

Learning Concept Prerequisite Relation via Global Knowledge Relation Optimization

Miao Zhang^{1,2}, Jiawei Wang^{1,3}, Kui Xiao^{1,3*},
Shihui Wang^{1,3}, Yan Zhang^{1,3*}, Hao Chen^{1,3}, Zhifei Li^{1,2}

¹School of Computer Science and Information Engineering, Hubei University, China

²Hubei Key Laboratory of Big Data Intelligent Analysis and Application, China

³Key Laboratory of Intelligent Sensing System and Security, Ministry of Education, China

zhangmiao@hubu.edu.cn, wangjw@stu.hubu.edu.cn, {xiaokui, wsh, zhangyan, ch, zhifei1993}@hubu.edu.cn

Abstract

Learning concept prerequisite relations helps better master and build a logically coherent knowledge structure. Many studies use graph neural networks to create heterogeneous knowledge networks that enhance concept representations. However, different types of relations in these networks can influence each other. Existing research often focuses solely on concept relations, neglecting other types of knowledge connections. To address this issue, this paper proposes a novel concept prerequisite relation learning model, named the **Global Knowledge Relation Optimization Model (GKROM)**. Specifically, we capture the impact of different knowledge relation types on document and concept semantic representations separately, integrating the document and concept semantic representations. Then, we introduce multi-objective learning to optimize the knowledge relation network from a global perspective. Through the above optimization, GKROM learns richer semantic representations for concepts and documents, improving the accuracy of concept prerequisite relation learning. Extensive experiments on public datasets demonstrate the effectiveness of our GKROM, achieving state-of-the-art performance in concept prerequisite relation learning.

Code — <https://github.com/wisejw/GKROM>

Introduction

In the age of information overload, the rapid pace of knowledge updating demands more from learners. To ensure effective learning and mastery of new knowledge, the fundamental concepts within educational resources must be thoughtfully structured and sequenced based on their interrelationships (Yu et al. 2024). One crucial type of relation is the concept prerequisite, which dictates that the understanding or learning of one concept must come before another related concept. For instance, in mathematics, students need to grasp “*the basic properties of fraction*” before tackling “*fraction operations*”, as these properties serve as the foundational prerequisites. Clearly identifying these prerequisite relations is essential for developing intelligent tutoring systems, and diagnosing students’ knowledge states (Chen et al.

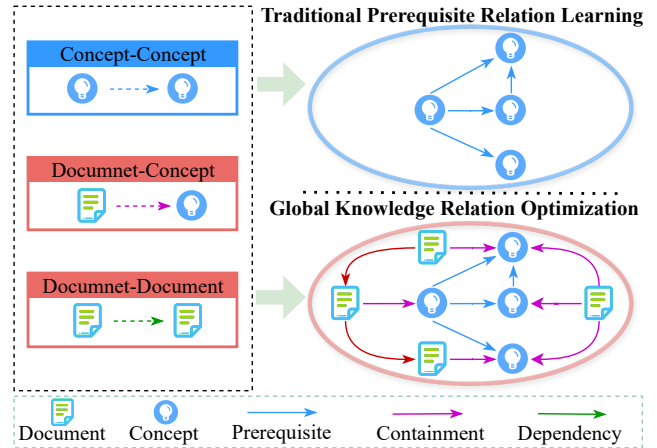


Figure 1: Comparison between our proposed Global Knowledge Relation Optimization and traditional prerequisite relation learning.

2018; Wang et al. 2024; Shen et al. 2024). Therefore, advancing more automated and precise methods for identifying these relations is crucial for enhancing personalized education and improving teaching outcomes.

Over the past decade, researchers have developed various methods to uncover concept prerequisite relations. Initially, these approaches rely on manually designed features and traditional machine learning techniques. With the advent of deep learning, attention shifted toward deep neural networks, which can automatically extract features, eliminating the need for manual design (Sun, Li, and Zhang 2022). More recently, the emergence of graph neural networks has allowed researchers to explore concept relations through graph-based structures (Zhang et al. 2022). However, current graph neural network-based methods still face challenges, particularly in addressing other types of relations within these heterogeneous networks.

In heterogeneous knowledge networks, documents and concepts are the two primary node types, both of which play critical roles (Jia et al. 2021). Similarly, the various types of relations—whether between concepts, between documents and concepts, or between documents—are all equally im-

*Corresponding Author.

portant. These relations are deeply interconnected and influence one another. For instance, dependencies concept prerequisite relations often stem from underlying concept prerequisite relations (Mazumder, Paik, and Basu 2023). However, as shown in Figure 1, traditional models for concept prerequisite learning, such as R-VGAE (Li et al. 2020) and MHAVGAE (Zhang et al. 2022), typically focus only on using neighboring nodes in heterogeneous knowledge networks to enhance the semantic representation of current concept node and predict concept relations. They often overlook the prediction about other types of knowledge relations within the networks. Yet, there is a strong connection between concept relations and other knowledge relations in heterogeneous networks. An effective model should not only learn concept relations but also optimize the learning about all types of knowledge relations from a global perspective.

We introduce a novel model for learning concept prerequisite relations, named the **Global Knowledge Relation Optimization Model (GKROM)**. GKROM optimizes the knowledge relation network by incorporating two auxiliary learning tasks: one for identifying containment relations between documents and concepts, and another for analyzing dependency relations between documents. These auxiliary tasks are designed to improve the accuracy of learning concept prerequisite relations. The key advantage of GKROM lies in its global optimization approach. By transforming traditional single-task learning into a multi-objective optimization problem, GKROM encourages the learning of richer semantic representations for both concepts and documents through the auxiliary tasks. To the best of our knowledge, GKROM is the first model to approach concept prerequisite relation learning from the perspective of global optimization within heterogeneous knowledge networks.

The contributions of this paper are as follows:

- We propose the GKROM model based on auxiliary tasks and global optimization strategy, which considers the impact of different types of knowledge relations on documents and concepts, thereby improving the accuracy of concept prerequisite relation learning.
- We use GKROM to obtain optimal embeddings of documents and concepts in a heterogeneous knowledge network by balancing losses across different tasks. This allows it to effectively predict document-concept containment relations and document-document dependencies.
- Extensive experiments on three public datasets validate the effectiveness of our proposed model, achieving state-of-the-art performance.

Related Work

Concept Prerequisite Relation Learning

The pioneering work by Talukdar and Cohen (Talukdar and Cohen 2012) marked a significant milestone in the study of concept prerequisite relations. Building on their foundation, the field progressed with the introduction of the hyperlink-based metric RefD (Liang et al. 2015), which effectively quantified reference distance between concepts. The PREREQ model (Roy et al. 2019) further advanced

the field by employing pairwise latent Dirichlet allocation and a Siamese network to enhance relation predictions. In the domain of graph networks for concept prerequisite prediction, the Variational Graph Auto-Encoder (VGAE) (Kipf and Welling 2016) was developed to capture complex connections between concepts and educational documents. This approach was extended by the Relational Variational Graph Auto-Encoder (R-VGAE), which predicted relations in graphs containing both concept and document nodes. The MHAVGAE model improved accuracy in identifying concept prerequisite relations by using multi-head attention and variational graph autoencoder techniques. Additionally, the HGAPNet model (Mazumder, Paik, and Basu 2023) leveraged node attention mechanisms on heterogeneous graphs to uncover prerequisite relations.

Existing studies often overlook other types of relations within heterogeneous knowledge networks. Effectively leveraging these additional relations is key to better capturing the semantic representations of concepts and documents.

Graph Neural Networks

Graph Neural Networks (GNNs) (Zhou et al. 2020; Mikolov et al. 2013) have become essential tools for analyzing graph-structured data, particularly in fields like social networks and knowledge bases (Zhang et al. 2023; Kipf and Welling 2017). GNNs effectively model complex relations by learning rich semantic representations of nodes, which is particularly valuable for understanding concept prerequisite relations in educational contexts. Notable advancements in this field include the Graph Auto-Encoder (GAE) (Kipf and Welling 2016) for unsupervised learning of graph node embeddings, and the Gated Graph Neural Network (GGNN) (Li et al. 2016), which utilizes gated recurrent units to effectively capture sequential dependencies in graph data. The Graph Attention Network (GAT) (Zhang et al. 2018) introduced attention mechanisms that assign varying importance to neighboring nodes, thereby enhancing the model’s ability to learn meaningful representations. However, these methods often treat specific types of relations in isolation, limiting their capacity to capture the broader, interconnected nature of knowledge networks.

To overcome these limitations, we extend GNNs into a global knowledge relation optimization network. Our model simultaneously captures and optimizes concept-concept, document-concept, and document-document relations. This comprehensive approach not only deepens the understanding of knowledge structures but also improves the accuracy and robustness of concept prerequisite relation learning.

Preliminaries

This paper aims to explore and address a key problem in the field of education: how to effectively learn and recognize concept prerequisite relations between concepts from a globally optimized perspective. In addition, we introduce two auxiliary tasks to improve the generalization ability and accuracy of the learning model. We first have a set of knowledge concepts $C = \{c_1, c_2, \dots, c_n\}$ and a set of documents $D = \{d_1, d_2, \dots, d_m\}$, where n is the number of knowledge

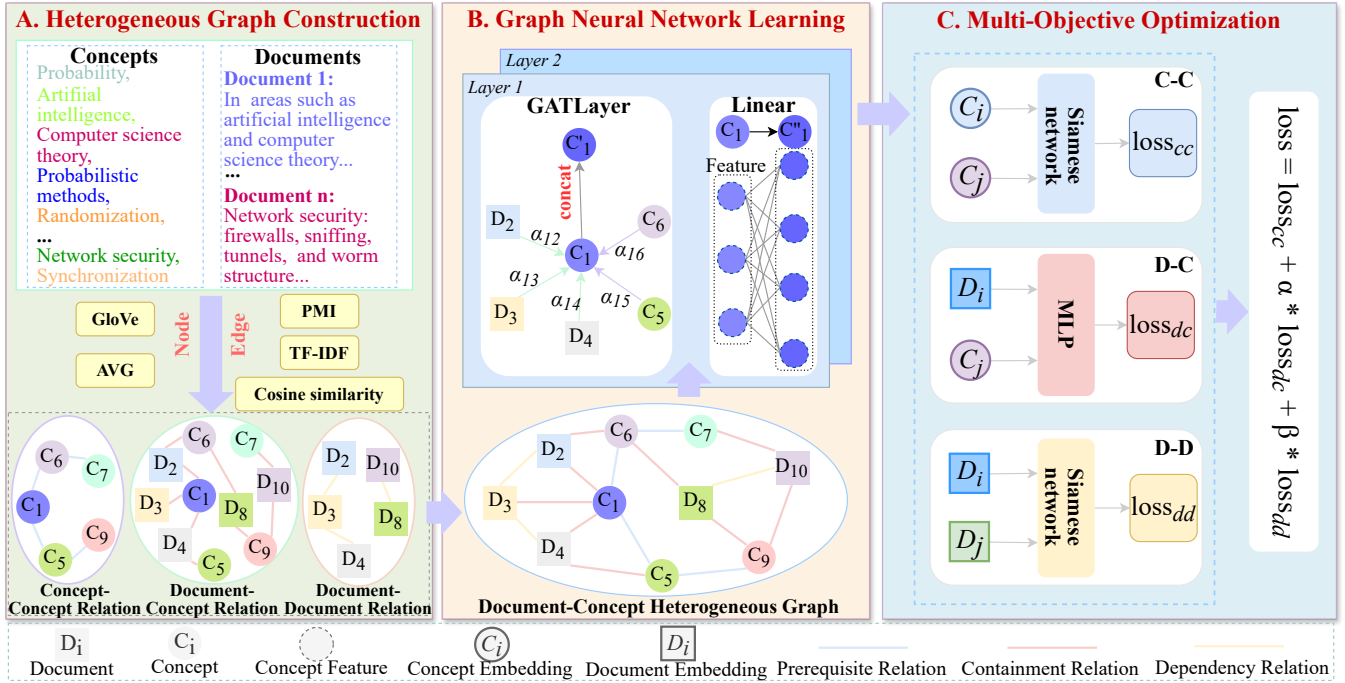


Figure 2: The overall structure of the proposed GKROM.

concepts and m is the number of documents. Concepts play an important role in learning a particular document, and each document contains one or more concepts.

Problem Definition: We formalize the concept prerequisite relation learning problem as a binary classification task. Given a pair of concepts (c_i, c_j) , our goal is to predict whether c_i is a prerequisite for c_j . This task can be defined as learning an objective function $F_{\theta_{cc}} : C \times C \rightarrow \{0, 1\}$. Specifically, if $F_{\theta_{cc}}(c_i, c_j) = 1$, then c_i is a prerequisite for c_j ; otherwise, c_i is not a prerequisite for c_j . The other two auxiliary tasks have similar objective functions, namely $F_{\theta_{dc}} : D \times C \rightarrow \{0, 1\}$ and $F_{\theta_{dd}} : D \times D \rightarrow \{0, 1\}$. The former represents the learning task for the document-concept containment relation, and the latter denotes the learning task for the document-document dependency relation.

Methodology

In this paper, we introduce an innovative model designed to deeply learn concept prerequisite relations and their associated knowledge relations through a global relation optimization strategy. GKROM leverages the powerful representation learning capabilities of graph neural networks and incorporates multi-perspective information from auxiliary tasks, resulting in a more comprehensive understanding of knowledge relations. As illustrated in Figure 2, the architecture of GKROM outlines the entire process from heterogeneous graph construction to multi-objective optimization. The model is organized into three main layers: Heterogeneous Graph Construction, Graph Neural Network Learning, and Multi-Objective Optimization.

Heterogeneous Graph Construction

We construct a heterogeneous graph $G = (V, E)$ using a set of documents and their associated knowledge concepts. The node set V includes two types of nodes: concept nodes and document nodes, while the edge set E consists of three types of edges: concept-concept edges, document-concept edges, and document-document edges. It’s important to note that the edges in the heterogeneous graph do not represent the prerequisite, dependency, or containment relations we aim to predict; instead, they represent the semantic correlations between nodes.

Node Embedding. For a single-word concept node, we use the GloVe model (Pennington, Socher, and Manning 2014) to generate 300-dimensional its embedding. For a multiple-words concept node, we average its word embeddings to obtain the final concept node embedding. Similarly, for a document node, we generate a 300-dimensional embedding by averaging the embeddings of all words contained in the document. In this context, a concept node embedding is represented as V_c , while a document node embedding is denoted as V_d .

Edge Generation. In constructing the adjacency matrix of the heterogeneous graph, we employ three distinct methods to determine the presence of edges between nodes. For the “concept-concept” relation (A_{cc}), we measure the similarity between two concepts using the Pointwise Mutual Information (Church and Hanks 1989), and an edge is formed only if the similarity exceeds a predefined threshold. The “document-concept” relation (A_{dc}) is quantified by the Term Frequency Inverse Document Frequency based on local ex-

tremum statistics (Mazumder, Paik, and Basu 2023). The decision to establish an edge between nodes is similarly based on a predefined threshold. The “document-document” relation (\mathbf{A}_{dd}) is determined by calculating the cosine similarity between the vectors of two documents, with an edge being created only if this similarity surpasses a specified threshold.

Graph Neural Network Learning

The primary objective of the graph neural network learning layer is to extract deep feature vectors for all nodes within the heterogeneous graph. We start by constructing the initial feature vectors $\vec{V} = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_N\} \in \mathbb{R}^{F \times N}$ and the adjacency matrix \mathbf{A} based on the entities defined above. Here, $\vec{v}_i \in \mathbb{R}^F$ represents the initial feature vector of a node, N indicates the total number of graph nodes, and F denotes the number of dimensions in each node’s initial feature vector. Next, we apply a two-layer graph attention network to integrate the feature information of neighboring nodes into each current node, producing a new embedding for all graph nodes $\vec{V}' \in \mathbb{R}^{F' \times N}$, where F' represents the number of dimensions in the updated feature vector of each node. The implementation details are described below.

For a given node v_i , we begin by using the self-attention mechanism $f : \mathbb{R}^{F'} \times \mathbb{R}^{F'} \rightarrow \mathbb{R}$ to calculate its attention coefficient with each neighboring node in the graph. For instance, considering a neighboring node v_j , the calculation is performed as follows:

$$e_{ij} = f(\mathbf{W}\vec{v}_i, \mathbf{W}\vec{v}_j), \quad (1)$$

here, $\mathbf{W} \in \mathbb{R}^{F' \times F}$ is the parameter matrix of the linear transformation.

To make the coefficients between different nodes easily comparable, we normalize all choices of node v_j using the softmax function. Note that we combine the adjacency matrix \mathbf{A} of the heterogeneous graph to take into account the graph structure information by performing masked attention to compute only the feature information of the node’s first-order neighboring nodes, while not taking into account the node’s own information.

$$\alpha_{ij} = \text{softmax}_j(e_{ij}). \quad (2)$$

The attention mechanism f described above is a single-layer feed-forward neural network parameterized by the weight vector $\vec{a} \in \mathbb{R}^{2F'}$ and applying the LeakyReLU nonlinear activation function. Overall, the detailed expanded form of the above equation α_{ij} is as follows:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\vec{a}^T [\mathbf{W}\vec{v}_i \parallel \mathbf{W}\vec{v}_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(\vec{a}^T [\mathbf{W}\vec{v}_i \parallel \mathbf{W}\vec{v}_k]))}, \quad (3)$$

where \mathcal{N}_i denotes the set of neighboring nodes of v_j , \cdot^T denotes transpose operation, and \parallel denotes concatenation operation.

The normalized attention coefficients are used to compute a weighted sum of the corresponding feature vectors, which is then nonlinearly transformed by the sigmoid function to generate the local feature representation for each node:

$$\vec{v}'_i = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{W}\vec{v}_j\right), \quad (4)$$

where σ is the sigmoid function. Formally, we define the graph attention network used throughout the architecture as the *GATLayer*.

The initial node embedding \vec{V} is then passed through a fully connected neural network, which considers the information of all the nodes to get the transformed node global feature representation as:

$$\vec{H} = \sigma(\mathbf{W}_v \vec{V} + \mathbf{b}_v), \quad (5)$$

where $\mathbf{W}_v \in \mathbb{R}^{NF' \times NF}$ and $\mathbf{b}_v \in \mathbb{R}^{NF'}$ are the weight matrix and bias vector of the fully connected neural network, respectively. σ represents the sigmoid function. Formally, we define this fully connected neural network, used to capture information from all nodes, as the *Linear*.

Finally, the output of the graph attention operation is combined with the result of the linear transformation and passed through an activation function to produce the first layer’s output $\vec{V}'(l_1)$. In a similar manner, the second layer of the GAT and the linear layer work together to produce a new feature representation $\vec{V}'(l_2)$ that integrates both local neighborhood information and global context.

$$\begin{cases} \vec{V}'(l_1) = \text{ReLU}(\text{GATLayer}(\vec{V}, \mathbf{A}) + \text{Linear}(\vec{V})), \\ \vec{V}'(l_2) = \text{GATLayer}(\vec{V}'(l_1), \mathbf{A}) + \text{Linear}(\vec{V}'(l_1)). \end{cases} \quad (6)$$

Multi-Objective Optimization

The GKROM model is centered around learning multiple knowledge relations from a global knowledge network. To achieve this, we divide the node embedding set $\vec{V}'(l_2)$ into two subsets based on the node types in the heterogeneous graph: the concept node embedding set \vec{V}'_c and the document node embedding set \vec{V}'_d . Each node’s embedding encapsulates both the node’s intrinsic information and the structural information of the network. These embeddings will later be used to predict various knowledge relations.

Concept-Concept Prediction Layer. After obtaining the final concept representation, GKROM uses the Siamese network to predict whether concept c_i is a prerequisite for concept c_j . The corresponding embeddings \vec{v}'_{c_i} and \vec{v}'_{c_j} of the two concepts are fed into two feedforward networks with shared weights, respectively.

$$\vec{c}_i = \text{ReLU}(\mathbf{W}_1 \vec{v}'_{c_i} + \mathbf{b}_1), \quad (7)$$

$$\vec{c}_j = \text{ReLU}(\mathbf{W}_1 \vec{v}'_{c_j} + \mathbf{b}_1), \quad (8)$$

where \vec{c}_i and \vec{c}_j represent the results after being projected by the feedforward network. \mathbf{W}_1 and \mathbf{b}_1 are the parameterized weight matrix and bias vector of the feedforward neural network, respectively. The two outputs are then employed for predicting concept prerequisite relation:

$$p(c_i, c_j) = \sigma(\mathbf{W}_2 [\vec{c}_i \parallel \vec{c}_j] \parallel ((\vec{c}_i - \vec{c}_j) \parallel (\vec{c}_i \odot \vec{c}_j)) + \mathbf{b}_2), \quad (9)$$

where \mathbf{W}_2 and \mathbf{b}_2 are the parameterized weight matrix and bias vector, respectively. σ is the sigmoid function, \odot and $-$ are the element-wise multiplication and subtraction operators, respectively, and \parallel denotes the concatenation of

vectors. Finally, we use the binary cross-entropy function (BCE) to compute the loss:

$$\mathcal{L}_{cc} = \frac{1}{|D_1|} \sum_{(c_i, c_j) \in D_1} \text{BCE}(p(c_i, c_j), y_{c_i c_j}), \quad (10)$$

where D_1 is the training dataset for the prerequisite relation learning task, and $y_{c_i c_j} \in \{0, 1\}$ indicates the ground truth relation of the concept pair.

Document-Concept Prediction Layer. After obtaining the final document and concept representations, we predict the document-concept relation, which involves the containment relation between heterogeneous entities. Instead of using a Siamese network, we employ a multi-layer perceptron (MLP) to predict whether the document d_i contains the concept c_j or not. The corresponding embeddings of the document and the concept, \vec{v}'_{d_i} and \vec{v}'_{c_j} , are fed into two separate feed-forward networks.

$$\vec{d}_i = \text{ReLU}(\mathbf{W}_3 \vec{v}'_{d_i} + \mathbf{b}_3), \quad (11)$$

$$\vec{c}_j = \text{ReLU}(\mathbf{W}_4 \vec{v}'_{c_j} + \mathbf{b}_4), \quad (12)$$

where \mathbf{W}_3 , \mathbf{W}_4 , \mathbf{b}_3 and \mathbf{b}_4 are the parameterized weight matrices and bias vectors of two independent feedforward neural networks. The two outputs are then connected for classification, which can be expressed as follows:

$$p(d_i, c_j) = \sigma(\mathbf{W}_5 [\vec{d}_i || \vec{c}_j] + \mathbf{b}_5), \quad (13)$$

where \mathbf{W}_5 and \mathbf{b}_5 are the parameterized weight matrix and bias vector, respectively. Similarly, we use BCE loss :

$$\mathcal{L}_{dc} = \frac{1}{|D_2|} \sum_{(d_i, c_j) \in D_2} \text{BCE}(p(d_i, c_j), y_{d_i c_j}), \quad (14)$$

where D_2 is the training dataset for the document-concept containment relation learning task, and $y_{d_i c_j} \in \{0, 1\}$ is the ground truth relation between a document and a concept.

Document-Document Prediction Layer. Similar to the Concept-Concept Prediction Layer, a Siamese network with the same structure but different parameters is used. This network predicts whether document d_j depends on document d_i . This process also results in a binary cross-entropy loss:

$$\mathcal{L}_{dd} = \frac{1}{|D_3|} \sum_{(d_i, d_j) \in D_3} \text{BCE}(p(d_i, d_j), y_{d_i d_j}), \quad (15)$$

where D_3 is the training dataset for the document-document dependency relation learning task, and $y_{d_i d_j} \in \{0, 1\}$ is the ground truth relation between the two documents.

Global Loss Function Our goal is to balance and simultaneously optimize three independent prediction tasks: concept prerequisite relation, document-concept containment relation, and document-document dependency relation. To achieve this goal, we define a global loss function that is a weighted sum of the individual task loss functions.

$$\mathcal{L} = \mathcal{L}_{cc} + \alpha \cdot \mathcal{L}_{dc} + \beta \cdot \mathcal{L}_{dd}, \quad (16)$$

where $\alpha \in (0, 1)$ and $\beta \in (0, 1)$ are hyperparameters used to balance losses across tasks. During model training, we use gradient backpropagation to minimize the global loss function. This updates the parameters of all shared and task-specific layers in the model.

Dataset	$ C $	$ D $	$ C_c $	$ D_c $	$ D_d $
UniversityCourse	407	654	4,347	3,076	861
LectureBank	320	277	821	2,572	1,210
MOOC	406	382	4,332	5,052	1,445

Table 1: Datasets statistics.

Experiment

Datasets

In order to fully evaluate the performance of our proposed model, we select three public datasets for a series of experiments. The following is a brief description of these datasets:

- **University Course**¹: This dataset (Liang et al. 2017) compiles information about computer science courses at U.S. universities.
- **LectureBank**²: This dataset (Li et al. 2019) includes five domains, with natural language processing and machine learning being among them.
- **MOOC**³: This dataset (Liang et al. 2017) focuses on online open courses and includes 406 concepts and video texts from 382 computer science courses.

The statistics of the datasets are presented in Table 1. It is important to highlight that we expanded the positive samples of prerequisite concept pairs in these three datasets by applying transitivity (i.e., if concept c_i is a prerequisite of concept c_j , and c_j is a prerequisite of concept c_k , then c_i is also considered a prerequisite of c_k). The number of concepts is represented by $|C|$, while the number of documents is represented by $|D|$. The number of concept prerequisite relations is denoted by $|C_c|$, the number of document-concept containment relations by $|D_c|$, and the number of document-document dependency relations by $|D_d|$. These relations are provided by the dataset and serve as ground truth during the model training process.

Several commonly used metrics are used to evaluate the effectiveness of our approach, including Accuracy (ACC), F1-score (F1), and the Area Under the ROC Curve (AUC).

Baseline Methods

We compare GKROM with nine currently popular concept prerequisite relation prediction models: RefD (Liang et al. 2015), M3 (Miaschi et al. 2019), GAE (Kipf and Welling 2016), VGAE (Kipf and Welling 2016), PREREQ (Roy et al. 2019), R-VGAE(T) (Li et al. 2020), R-VGAE(P) (Li et al. 2020), MHAVGAE (Zhang et al. 2022), and HGAP-Net (Mazumder, Paik, and Basu 2023). Additionally, several traditional classification methods are also considered, namely NB, SVM and RF.

Implementation Details

To ensure the fairness and accuracy of model evaluation, we follow the ratio of 8:1:1 to divide the dataset into training,

¹<https://github.com/suderoy/PREREQ-IAAI-19>

²<https://github.com/Yale-LILY/LectureBank>

³<https://github.com/suderoy/PREREQ-IAAI-19>

Method	University Course			LectureBank			MOOC		
	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
NB	0.6186	0.6554	0.6175	0.5585	0.6683	0.5552	0.5159	0.4113	0.5161
SVM	0.7008	0.7030	0.7008	0.6511	0.6216	0.6519	0.5761	0.3478	0.5800
RF	0.7647	0.7654	0.7648	0.8066	0.7998	0.8070	0.7629	0.7572	0.7632
RefD	<u>0.7620</u>	<u>0.7110</u>	<u>0.7520</u>	<u>0.3900</u>	<u>0.5570</u>	<u>0.4760</u>	<u>0.5870</u>	<u>0.4140</u>	<u>0.4900</u>
M3	0.8250	0.8250	0.8910	0.7180	0.6940	0.7860	0.7910	0.7750	0.8750
GAE	0.6642	0.6631	0.6955	0.6877	0.6864	0.7651	0.6721	0.6700	0.6995
VGAE	0.6933	0.6927	0.7552	0.6907	0.6898	0.7534	0.6664	0.6656	0.7144
PREREQ	0.5433	0.5866	0.6702	0.4975	0.5130	0.5557	0.5429	0.5746	0.6248
R-VGAE(T)	0.6849	0.6618	0.7646	0.6660	0.6345	0.7942	0.5926	0.5435	0.6602
R-VGAE(P)	0.7369	0.7220	0.8325	0.5678	0.4714	0.8107	0.5344	0.4108	0.8730
MHAVGAE	0.7875	0.7952	0.8645	0.7263	0.7401	0.8213	0.7475	0.7642	0.8759
HGAPNet	<u>0.8687</u>	<u>0.8710</u>	0.9456	<u>0.8110</u>	<u>0.8098</u>	<u>0.9017</u>	<u>0.8750</u>	<u>0.8787</u>	<u>0.9388</u>
GKROM (Ours)	0.8698	0.8740	<u>0.9395</u>	0.8232	0.8199	0.9054	0.8796	0.8847	0.9417
Improve rate	0.11% ↑	0.30% ↑	0.61% ↓	1.22% ↑	1.01% ↑	0.37% ↑	0.46% ↑	0.60% ↑	0.29% ↑

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.

Method	LectureBank		
	ACC	F1	AUC
Ours	0.8232	0.8199	0.9054
Ours w/o DC	<u>0.8171</u>	<u>0.8181</u>	<u>0.8993</u>
Ours w/o DD	0.8110	0.8075	0.9017
Ours w/o (DC & DD)	0.8049	0.8025	0.8978

Table 3: The ablation study of GKROM.

validation, and test sets. Additionally, we obtain the same number of negative samples as positive samples by random sampling. The codes of RefD, M3, MHAVGAE, and HGAPNet are provided by their original papers, while the codes of other baseline methods are implemented by ourselves. Our proposed model is trained using the Adam optimizer. The random seed, learning rate, weight decay, and batch size are set to 25, 10^{-6} , 10^{-5} , and 4, respectively. The feature dimensions of the initial embeddings of both concepts and documents are 300. The first layer of the graph neural network contains 128 hidden units, and the second layer contains 512 hidden units. The loss weights α and β in the global loss function are set to 0.5 and 0.5, respectively.

The multi-objective optimization mechanism increases the number of model parameters by about 27%, with an average training time of 5.2 hours, an increase of less than 9%. All models are implemented in PyTorch, and the construction of the graph neural network is accomplished using the PyTorch Geometric library. Experiments are performed on the server with one RTX 4090D 24GB GPU.

Experimental Results

Overall Performance. The experimental results in Table 2 demonstrate that the proposed GKROM model excels across all evaluation metrics. On the three datasets, GKROM

outperforms all baseline models on most metrics, showcasing its strong ability to learn concept prerequisite relations. Specifically, in the University Course dataset, GKROM achieves an F1 of 0.8740, about 0.3% higher than the second-best HGAPNet model. In the LectureBank dataset, GKROM reaches an ACC of 0.8232, surpassing HGAPNet by 1.22%. In the MOOC dataset, GKROM records an AUC of 0.9417, approximately 0.3% higher than HGAPNet. Although GKROM’s AUC is slightly lower than HGAPNet’s in the University Course dataset, this is due to the model’s focus on optimizing overall performance rather than a single metric. The global knowledge relation optimization strategy allows GKROM to perform strongly across multiple metrics, resulting in better overall outcomes. These results indicate that GKROM not only achieves top performance across various metrics but also demonstrates stability and robustness across different datasets.

The comparison results lead to the following observations: (1) GKROM outperforms baseline models on multiple metrics, indicating that the global knowledge relation optimization strategy significantly enhances the learning of concept prerequisite relations. (2) GKROM shows particularly strong performance on AUC across different datasets, highlighting its ability to effectively distinguish between positive and negative samples and model complex relations. (3) Compared to baseline models like MHAVGAE and HGAPNet, GKROM demonstrates higher accuracy and generalization in capturing complex relations between concepts and documents, thanks to its global optimization strategy.

Ablation Experiment. The ablation study results in Table 3 show the impact of the two auxiliary tasks on the model’s performance. (1) Removing the Document-Concept containment relation (DC) drops the AUC to 0.8993, highlighting its critical role in enhancing discriminative ability. Accuracy and F1 also decrease slightly, underscoring the importance of the DC task. (2) Removing the Document-

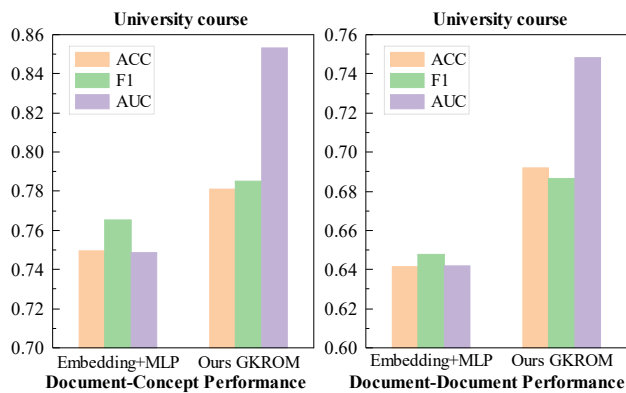


Figure 3: The performance of GKROM on auxiliary tasks, compared with the baseline prediction model.

Document dependency relation (DD) causes a more significant decline, with AUC falling to 0.9017 and F1 to 0.8075, indicating the DD task’s vital role in generalization, particularly in capturing document dependencies in complex relations. (3) Removing both auxiliary tasks (DC and DD) leads to a marked drop in all metrics, with AUC decreasing to 0.8978, showing that their combined influence is essential for maximizing performance.

The study demonstrates that the Document-Document dependency relation and the Document-Concept containment relation can be regarded as implicit features. By optimizing the global knowledge relation network, the model can obtain more accurate concept feature representations, which better support the learning of concept prerequisite relations. This validates the effectiveness of the global knowledge relation optimization strategy.

Performance on Auxiliary Tasks. As shown in Figure 3, we evaluate GKROM’s performance on the DC and DD tasks, comparing it to a prediction model that utilizes only embeddings and an MLP. GKROM shows significant advantages over the comparison model in both tasks.

- **DC Task Performance:** GKROM surpasses the baseline in ACC, F1, and AUC on the MOOC dataset, with a particularly strong performance in AUC, highlighting its superior ability to distinguish between positive and negative samples. This confirms the effectiveness of the global knowledge relation optimization strategy in enhancing DC predictions.
- **DD Task Performance:** GKROM also excels in predicting DD, significantly outperforming the baseline in ACC, F1, and AUC. This demonstrates GKROM’s effectiveness in capturing dependencies between documents and its higher accuracy and stability in managing complex document relations.

Parameter Sensitivity Analysis. As shown in Figure 4, we conducted a comprehensive evaluation of the model’s performance under various hyperparameter settings, using a learning rate of 1E-6, weight decay of 1E-5, and a batch size of 4. The global loss weights α and β are both set to 0.5.

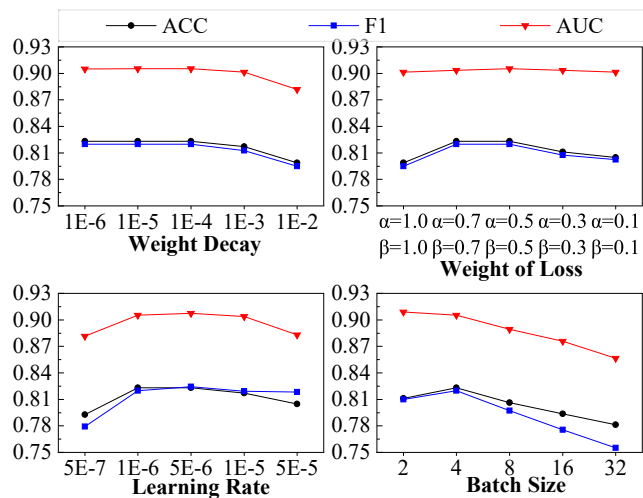


Figure 4: The parameter sensitivity analysis of GKROM.

- **Effect of Weight Decay:** On the University Course dataset, increasing weight decay reduces ACC, F1, and AUC, with a sharp drop at 1E-2, indicating that excessive weight decay harms generalization. The best performance occurs at a weight decay of 1E-5.
- **Effect of Loss Weights:** The model performs best with balanced loss weights ($\alpha = 0.5, \beta = 0.5$), achieving the highest ACC, F1, and AUC scores. This balance is crucial for maximizing overall model performance.
- **Effect of Learning Rate:** The model achieves peak AUC at a learning rate of 1E-6, highlighting that lower learning rates promote stable convergence.
- **Effect of Batch Size:** Performance improves with increasing batch size up to 4, after which it declines. A batch size of 4 optimizes ACC, F1, and AUC, indicating that smaller batches are beneficial, though computationally intensive. Notably, the same batch size is used for both auxiliary and main tasks, with auxiliary task data closely aligned with the current main task batch.

Conclusion

In this paper, we focus on solving the global knowledge optimization problem of multiple relations in knowledge networks. We propose a simple and effective model that combines the powerful representation capabilities of graph neural networks with the advantages of multi-task learning. Our model not only accurately identifies concept prerequisite relations, but also simultaneously predicts document-concept containment relations and document-document dependency relations. After extensive experiments on multiple datasets, our model achieves state-of-the-art performance in concept prerequisite relation learning, surpassing existing techniques and demonstrating its superiority in dealing with complex knowledge network relations.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grants No.62377009, 62407013).

References

- Chen, P.; Lu, Y.; Zheng, V. W.; and Pian, Y. 2018. Prerequisite-Driven Deep Knowledge Tracing. In *IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 17-20, 2018*, 39–48. IEEE Computer Society.
- Church, K. W.; and Hanks, P. 1989. Word Association Norms, Mutual Information and Lexicography. In Hirschberg, J., ed., *27th Annual Meeting of the Association for Computational Linguistics, 26-29 June 1989, University of British Columbia, Vancouver, BC, Canada, Proceedings*, 76–83. ACL.
- Jia, C.; Shen, Y.; Tang, Y.; Sun, L.; and Lu, W. 2021. Heterogeneous Graph Neural Networks for Concept Prerequisite Relation Learning in Educational Data. In Toutanova, K.; Rumshisky, A.; Zettlemoyer, L.; Hakkani-Tür, D.; Beltagy, I.; Bethard, S.; Cotterell, R.; Chakraborty, T.; and Zhou, Y., eds., *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*, 2036–2047. Association for Computational Linguistics.
- Kipf, T. N.; and Welling, M. 2016. Variational Graph Auto-Encoders. *CoRR*, abs/1611.07308.
- Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.
- Li, I.; Fabbri, A. R.; Hingmire, S.; and Radev, D. R. 2020. R-VGAE: Relational-variational Graph Autoencoder for Un-supervised Prerequisite Chain Learning. In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, 1147–1157. International Committee on Computational Linguistics.
- Li, I.; Fabbri, A. R.; Tung, R. R.; and Radev, D. R. 2019. What Should I Learn First: Introducing LectureBank for NLP Education and Prerequisite Chain Learning. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, 6674–6681. AAAI Press.
- Li, Y.; Tarlow, D.; Brockschmidt, M.; and Zemel, R. S. 2016. Gated Graph Sequence Neural Networks. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.
- Liang, C.; Wu, Z.; Huang, W.; and Giles, C. L. 2015. Measuring Prerequisite Relations Among Concepts. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, 1668–1674. The Association for Computational Linguistics.
- Liang, C.; Ye, J.; Wu, Z.; Pursel, B.; and Giles, C. L. 2017. Recovering Concept Prerequisite Relations from University Course Dependencies. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, 4786–4791. AAAI Press.
- Mazumder, D.; Paik, J. H.; and Basu, A. 2023. A Graph Neural Network Model for Concept Prerequisite Relation Extraction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*, 1787–1796. ACM.
- Miaschi, A.; Alzetta, C.; Cardillo, F. A.; and Dell’Orletta, F. 2019. Linguistically-Driven Strategy for Concept Prerequisites Learning on Italian. In Yannakoudakis, H.; Kochmar, E.; Leacock, C.; Madnani, N.; Pilán, I.; and Zesch, T., eds., *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications, BEA@ACL 2019, Florence, Italy, August 2, 2019*, 285–295. Association for Computational Linguistics.
- Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, 3111–3119.
- Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, 1532–1543. ACL.
- Roy, S.; Madhyastha, M.; Lawrence, S.; and Rajan, V. 2019. Inferring Concept Prerequisite Relations from Online Educational Resources. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, 9589–9594. AAAI Press.
- Shen, J.; Qian, H.; Zhang, W.; and Zhou, A. 2024. Symbolic Cognitive Diagnosis via Hybrid Optimization for Intelligent Education Systems. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, February 20-27, 2024, Vancouver, Canada*, 14928–14936. AAAI Press.
- Sun, H.; Li, Y.; and Zhang, Y. 2022. ConLearn: Contextual-knowledge-aware Concept Prerequisite Relation Learning with Graph Neural Network. In *Proceedings of the 2022 SIAM International Conference on Data Mining, SDM 2022, Alexandria, VA, USA, April 28-30, 2022*, 118–126. SIAM.
- Talukdar, P. P.; and Cohen, W. W. 2012. 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*, 307–315. The Association for Computer Linguistics.

- Wang, S.; Zeng, Z.; Yang, X.; Xu, K.; and Zhang, X. 2024. Boosting Neural Cognitive Diagnosis with Student's Affective State Modeling. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, February 20-27, 2024, Vancouver, Canada*, 620–627. AAAI Press.
- Yu, S.; Zeng, Y.; Yang, F.; and Pan, Y. 2024. Causal-Driven Skill Prerequisite Structure Discovery. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, February 20-27, 2024, Vancouver, Canada*, 20604–20612. AAAI Press.
- Zhang, J.; Lin, N.; Zhang, X.; Song, W.; Yang, X.; and Peng, Z. 2022. Learning Concept Prerequisite Relations from Educational Data via Multi-Head Attention Variational Graph Auto-Encoders. In *WSDM '22: The Fifteenth ACM International Conference on Web Search and Data Mining, Virtual Event / Tempe, AZ, USA, February 21 - 25, 2022*, 1377–1385. ACM.
- Zhang, J.; Shi, X.; Xie, J.; Ma, H.; King, I.; and Yeung, D. 2018. GaAN: Gated Attention Networks for Learning on Large and Spatiotemporal Graphs. In *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI 2018, Monterey, California, USA, August 6-10, 2018*, 339–349. AUAI Press.
- Zhang, J.; Wang, S.; Sun, Y.; and Peng, Z. 2023. Prerequisite-driven Fair Clustering on Heterogeneous Information Networks. *Proc. ACM Manag. Data*, 1(2): 122:1–122:27.
- Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; and Sun, M. 2020. Graph neural networks: A review of methods and applications. *AI Open*, 1: 57–81.