ID: 6181

DGCPL: Dual Graph Distillation for Concept Prerequisite Relation Learning

IJCAI 2025 MONTREAL

Miao Zhang[1,2], Jiawei Wang[1,3], Jinying Han[1,3], Kui Xiao[1,3 ✉],
Zhifei Li[1,2], Yan Zhang[1,3], Hao Chen[1,3], Shihui Wang[1,3]

[1]School of Computer Science, Hubei University, China
[2]Hubei Key Laboratory of Big Data Intelligent Analysis and Application, China
[3]Key Laboratory of Intelligent Sensing System and Security, Ministry of Education, China

Code and Appendix

🔥 Introduction

Concept prerequisite relations determine the learning order of knowledge concepts in one domain, which has an important impact on teachers' course design and students' personalized learning. Current research usually predicts concept prerequisite relations from the perspective of knowledge, and rarely pays attention to the role of learners' learning behavior. We propose a **D**ual **G**raph Distillation Method for **C**oncept **P**rerequisite Relation **L**earning (DGCPL). Specifically, DGCPL constructs a dual graph structure from both the knowledge and learning behavior perspectives, and captures the high-order knowledge features and learning behavior features through the concept-resource hypergraph and the learning behavior graph respectively. In addition, we introduce a gated knowledge distillation to fuse the structural information of concept nodes in the two graphs, so as to obtain a more comprehensive concept embedding representation and achieve accurate prediction of prerequisite relations. On three public benchmark datasets, we compare DGCPL with eight graph-based baseline methods and five traditional classification baseline methods. The experimental results show that DGCPL achieves state-of-the-art performance in learning concept prerequisite relations. **Our code is available at https://github.com/wisejw/DGCPL.**
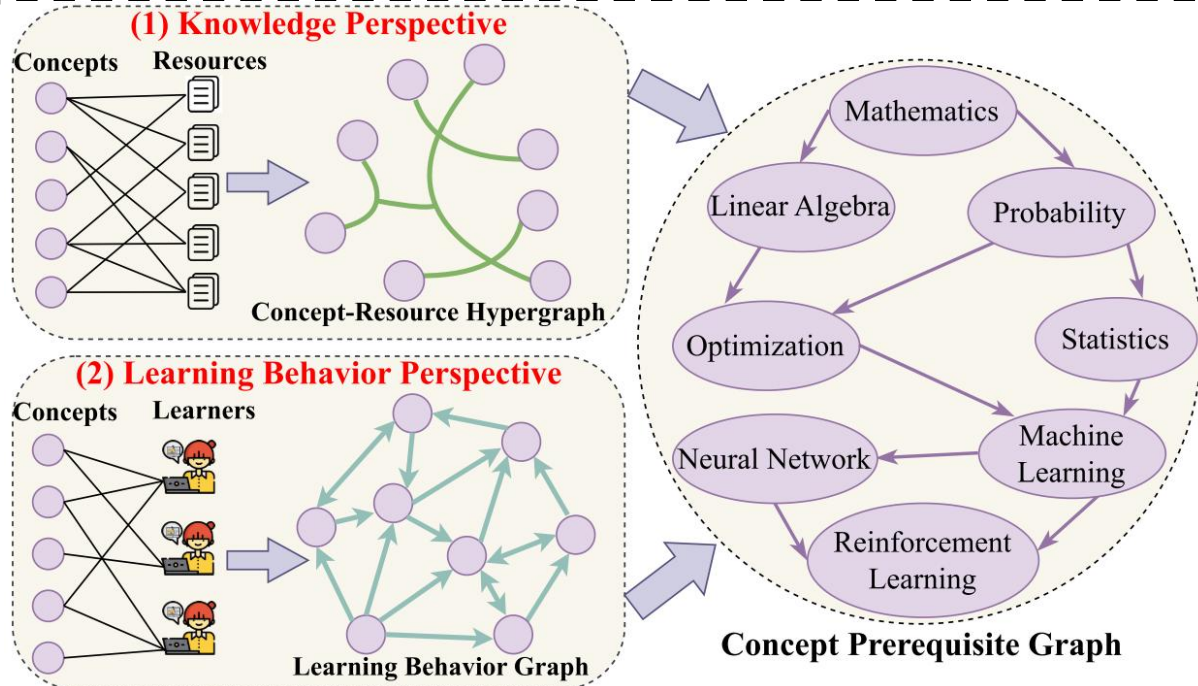

Figure 1: Learning concept prerequisite relations from the perspective of knowledge and the perspective of learning behavior.

👍 Main contributions:

➢ *Dual Graph Structure.*
We propose a **D**ual **G**raph Distillation Method for **C**oncept **P**rerequisite Relation **L**earning (DGCPL), which predicts concept prerequisite relations by constructing a dual graph from the knowledge and learning behavior perspectives.

➢ *Gated Knowledge Distillation.*
We introduce gated knowledge distillation that adaptively integrates higher-order knowledge relations and learning behavior features information in the dual graph through a gating mechanism, resulting in more comprehensive representations of concept embeddings.

➢ *Extensive experiments and SOTA Model.*
On three publicly available benchmark datasets, we compare DGCPL with eight graph-based baseline methods and five traditional classification baseline methods. The experimental results demonstrate that DGCPL achieves state-of-the-art performance.
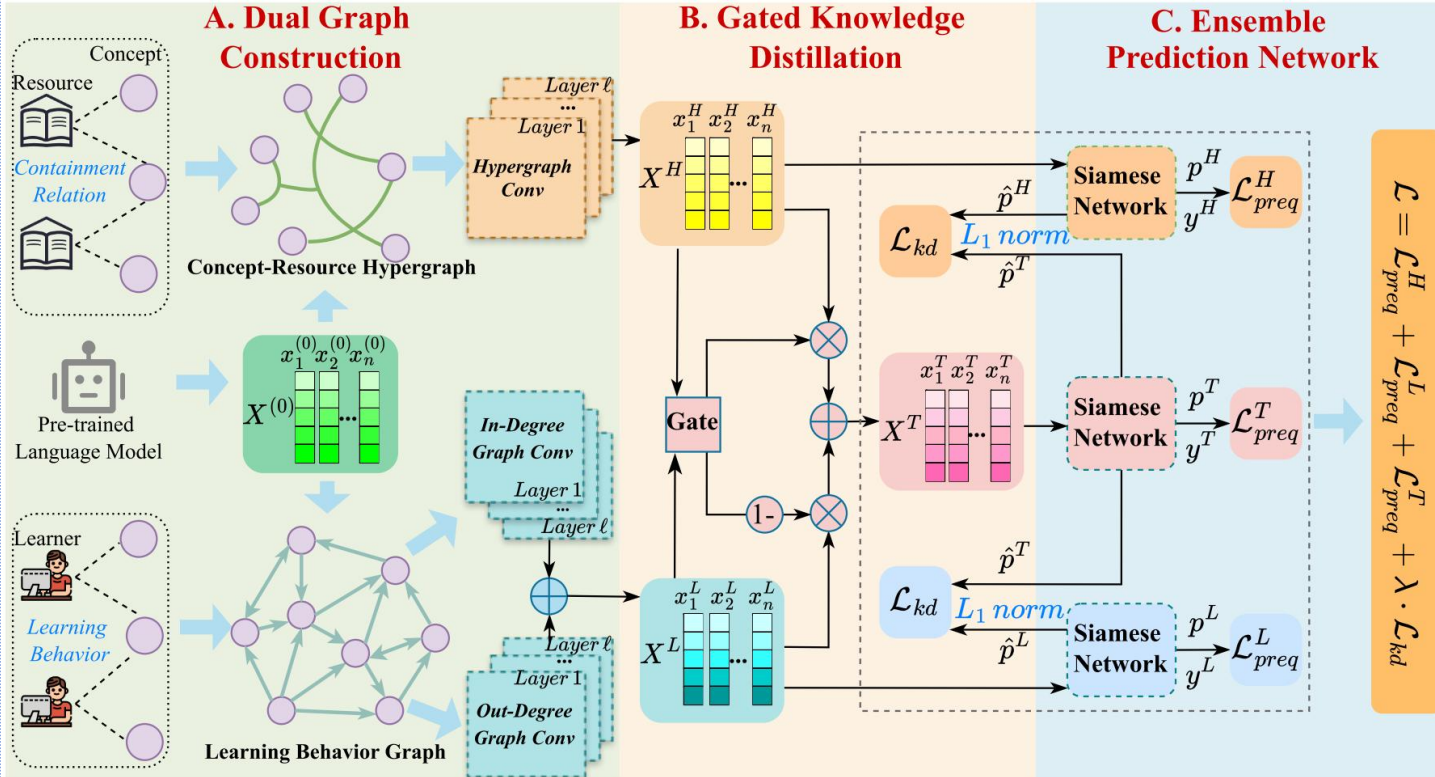
Method

Figure 2: The overall structure of the proposed DGCPL.

A. Dual Graph Construction:

Concept-Resource Hypergraph
We adopt a hypergraph to *capture the high-order knowledge relations between knowledge concepts and learning resources*, thereby learning concept embeddings within the hypergraph.

$$\mathbf{x}_i^{(l+1)} = \phi(\mathbf{D}^{-1/2}\mathbf{HWB}^{-1}\mathbf{H}^T\mathbf{D}^{-1/2}\mathbf{x}_i^{(l)}\Theta^{(l+1)} + \text{Linear}(\mathbf{x}_i^{(l)}))$$

Learning Behavior Graph
We *utilize a directed graph convolutional network to learn concept embeddings* from the learning behavior graph. Nodes can not only propagate information to other nodes but also receive information from other nodes.

$$\mathbf{x}_i^{(l+1),-} = \phi(\sum_{v_j \in N(i)^-} \frac{A_{j,i}^{(in)}}{\sqrt{deg_i^{(in)} \cdot deg_j^{(out)}}}\Theta_N^{(l+1),-}\mathbf{x}_j^{(l)}), \quad \mathbf{x}_i^{(l+1),+} = \phi(\sum_{v_j \in N(i)^+} \frac{A_{i,j}^{(out)}}{\sqrt{deg_i^{(out)} \cdot deg_j^{(in)}}}\Theta_N^{(l+1),+}\mathbf{x}_j^{(l)})$$

$$\mathbf{x}_i^{(l+1)} = \phi(\text{Linear}(\mathbf{x}_i^{(l+1),+}) + \text{Linear}(\mathbf{x}_i^{(l+1),-})).$$

B. Gated Knowledge Distillation:

We treat the concept-resource hypergraph and the learning behavior graph as two independent student models. Subsequently, we regard these two graph modules as *equivalent student models* and employ a *gating mechanism* to achieve their effective integration.

$$\mathbf{x}_i^T = g \odot \mathbf{x}_i^H + (1-g) \odot \mathbf{x}_i^L, \quad g = \sigma(\mathbf{W}_f[\mathbf{x}_i^H, \mathbf{x}_i^L] + \mathbf{b}_f).$$

C. Ensemble Prediction Network:

Through three Siamese networks with the same structure but different parameters, we can *obtain the predicted probabilities of the two student models and the teacher model*. For these three models, we use the binary cross-entropy loss function(BCE) to calculate the *prediction loss*. Then, we calculate the *distillation loss* to encourage each student model to align its predictions with the teacher model's predictions. Finally, the **overall loss function** of the model is defined as:

$$\mathcal{L} = \mathcal{L}_{preq}^H + \mathcal{L}_{preq}^L + \mathcal{L}_{preq}^T + \lambda \cdot \mathcal{L}_{kd}.$$

Acknowledgement

Results

RQ1. Overall Performance

Method	UCD			LectureBank			MOOC		
	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
NB	0.5495	0.5845	0.5369	0.5207	0.5323	0.5262	0.5124	0.5288	0.5357
SVM	0.5743	0.6161	0.5601	0.5455	0.5669	0.5361	0.5821	0.6000	0.5923
RF	0.6683	0.6455	0.7419	0.6777	0.6723	0.8245	0.7363	0.7135	0.8084
RefD	0.7620	0.7110	0.7520	0.3900	0.5570	0.4760	0.5870	0.4140	0.4900
PREREQ	0.5433	0.5866	0.6702	0.4975	0.5130	0.5557	0.5429	0.5746	0.6248
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
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
ConLearn	0.7822	0.7684	0.8529	0.8017	0.7931	0.8541	0.7562	0.7200	0.8472
MHAVGAE	0.7875	0.7952	0.8645	0.7263	0.7401	0.8213	0.7475	0.7642	0.8759
HGAPNet	0.8200	0.8043	0.8998	0.8167	0.8136	**0.8803**	0.8550	0.8497	0.9014
LCPRE	0.8366	0.8216	0.8884	0.8182	0.8000	0.8514	0.8258	0.8223	0.8898
DGCPL (Ours)	**0.8564**	**0.8557**	**0.9053**	**0.8347**	**0.8246**	0.8795	**0.8756**	**0.8718**	**0.9236**
Improve rate	1.98% ↑	3.41% ↑	0.55% ↑	1.65% ↑	1.10% ↑	0.08% ↓	2.06% ↑	2.21% ↑	2.22% ↑

RQ2. Ablation Experiment

Method	UCD	LectureBank	MOOC
Ours DGCPL	**0.8557**	**0.8246**	**0.8718**
Ours w/o GKD	0.8374	0.8226	0.8528
Ours w/o CRHG	0.8177	0.8160	0.8168
Ours w/o LBG	0.8235	0.8130	0.8469

✓ **w/o GKD**: Removing the Gated Knowledge Distillation.
✓ **w/o CRHG**: Removing the Concept Resource Hypergraph.
✓ **w/o LBG**: Removing the Learning Behavior Graph.

RQ3. Quality Analysis

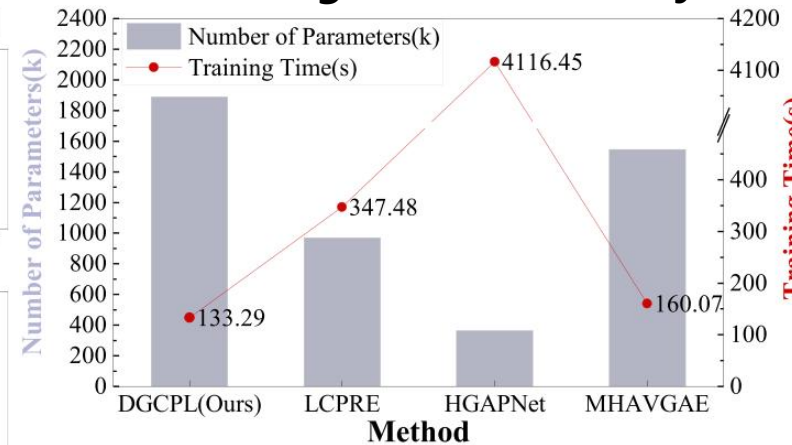| Degree Level	Degree Range
Low Degree	0<Degree<=3
Medium Degree	3<Degree<=6
High Degree	6<Degree
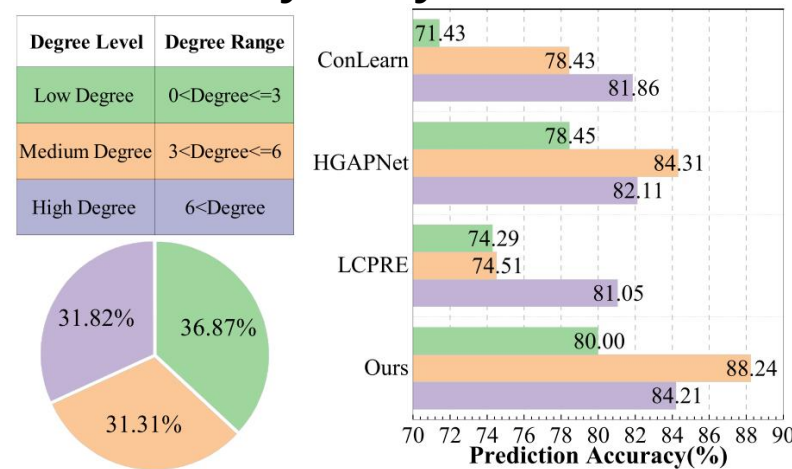
RQ4. Hyperparameter Analysis
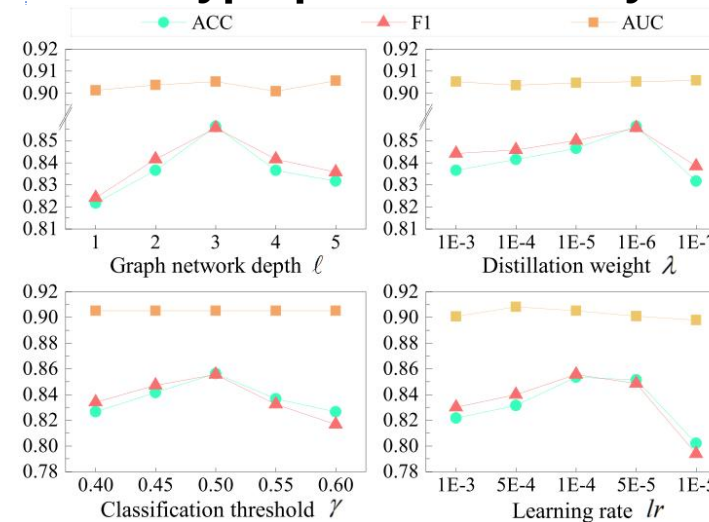
RQ5. Training Overhead Analysis

Dual Graph Structure + Gated Knowledge Distillation -> more efficient and effective concept embedding learning

RQ6 (Appendix). Impact of Directed Graph Feature Fusion Methods

Operation	UCD			LectureBank			MOOC		
	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
Addition (Ours)	**0.8564**	**0.8557**	0.9053	**0.8347**	**0.8246**	0.8795	**0.8756**	**0.8718**	**0.9236**
Concat+MLP	0.8416	0.8416	**0.9074**	0.8099	0.8099	0.8776	0.8557	0.8497	0.9211
Maximum	0.8515	0.8515	0.9048	0.8264	0.8205	0.8724	0.8657	0.8601	0.9190
Average	0.8416	0.8384	0.9011	0.8017	0.7778	**0.8842**	0.8657	0.8643	0.9214
Minimum	0.8515	0.8469	0.9058	0.7851	0.7636	0.8727	0.8507	0.8485	0.9134

RQ7 (Appendix). Impact of pre-training language model

Method	UCD			LectureBank			MOOC		
	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
DGCPL (Ours)	**0.8564**	**0.8557**	**0.9053**	**0.8347**	**0.8246**	**0.8795**	**0.8756**	**0.8718**	**0.9236**
LCPRE	0.8366	0.8216	0.8884	0.8182	0.8000	0.8514	0.8258	0.8223	0.8898
HGAPNet	0.8400	0.8333	0.9044	0.7833	0.7759	0.8672	0.7900	0.7835	0.8573
ConLearn	0.7822	0.7684	0.8529	0.8017	0.7931	0.8541	0.7562	0.7200	0.8472

IJCAI 2025