final output for downstream tasks.

$$\widetilde{\mathbf{c}}_i^{trd'} = \text{Mask}\left(\mathbf{c}_i^{trd'}, \rho\right). \tag{10}$$

Then the contrastive learning optimizes spatial alignment, with the loss defined as:

$$L^{\text{recon}} = -\log \frac{\exp\left(\text{sim}\left(\widetilde{\mathbf{c}}_i^{\text{trd}}, \mathbf{c}_i^{\text{sem}}\right)/\tau\right)}{\sum_{j=1}^{N} \exp\left(\text{sim}\left(\widetilde{\mathbf{c}}_i^{\text{trd}}, \mathbf{c}_j^{\text{sem}}\right)/\tau\right)}, \tag{11}$$

where $\text{sim}(\cdot)$ represents the cosine similarity function, and τ is the temperature hyperparameter.

Adaptive Dual-Space Fusion

After alignment, we obtain two types of concept embeddings: one from the semantic space and one from the relational space. Simply averaging these two embeddings to fuse them presents a problem, because not all concepts require both types of information equally.

To fully utilize the complementary information from both space embeddings, we designed an adaptive fusion module. Adaptive fusion allows the model to decide how to reasonably allocate these two types of information for each concept. Using a gating network [Dhingra *et al.*, 2016], the model observes the two embeddings of a concept and then computes a fusion weight. This design ensures that the final concept representation contains both semantic understanding and relational information, combined in the most suitable way for each specific concept.

For the two types of concept embeddings, the fusion weight is computed via a gating network:

$$w_i^{\text{sem}} = \sigma\left(\mathbf{W}_2 \cdot \tanh \cdot \left(\mathbf{W}_1 \cdot \left(\left[\mathbf{c}_i^{\text{trd}'} \oplus \mathbf{c}_i^{\text{sem}}\right]\right) + \mathbf{b}_1\right) + \mathbf{b}_2\right), \tag{12}$$

where $\oplus$ indicates concatenation operation.

After nonlinear enhancement of the relational space embedding, the fused embedding is:

$$\mathbf{c}_i^{\text{fused}} = w_i^{\text{sem}} \cdot \mathbf{c}_i^{\text{sem}} + (1 - w_i^{\text{sem}}) \cdot \text{GELU}\left(\mathbf{c}_i^{trd'}\right). \tag{13}$$

Using a fusion regularization loss to ensure the model does not go to extremes or completely ignore one space's information. The regularization term formula for L_2 expresses the fusion loss as:

$$L^{\text{fused}} = \lambda \cdot \left(w_i^{\text{sem}} - w_{\text{target}}\right)^2, \tag{14}$$

where w_{target} represents the target weight.

4.3 Concept Prerequisite Relation Prediction

For two concepts $\mathbf{c}_i^{\text{fused}}$ and $\mathbf{c}_j^{\text{fused}}$, we construct multi-dimensional features: difference features $x_{\text{diff}} = \mathbf{c}_i^{\text{fused}} - \mathbf{c}_j^{\text{fused}}$ and product features $x_{\text{hadamard}} = \mathbf{c}_i^{\text{fused}} \odot \mathbf{c}_j^{\text{fused}}$. These features are concatenated and then compute a logit via a multilayer perceptron.

$$\text{logit}_{ij} = \text{MLP}\left(\left[\mathbf{c}_i^{\text{fused}}; \mathbf{c}_j^{\text{fused}}; \mathbf{x}_{\text{diff}}; \mathbf{x}_{\text{hadamard}}\right]\right), \tag{15}$$

where [;] indicates concatenation of vectors, and $\odot$ indicates the Hadamard product.

The final predicted probability of the prerequisite relation between the two concepts is:

$$P\left(\mathbf{c}_i^{\text{fused}}, \mathbf{c}_j^{\text{fused}}\right) = \sigma\left(\text{logit}_{ij}\right), \tag{16}$$

where σ is the Sigmoid function. Finally, we use binary cross-entropy loss to compute the loss, with the concept prerequisite relation prediction loss being:

$$\begin{aligned} L^{\text{pred}} = -\frac{1}{N_C} \sum_{\left(\mathbf{c}_i^{\text{fused}}, \mathbf{c}_j^{\text{fused}}\right) \in \mathcal{C}} \Big[& y_{ij} \log P\left(\mathbf{c}_i^{\text{fused}}, \mathbf{c}_j^{\text{fused}}\right) \\ & + (1 - y_{ij}) \log\left(1 - P\left(\mathbf{c}_i^{\text{fused}}, \mathbf{c}_j^{\text{fused}}\right)\right)\Big]. \end{aligned} \tag{17}$$

where y_{ij} is the true label, which can only be 0 or 1.

The total loss of our model is:

$$L_{\text{total}} = L^{\text{pred}} + \lambda_{\text{recon}} \cdot L^{\text{recon}} + L^{\text{fused}}, \tag{18}$$

where λ_{recon} is the contrastive loss strength coefficient.

5 Experiments

5.1 Experimental Setup

Datasets: To evaluate the effectiveness of our proposed model, we select three public benchmark datasets. The detailed dataset information is shown in Table 1.

University Course Dataset (UCD) [Liang *et al.*, 2017]: This dataset compiles course information from 377 universities in the United States, covering 376 courses in the field of computer science.

LectureBank [Li *et al.*, 2019]: This dataset originates from an online education platform. It includes a variety of course materials, such as lecture notes and video subtitles.

MOOC [Liang *et al.*, 2017]: This dataset is derived from video playlists in a large-scale academic-video corpus. It aggregates video subtitle texts from 38 carefully curated playlists contributed by 386 computer-science departments worldwide.

Dataset	$\lvert C\rvert$	$\lvert D\rvert$	C_{pre}^{+}	C_{pre}^{-}
UCD	407	654	1007	1007
LectureBank	246	277	601	601
MOOC	406	381	1003	1003

Table 1: Datasets statistics. $|C|$ represents the number of concepts, $|D|$ represents the number of resources, C_{pre}^{+} and C_{pre}^{-} denote the number of positive and negative examples, respectively.

Model	LectureBank			MOOC			UCD		
	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1
Qwen2-7B	0.7842	0.8358	0.7906	0.7671	0.8369	0.7689	0.8045	0.8345	0.7986
GPT-4.0 mini(8B)	0.7934	0.8544	0.8062	0.8060	0.8483	0.8040	0.7772	0.8325	0.7644
GAE	0.6877	0.7651	0.6864	0.6721	0.6995	0.6700	0.6642	0.6955	0.6631
VGAE	0.6907	0.7534	0.6898	0.6664	0.7144	0.6656	0.6933	0.7552	0.6927
PREREQ	0.4975	0.5557	0.5130	0.5429	0.6248	0.5746	0.5433	0.6702	0.5866
ConLearn	0.8017	0.8041	0.7104	0.7562	0.8272	0.7156	0.7522	0.8329	0.7500
MHAVGAE	0.7263	0.8213	0.7401	0.7475	0.8759	0.7642	0.7875	0.8645	0.7952
HGAPNet	0.8167	0.8803	0.8136	0.8550	0.9014	0.8397	0.8200	0.8998	0.8043
LCPRE	0.8182	0.8514	0.8359	0.8258	0.8898	0.8223	0.8366	0.8884	0.8216
GKROM	0.8358	0.8866	0.8070	0.8522	0.9040	0.8409	**0.8675**	0.9070	0.8326
Ours	**0.8595**	**0.9230**	**0.8595**	**0.8905**	**0.9497**	**0.8889**	0.8614	**0.9228**	**0.8542**
Improve Rate	2.37%↑	3.64%↑	2.36%↑	3.55%↑	4.57%↑	4.80%↑	0.61%↓	1.58%↑	2.16%↑

Table 2: Performance Comparison. The best performance is highlighted in bold, and the runner-up is underlined. ↑(↓) indicates the improvement (decline) of our model compared to the best baseline.

Baseline Methods. We compared our proposed model with eight deep learning-based concept prerequisite relation prediction methods: GAE [Kipf and Welling, 2016], VGAE [Kipf and Welling, 2016], PREREQ [Roy *et al.*, 2019], ConLearn [Sun *et al.*, 2022], MHAVGAE [Zhang *et al.*, 2022b], HGAPNet [Mazumder *et al.*, 2023], LCPRE [Sun *et al.*, 2024], and GKROM [Zhang *et al.*, 2025b]. Additionally, two LLM-based [Dong *et al.*, 2025] methods for prerequisite relation prediction are also considered: Qwen2-7B, GPT-4.0 mini.

Evaluation Metrics. We use ACC , F1 score , and AUC to evaluate the performance of the model in predicting concept prerequisite relations. ACC refers to the proportion of correctly predicted samples out of the total number of samples. The F1 score is the harmonic mean of precision and recall. AUC refers to the area under the receiver operating characteristic curve.

Implementation Details. We follow GKROM's setup, dividing the dataset into training, validation, and test sets into 8:1:1 ratio. To ensure fair and accurate evaluation, all baseline models are reproduced and compared under the same data split. We use the Adam optimizer to optimize our model, with 50 training epochs. The random seed and weight decay coefficient are set to 42 and 1E-4, respectively. Additionally, to enrich the text descriptions of concepts, we use the LLM Qwen2-7B-Instruct, employing a two-stage prompt learning process to generate effective text content for current concepts, then use the word vector encoder GloVe to generate 300-dimensional semantic embeddings of concepts. All experiments are implemented using the PyTorch framework, and the experimental environment is a Linux server equipped with an RTX 4090 GPU.

5.2 Overall Performance

In Table 2, we present the performance comparison results of all baseline methods on the three datasets. Our model usually outperformed all other baseline methods on three datasets. In summary, we draw the following conclusions: (1). Our model performs better than methods that only use graph structures (GAE, VGAE, MHAVGAE, HGAPNet, etc.), demonstrating the effectiveness of extracting the high-quality concept descriptions and context evidence from raw documents through "two-stage prompting". Rather than asking the LLM to produce a concept definition in one shot, the two-stage prompting progressively distills document context into pedagogically unambiguous and verifiable semantic snippets. (2). LECDPR also outperforms methods that only use text (LCPRE, Qwen2-7B, etc.), this is because we explicitly model the rich concept–document interactions: multiple concepts can appear in one document, one concept can span multiple documents, and concept pairs with prerequisite relations tend to reside in documents that themselves exhibit dependency links, which validates the high-order structural modeling capability of graph neural networks. (3). For the fine-grained educational mining task of concept prerequisite relations, integrating the deep semantic generation ability of large language models, the high-order structural modeling capability of graph neural networks, and the cross-space coordination ability of contrastive alignment mechanisms in a closed-loop manner can truly approach the performance limit.

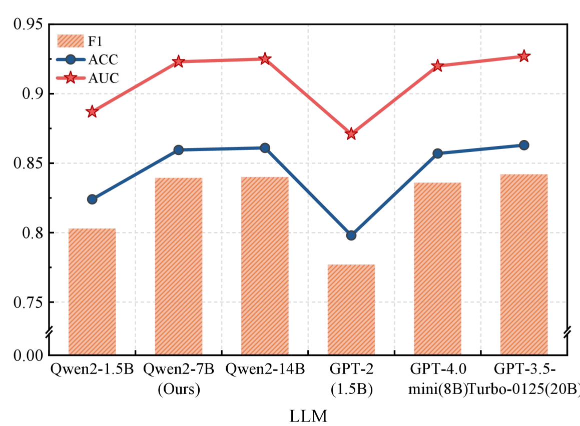

Figure 5: The Impact of LLM Scale on the LectureBank dataset.

5.3 Impact of LLM Scale on Prerequisite Relation Prediction

To verify the cross-model generalizability of our approach, we replaced the text-generation LLM Qwen2-7B with (i) other models from the same family but with different parameter counts and (ii) models of comparable size yet belonging to entirely different architectures. Figure 5 reports the results on LectureBank for both settings. We observe that once the Qwen2 family exceeds 7B parameters the marginal gain vanishes, indicating that the proposed prompting strategy already exhausts the available semantic evidence at a medium scale. Conversely, experiments with Qwen2-1.5B show that LECDPR retains strong performance even when only a small LLM is available, demonstrating the effectiveness of our two-stage prompting across varying model scales. Furthermore, when we substitute the backbone with a same-scale GPT-series model, the outcomes remain stable and competitive, confirming that the two-stage prompt template is robust and transferable across entirely different LLM lineages. Experiments conclusively demonstrate that the two-stage prompting strategy is insensitive to the LLM family, model size, or tokenizer; as long as our two-stage prompting pipeline is followed, high-quality concept descriptions are consistently produced. Consequently, LECDPR maintains robust and state-of-the-art performance across the entire spectrum, achieving genuine cross-model generalization.

5.4 Ablation Experiment

To quantify the individual contributions of each module in LECDPR, we conducted four ablation tests under unified data splits and hyperparameter settings. Table 3 shows the results on UCD. (1) Removing mask alignment (MA). When the contrastive alignment between semantic and relational spaces is disabled, the model degrades to a simple late fusion baseline. This confirms the key role of "forcing the graph encoder to reconstruct semantics generated by the LLM". (2) Removing large language model generation (LLM). Eliminating the two-stage prompt of the LLM causes the semantic space to lose definition and sequence cues, verifying that evidence refined by the LLM is indispensable for fine-grained prerequisite relation discrimination. (3) Removing adaptive gated fusion (AG). Without adaptive gating, the model cannot assign appropriate semantic and structural proportions to different concepts. (4) Removing graph embedding and fusion (GE). If only the LLM semantic embedding is retained, the model reduces to a pure text-matching classifier, with performance dropping, indicating that graph structure and semantic information are complementary.

Method	ACC	AUC	F1
Ours LECDPR	**0.8614**	**0.9228**	**0.8542**
Ours w/o MA	0.8311	0.9070	0.8366
Ours w/o LLM	0.8511	0.9004	0.8389
Ours w/o AG	0.8267	0.8867	0.8205
Ours w/o GE	0.8069	0.8510	0.8098

Table 3: The ablation study of LECDPR.

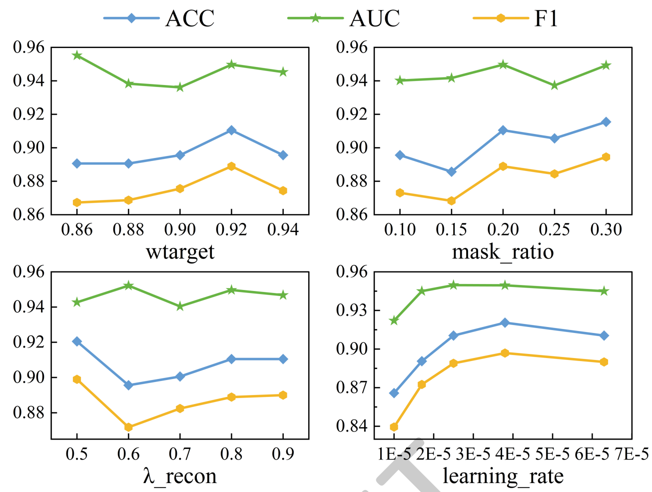

Figure 6: The hyperparameter analysis of LECDPR on the MOOC.

5.5 Hyperparameter Analysis

As shown in Figure 6, we conducted four sets of key hyperparameters for LECDPR on the MOOC dataset. (1) When w_{target} increases from 0.5 to 0.9, performance rises and then drops, peaking at 0.88. A small target weight overemphasizes relational space and suppresses semantic cues, while an overly large weight weakens structural constraints. Thus, $w_{\text{target}} = 0.88$ best balances semantic sufficiency and structural consistency. (2) Raising mask_ratio from 0.15 to 0.25 improves performance, peaking at 0.22. Too few masks push the mapping network toward an identity function and weaken correspondence learning; too many make reconstruction overly difficult, increase gradient noise, and destabilize convergence. (3) As the contrastive loss weight λ_{recon} increases from 0.5 to 1.0, metrics improve; further increasing to 1.2 causes a decline. Properly increasing contrast strength can more tightly inject LLM knowledge into the graph encoder, but too strong a contrast suppresses the main classification loss, causing gradient conflicts. (4) The learning rate lr varies within 1e-5 to 7e-5, with ACC and F1 forming an inverted U shape curve, peaking at 2.8e-5. When less than 1e-5, convergence is slow with high ten-fold average variance; above 5e-5, the loss oscillates.

6 Conclusion

We target prerequisite relation identification and propose LECDPR, an LLM Enhancement and Concept-Document Interactive Modeling for Prerequisite Relation Prediction model. It uses two-stage prompting to extract high-quality concept text from . A complementary representation mechanism of semantic space and relational space is constructed, utilizing random masking and contrastive learning to achieve fine-grained cross-space alignment; simultaneously, an adaptive gating fusion is designed to dynamically allocate semantic and structural weights to different concepts, balancing semantic sufficiency and structural consistency, thereby enabling more accurate prediction of concept prerequisite relations.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant No.62377009, 62407013), the Natural Science Foundation of Hubei Province of China (No.2026AFB746), the Hubei Provincial Key Laboratory of Artificial Intelligence and Smart Learning (No.2025AISL001), and the Key Laboratory of Intelligent Sensing System and Security (No.KLISSS202504).

References

[Agrawal *et al.*, 2016] Rakesh Agrawal, Behzad Golshan, and Evangelos Elias Papalexakis. Toward data-driven design of educational courses: A feasibility study. In *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016, Raleigh, North Carolina, USA, June 29 - July 2, 2016*, page 6. International Educational Data Mining Society (IEDMS), 2016.

[Allègre *et al.*, 2023] Olivier Allègre, Amel Yessad, and Vanda Luengo. Discovering prerequisite relationships between knowledge components from an interpretable learner model. In *Proceedings of the 16th International Conference on Educational Data Mining, EDM 2023, Bengaluru, India, July 11-14, 2023*. International Educational Data Mining Society, 2023.

[Bai *et al.*, 2025] Youheng Bai, Zitao Liu, Teng Guo, Mingliang Hou, and Kui Xiao. Prerequisite relation learning: A survey and outlook. *ACM Comput. Surv.*, 57(11):279:1–279:28, 2025.

[Becerra *et al.*, 2024] Luccas Rojas Becerra, Juan Romero Colmenares, and Rubén Manrique. Enhancing autonomous task execution in social robots with large language models. In *10th International Conference on Automation, Robotics and Applications, ICARA 2024, Athens, Greece, February 22-24, 2024*, pages 40–44. IEEE, 2024.

[Chanaa and Faddouli, 2024] Abdessamad Chanaa and Nour-Eddine El Faddouli. Prerequisites-based course recommendation: recommending learning objects using concept prerequisites and metadata matching. *Smart Learn. Environ.*, 11(1):16, 2024.

[Chen *et al.*, 2018] Weiyu Chen, Andrew Scott Lan, Da Cao, Christopher George Brinton, and Mung Chiang. Behavioral analysis at scale: Learning course prerequisite structures from learner clickstreams. In *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018, Buffalo, NY, USA, July 15-18, 2018*. International Educational Data Mining Society (IEDMS), 2018.

[Dhingra *et al.*, 2016] Bhuwan Dhingra, Hanxiao Liu, William William Cohen, and Ruslan Salakhutdinov. Gated-attention readers for text comprehension. *CoRR*, abs/1606.01549, 2016.

[Dong *et al.*, 2024] Yao Dong, Yuxi Liu, Yongfeng Dong, Yacong Wang, and Min Chen. Multi-knowledge enhanced graph convolution for learning resource recommendation. *Knowledge-Based Systems*, 291:111521, 2024.

[Dong *et al.*, 2025] Zhiang Dong, Jingyuan Chen, and Fei Wu. Knowledge is power: Harnessing large language models for enhanced cognitive diagnosis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 164–172, 2025.

[He *et al.*, 2022] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.

[Jia *et al.*, 2021] Chenghao Jia, Yongliang Shen, Yechun Tang, Lu Sun, and Weiming Lu. 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, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 2036–2047. Association for Computational Linguistics, 2021.

[Kipf and Welling, 2016] Thomas Niels Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.

[Kipf, 2016] Thomas Niels Kipf. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[Li *et al.*, 2019] Irene Li, Alexander R Fabbri, Robert R Tung, and Dragomir R Radev. What should i learn first: Introducing lecturebank for nlp education and prerequisite chain learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6674–6681, 2019.

[Liang *et al.*, 2017] Chen Liang, Jianbo Ye, Zhaohui Wu, Bart Pursel, and C Giles. Recovering concept prerequisite relations from university course dependencies. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[Manrique *et al.*, 2018] Rubén Manrique, Christian Grévisse, Olga Mariño, and Steffen Rothkugel. Knowledge graph-based core concept identification in learning resources. In *Semantic Technology - 8th Joint International Conference, JIST 2018, Awaji, Japan, November 26-28, 2018, Proceedings*, volume 11341 of *Lecture Notes in Computer Science*, pages 36–51. Springer, 2018.

[Mazumder *et al.*, 2023] Debjani Mazumder, Jiaul H Paik, and Anupam Basu. A graph neural network model for concept prerequisite relation extraction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 1787–1796, 2023.

[Pecháč *et al.*, 2024] Matej Pecháč, Michal Chovanec, and Igor Farkaš. Self-supervised network distillation: An effective approach to exploration in sparse reward environments. *Neurocomputing*, 599:128033, 2024.

[Roy *et al.*, 2019] Sudeshna Roy, Meghana Madhyastha, Sheril Lawrence, and Vaibhav Rajan. Inferring concept prerequisite relations from online educational resources. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 9589–9594, 2019.

[Sun *et al.*, 2022] Hao Sun, Yuntao Li, and Yan Zhang. Conlearn: Contextual-knowledge-aware concept prerequisite relation learning with graph neural network. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, pages 118–126. SIAM, 2022.

[Sun *et al.*, 2024] Jingwen Sun, Yu He, Yiyu Xu, Jingwei Sun, and Guangzhong Sun. A learning-path based supervised method for concept prerequisite relations extraction in educational data. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 2168–2177, 2024.

[Talukdar and Cohen, 2012] Partha Pratim Talukdar and William Weston Cohen. Crowdsourced comprehension: Predicting prerequisite structure in wikipedia. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP, BEA@NAACL-HLT 2012, June 7, 2012, Montréal, Canada*, pages 307–315. The Association for Computer Linguistics, 2012.

[Tang *et al.*, 2020] Zeyun Tang, Yongliang Shen, Xinyin Ma, Wei Xu, Jiale Yu, and Weiming Lu. Multi-hop reading comprehension across documents with path-based graph convolutional network. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 3905–3911. ijcai.org, 2020.

[Wang and Liu, 2016] Shuting Wang and Lei Liu. Prerequisite concept maps extraction for automaticassessment. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 519–521, 2016.

[Wang *et al.*, 2021] Chao Wang, Hengshu Zhu, Peng Wang, Chen Zhu, Xi Zhang, Enhong Chen, and Hui Xiong. Personalized and explainable employee training course recommendations: A bayesian variational approach. *ACM Transactions on Information Systems (TOIS)*, 40(4):1–32, 2021.

[Wang *et al.*, 2024a] Zhichun Wang, Yifeng Shao, Boci Peng, Bangui Li, Yun Li, Qianren Wang, and Nijun Li. Multi-view transformer-based network for prerequisite learning in concept graphs. In *International Semantic Web Conference*, pages 67–86. Springer, 2024.

[Wang *et al.*, 2024b] Zhichun Wang, Yifeng Shao, Boci Peng, Bangui Li, Yun Li, Qianren Wang, and Nijun Li. Multi-view transformer-based network for prerequisite learning in concept graphs. In *International Semantic Web Conference*, pages 67–86. Springer, 2024.

[Wei *et al.*, 2020] Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, and Tat-Seng Chua. Graph-refined convolutional network for multimedia recommendation with implicit feedback. In *MM '20: The 28th ACM International Conference on Multimedia, Virtual Event / Seattle, WA, USA, October 12-16, 2020*, pages 3541–3549. ACM, 2020.

[Wei *et al.*, 2023] Jerry William Wei, Jason Wei, Yi Tay, Dustin Tran, Albert Webson, Yifeng Lu, Xinyun Chen, Hanxiao Liu, Da Huang, Denny Zhou, and Tengyu Ma. Larger language models do in-context learning differently. *CoRR*, abs/2303.03846, 2023.

[Wu *et al.*, 2018] Yuexin Wu, Hanxiao Liu, and Yiming Yang. Graph convolutional matrix completion for bipartite edge prediction. In *Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K 2018, Volume 1: KDIR, Seville, Spain, September 18-20, 2018*, pages 49–58. SciTePress, 2018.

[Xianghan Christine and Stern, 2022] O'Dea Xianghan Christine and Julian Stern. Virtually the same?: Online higher education in the post covid-19 era. *British Journal of educational technology*, 53(3):437, 2022.

[Yang *et al.*, 2023] Chengrun Yang, Xuezhi Wang, Yifeng Lu, Hanxiao Liu, Quoc Viet Le, Denny Zhou, and Xinyun Chen. Large language models as optimizers. *CoRR*, abs/2309.03409, 2023.

[Yu *et al.*, 2024] Shenbao Yu, Yifeng Zeng, Fan Yang, and Yinghui Pan. Causal-driven skill prerequisite structure discovery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 20604–20612, 2024.

[Zhang *et al.*, 2022a] Juntao Zhang, Nanzhou Lin, Xuelong Zhang, Wei Song, Xiandi Yang, and Zhiyong Peng. Learning concept prerequisite relations from educational data via multi-head attention variational graph auto-encoders. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 1377–1385, 2022.

[Zhang *et al.*, 2022b] Juntao Zhang, Nanzhou Lin, Xuelong Zhang, Wei Song, Xiandi Yang, and Zhiyong Peng. Learning concept prerequisite relations from educational data via multi-head attention variational graph auto-encoders. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 1377–1385, 2022.

[Zhang *et al.*, 2024] JiaYu Zhang, XiaoYan Zhang, XiaoFeng Du, and TianBo Lu. Ebcpl: A novel evidence-based method for concept prerequisite relation learning. In *International Conference on Database Systems for Advanced Applications*, pages 390–405. Springer, 2024.

[Zhang *et al.*, 2025a] Miao Zhang, Jiawei Wang, Jinying Han, Kui Xiao, Zhifei Li, Yan Zhang, Hao Chen, and Shihui Wang. Dgcpl: Dual graph distillation for concept prerequisite relation learning. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 8366–8374, 2025.

[Zhang *et al.*, 2025b] Miao Zhang, Jiawei Wang, Kui Xiao, Shihui Wang, Yan Zhang, Hao Chen, and Zhifei Li. Learning concept prerequisite relation via global knowledge relation optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 1638–1646, 2025.