Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)
Special Track on AI and Social Good

LLM-Enhanced Knowledge and Learning Path Understanding for Graph-based Educational Recommendation

Qingqing Liang[1,2] , **Shuyan Zheng**[1] , **Peiwei Xia**[1] , **Chunyang Wang**[1] , **Xuesong Lu**[1*] and **Aoying Zhou**[1]

[1]School of Data Science and Engineering, East China Normal University
[2]College of Computer Science and Technology, Guizhou University
{qqliang.dase, shuyanzheng, peiwei.xia}@stu.ecnu.edu.cn, {cywang, xslu, ayzhou}@ecnu.edu.cn

Abstract

Educational recommendations empower personalized learning by suggesting suitable learning resources to learners, and the graph-based recommenders are widely adopted. Existing methods are mainly ID-based, which initialize learners and resources with trainable identifiers and optimize their representations solely from the interaction graph. As a result, the lack of semantic understanding of learning resources and learning paths hinders further improvements in recommendation accuracy. To alleviate the problem, we propose KLU4EduRec, which leverages large language models (LLMs) to understand resource knowledge and learning paths, thereby enhancing traditional graph-based educational recommenders. Specifically, for learning path understanding, we segment a learning path by detecting learning pattern drift in resource knowledge sequence, and prompt LLMs to infer learners' learning patterns within each segment. The segment-level patterns are then chronologically aggregated to represent the overall learning path. Besides, we prompt LLMs to summarize the core knowledge of learning resources from their content as complementary semantic signals. Finally, the resulting semantic representations are aligned and fused with structural representations learned by a graph-based recommender to enable more accurate recommendations. Extensive experiments show that KLU4EduRec greatly outperforms existing methods, including traditional ID-based methods and recent LLM-powered methods. A case study shows how the understanding of pattern drift in a learning path leads to more suitable recommendations. Our codes are available at https://github.com/DaSESmartEdu/KLU4EduRec.

1 Introduction

Large Language Models (LLMs) have been leveraged to enhance the semantic understanding of items and users in recommender systems [Bao *et al.*, 2023; Liu *et al.*, 2024; Shan *et al.*, 2025]. The extra semantic information enriches the structural representations learned by traditional ID-based models, therefore further improving the recommendation accuracy as well as enhancing the interpretability of recommendation results. Compared to commercial recommendations, educational recommender systems [Gong *et al.*, 2020; Wang *et al.*, 2023; Jiang *et al.*, 2023; Liang *et al.*, 2023; Gu *et al.*, 2024] are expected to have greater interpretability and higher accuracy, as there is much less room for making irrational or incorrect recommendations for learners. This relies on a precise understanding of the knowledge carried by the learning resources, as well as the learning patterns of the learners. Recently, Li et al. [2024] are among the first to introduce LLMs into educational recommendations. They leverage factual knowledge from LLMs and relationships between concepts obtained from knowledge graphs to construct textual representations of concepts, which are then adapted to the concept recommendation task. Ozyurt et al. [2025] use LLMs to annotate questions with their corresponding knowledge concepts and learn their semantic representations. The representations are used to train a knowledge tracing [Abdelrahman *et al.*, 2023] model, which serves as the environment of reinforcement learning for recommending exercises.

Despite the progress, we observe that existing LLM-assisted educational recommender systems often focus on obtaining the semantic representations of learning resources via LLMs, and overlook the semantic understanding of learners' behavioral patterns, thereby resulting in suboptimal matching between recommended resources and learners. Also, rich textual information, such as video subtitles and question descriptions, has not been fully utilized for generating the semantic representations of learning resources. To bridge the gap, we propose an LLM-enhanced method **KLU4EduRec**, which strengthens **K**nowledge and **L**earning path **U**nderstanding via LLMs for graph-based **Edu**cational **Rec**ommendation. For each learning resource, we prompt an LLM to generate a concise summary of core knowledge by providing it with detailed textual information such as resource title, content and related concepts, which is adapted and fed into a GNN encoder to obtain the semantic representation. For each learner, we segment the learning path according to the semantic drift of knowledge interacted with by the learner and ask an LLM to summarize the learning pattern entailed in each segment. Then we obtain the semantic representation of the learning

*Corresponding authors

pattern in each segment via an adapter, and use a GRU to chronologically aggregate the semantic representations of the entire learning path, which represents the overall behavioral pattern of the learner. The hierarchical representation extraction enables a deeper understanding of the learning path, thereby generating a more accurate semantic representation of the learner. Finally, the semantic representations of resources and learners are aligned and fused with the structural representations obtained via a GNN encoder, which are used to compute the ratings of the learner to the learning resources.

Experiments against representative ID-based and LLM-powered recommenders demonstrate the superiority of KLU4EduRec. Ablation and case studies further validate the effectiveness of the segmentation-then-aggregation paradigm for learning path understanding.

2 Related Work

Traditional educational recommender systems are mainly ID-based and leverage collaborative filtering algorithms, graph neural networks (GNNs) and reinforcement learning (RL) algorithms. In early studies, Jing et al. [2017] and Zhao et al. [2020] design collaborative filtering algorithms to learn learners' interests and recommend suitable courses to them. Zhang et al. [2019] use RL to remove noisy courses in the learner profile when a learner has interests in many different courses. Zhang et al. [2023] and Deng et al. [2023] build course graphs and learn the representations using GNNs. Huang et al. [2019] and Liu et al. [2023] propose RL-based methods for exercise recommendations, which devise proper rewards to guide recommendation actions. Some recent studies investigate more fine-grained knowledge concept recommendations. Gong et al. [2020] utilize graph convolutional networks to learn knowledge concept representations in a heterogeneous information network (HIN), and predict the ratings of learners on knowledge concepts. Gong et al. [2023] use RL to capture dynamic interactions between learners and knowledge concepts, and employ GNNs to represent learners in an HIN. Gu et al. [2024] utilize GNNs to represent the explicit and implicit relations in an HIN, and use contrastive learning to optimize the representations. Liang et al. [2025] develop a dual dynamic GNN module to disentangle preference and knowledge status from a continuous-time HIN.

Due to the strong comprehension and reasoning abilities, large language models (LLMs) have been recently explored to reshape traditional recommender systems. Existing studies generally follow two paradigms: LLM-as-a-recommender and LLM-as-an-enhancer. In the former paradigm, TALLRec [Bao *et al.*, 2023] fine-tunes LLMs with recommendation-specific data using instruction tuning. EXP3RT [Kim *et al.*, 2025] leverages rich preference information contained in user and item reviews, and distills from a teacher LLM to perform preference extraction, profile construction and textual reasoning. SPRec [Gao *et al.*, 2025] mitigates over-recommendation and improves fairness of LLMs using supervised fine-tuning and Direct Preference Optimization. In the latter paradigm, RLMRec [Ren *et al.*, 2024] employs LLMs for user and item profiling. The semantic space of LLMs is aligned with collaborative relational signals through cross-view alignment. LLM2Rec [He *et al.*, 2025] adapts LLMs to infer item relationships according to historical interactions, and refines the specialized LLMs into structured item embedding models. The current use of LLMs in educational recommendations follows the LLM-as-an-enhancer paradigm. For example, SKarREC [Li *et al.*, 2024] uses LLMs as a factual knowledge source to enrich concept text and enhance textual representations of concepts. The anisotropic text embeddings are adapted to the concept recommendation task via an adapter. ExRec [Ozyurt *et al.*, 2025] uses LLMs to generate a step-by-step solution for each question and map concepts to the solution steps. The mapping serves representation learning of questions and concepts, which are further used for RL-based exercise recommendation. Our KLU4EduRec also belongs to the LLM-as-an-enhancer paradigm, where we are the first to leverage LLMs to draw semantic representations of learners for educational recommendation.

3 Preliminaries

3.1 The Task of Educational Recommendation

Let U and R denote the sets of learners and learning resources, respectively. Given an observed learning path $\mathcal{P}_u = [r_1^u, r_2^u, \cdots, r_t^u]$, which records the chronological sequence of resources accessed by learner u, the educational recommendation task aims to predict a likelihood score $\hat{y}_{ur}$ for each resource $\{r | r \in R \wedge r \notin \mathcal{P}_u\}$ and generate a list of top-k resources that u is most likely to learn.

3.2 Graph-based Educational Recommenders

The common data on learning platforms include learning resources, knowledge concepts, learners and their interactions, which naturally form a complex graph structure. Consequently, graph-based recommendation methods have become a mainstream paradigm in educational recommendation.

Formally, an educational graph is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the node set $\mathcal{V}$ contains heterogeneous educational entities, such as learners, learning resources (e.g., videos and questions), courses and knowledge concepts. The edge set $\mathcal{E}$ contains multiple types of relations, such as learner-resource interactions derived from learning paths, course-resource relations, resource-concept relations. All entities and relations are typically initialized with identifiers. Graph neural networks (GNNs) are commonly used to learn low-dimensional representations for learners and resources from $\mathcal{G}$, which can be abstractly formulated as:

$$\mathbf{e}_u, \mathbf{e}_r = \mathcal{F}(\mathbf{e}_u^0, \mathbf{e}_r^0, \mathcal{G}), \tag{1}$$

where $\mathbf{e}_u^0$ and $\mathbf{e}_r^0$ are the initial embeddings of learner u and resource r, respectively, $\mathcal{F}(\cdot)$ represents a GNN encoder that propagates information over $\mathcal{G}$, and $\mathbf{e}_u$ and $\mathbf{e}_r$ are the learned representations. The likelihood score between u and r is usually computed by inner product $\hat{y}_{ur} = \mathbf{e}_u \cdot \mathbf{e}_r$, and the model is typically optimized using the Bayesian Personalized Ranking (BPR) loss:

$$\mathcal{L}_{rec} = \sum_{(u,r^+,r^-)\in\mathcal{O}} -\ln \sigma(\hat{y}_{u,r^+} - \hat{y}_{u,r^-}), \tag{2}$$

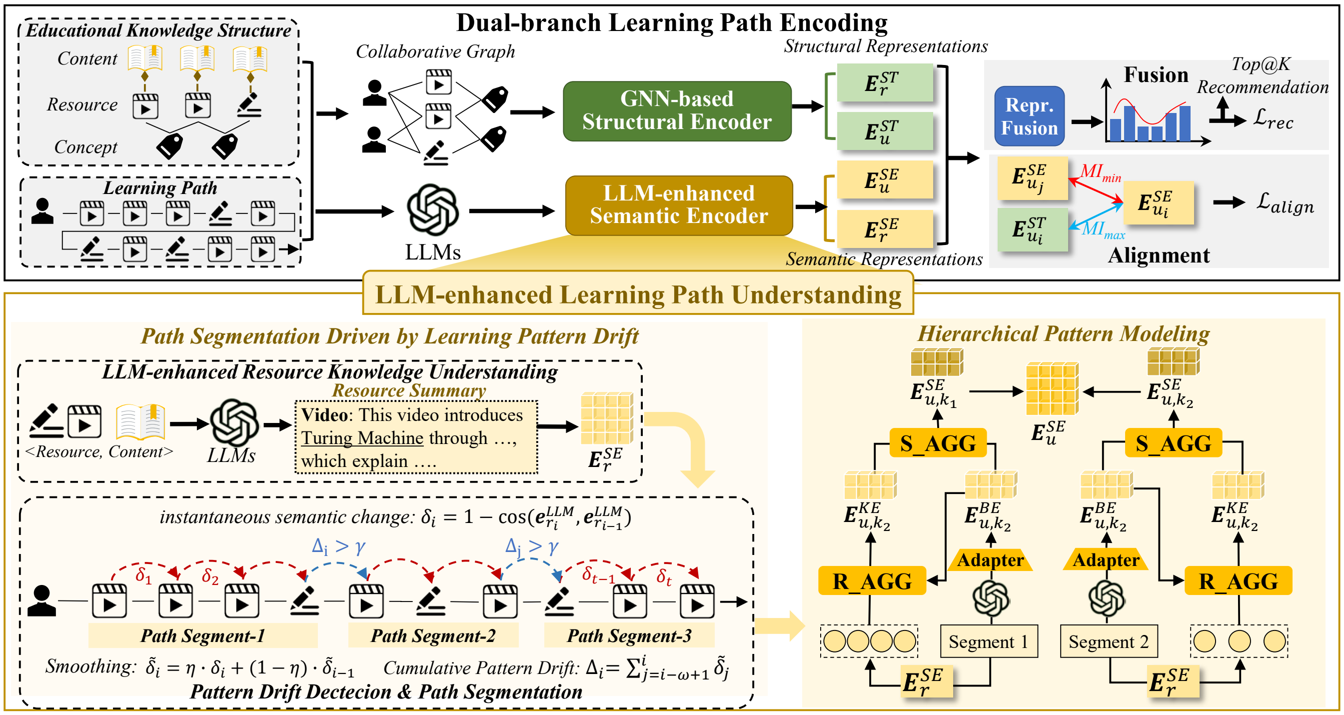

Figure 1: The overall architecture of KLU4EduRec.

where r^+ and r^- denote the observed (positive) and unobserved (negative) resources of u, respectively. Such ID-based graph recommenders rely solely on graph structures and lack explicit semantic understanding of resources and learning paths, which motivates our semantic-enhanced framework.

4 The KLU4EduRec Method

KLU4EduRec has a dual-branch architecture for modeling learning patterns from both structural and semantic perspectives. As illustrated in the upper part of Figure 1, the green branch employs a GNN-based encoder to capture structural representations from an educational graph, and the yellow branch leverages an LLM-enhanced encoder to model semantic representations from learning paths. In particular, as shown in the lower part, the LLM-enhanced branch extracts resource knowledge, segments a learning path based on the drift of learning patterns, and hierarchically aggregates the semantic representations to model the entire path. Finally, the structural and semantic representations learned by the two branches are aligned and fused to support recommendations.

4.1 Dual-branch Learning Path Encoding

GNN-based Structural Encoder

The GNN-based structural branch models learners' learning paths from an interaction-driven perspective. Given an educational collaborative graph $\mathcal{G}$ that describes learners' interactions with learning resources, the structural representations of learner u and resource r are obtained as:

$$\mathbf{e}_u^{ST}, \mathbf{e}_r^{ST} = \mathcal{F}_{GNN}(\mathbf{e}_u^{ID}, \mathbf{e}_r^{ID}, \mathcal{G}), \tag{3}$$

where $\mathbf{e}_u^{ID}$ and $\mathbf{e}_r^{ID}$ are ID-based initial embeddings and $\mathcal{F}_{GNN}(\cdot)$ denotes a GNN-based encoder. We instantiate $\mathcal{F}_{GNN}$ with LightGCN [He *et al.*, 2020].

LLM-enhanced Semantic Encoder

Complementary to the purely ID-based structural branch, the semantic branch models learners' learning paths from a semantic perspective using LLMs, which captures both resource knowledge semantics and learners' learning pattern semantics. Specifically, LLMs are applied to extract knowledge semantics that a resource conveys and infer the learning pattern of a learning path. Abstractly, the semantic representations of u and r are obtained as:

$$\mathbf{e}_u^{SE}, \mathbf{e}_r^{SE} = \mathcal{F}_{LLM}(\mathcal{C}_r, \mathcal{P}_u), \tag{4}$$

where $\mathcal{C}_r$ denotes the detailed content of resource r and $\mathcal{P}_u$ denotes the learning path of learner u. This branch provides a high-level semantic interpretation of learning paths, which is further refined in Section 4.2.

Representation Fusion and Alignment

The dual-branch architecture produces two complementary representations for each learner and each learning resource. Effectively integrating these two representations is therefore critical for accurate recommendations.

We adopt a gated fusion mechanism to combine structural and semantic representations. For each entity x, which represents either a learner u or a learning resource r, the fused representation is computed as:

$$\mathbf{e}_x = g_x \cdot \mathbf{e}_x^{ST} + (1 - g_x) \cdot \mathbf{e}_x^{SE}, \tag{5}$$

where $\mathbf{e}_x^{ST}$ and $\mathbf{e}_x^{SE}$ denote the structural and semantic representations of x, and g_x is a learnable fusion gate that adaptively balances the two representations. The fused representations $\mathbf{e}_u$ and $\mathbf{e}_r$ are used to compute the BPR loss $\mathcal{L}_{rec}$ as described in Equations 1 and 2.

Additionally, the two types of representations are learned from fundamentally different modeling paradigms. The structural representations rely on graph-based collaborative learning, whereas the semantic representations are obtained via LLM-based semantic reasoning. This discrepancy may lead to representation inconsistency and hinder effective fusion. To mitigate the issue, we introduce a structural-semantic alignment objective based on an InfoNCE contrastive loss, which aligns structural and semantic representations of the same entity while discriminating different ones:

$$\mathcal{L}_{align}^{x} = \sum_{x} -\log \frac{\exp(\mathrm{sim}(\mathbf{e}_x^{ST}, \mathbf{e}_x^{SE})/\tau)}{\sum_{x'} \exp(\mathrm{sim}(\mathbf{e}_x^{ST}, \mathbf{e}_{x'}^{SE})/\tau)} \quad (6)$$

where $\mathrm{sim}(\cdot,\cdot)$ denotes cosine similarity, τ is a temperature hyperparameter, and $x' \in U$ or $x' \in R$ depends on whether x denotes u or r.

The overall training objective is defined as:

$$\mathcal{L} = \mathcal{L}_{rec} + \alpha \cdot \mathcal{L}_{align}^{u} + \beta \cdot \mathcal{L}_{align}^{r} + \lambda \cdot \|\Phi\|^2 \quad (7)$$

where $\mathcal{L}_{rec}$ is the recommendation loss defined in Equation 1, $\mathcal{L}_{align}^{u}$ and $\mathcal{L}_{align}^{r}$ denote the alignment losses for learners and resources, respectively, Φ denotes the trainable parameters, and α, β and λ are hyperparameters.

4.2 LLM-enhanced Learning Path Understanding

A naive way of applying LLMs for learning path understanding is to directly input the entire learning path and prompt an LLM to summarize the learning pattern. However, it faces two key limitations: (1) a learning path is often long and the information exceeds the context capacity of LLMs; (2) modeling the entire path directly leads to a coarse representation that obscures the changes within the learning process.

To overcome the issues, we propose a hierarchical framework for learning path understanding driven by learning pattern drift. Specifically, we define "learning pattern" as the content (i.e., *what*) a learner learns and the style (i.e., *how*) a learner learns in. Based on the definition, we first extract the semantic representations of learning resources, then segment a learning path according to the pattern drift reflected by the semantic changes of learning resources within the path, and finally aggregate the semantic representations in a hierarchical manner to represent the entire path.

Resource Knowledge Understanding

Learning resources, such as videos and questions, contain rich semantic information that reflects the underlying knowledge. However, the raw information is often noisy and verbose, which may hinder effective semantic modeling.

To obtain concise and knowledge-focused representations, we design a resource-specific prompt p_r to guide an LLM to summarize the core knowledge of r, as shown in Figure 2. The summarized knowledge is encoded and projected into the recommendation space and propagated over the educational graph, which is calculated as:

$$\begin{aligned} \mathbf{e}_r^{LLM} &= \mathrm{LLM}_{emb}(\mathrm{LLM}(p_r)), \\ \mathbf{e}_r^{SE} &= \mathrm{GNN}(\mathrm{ADPT}_r(\mathbf{e}_r^{LLM}), \mathcal{G}), \end{aligned} \quad (8)$$

where LLM_{emb} denotes a text-embedding model and ADPT_r is a lightweight adapter. The LLM-generated semantic representation $\mathbf{e}_r^{LLM}$ is frozen to preserve the knowledge semantics inferred by the LLM. The resulting $\mathbf{e}_r^{SE}$ serves as a structure-and-knowledge-aware resource representation for subsequent learning pattern modeling.

Assume you are an educational content analysis expert. Below is the information of a learning resource.

Resource Type: [video / question]
Resource Title: [video / question title]
Resource Content: [video subtitles / question description]
Belong Course: [course name]
Related Concepts: [concept_1, concept_2, ...]

Please provide a concise summary of the core knowledge covered by the resource, ignoring redundant or irrelevant information. The summary should be no more than 80 words.

Figure 2: Resource knowledge summarization prompt.

Path Segmentation Driven by Learning Pattern Drift

Educational motivation theory [Schunk *et al.*, 2014] points out that students' learning motivation typically evolves in a stage-wise manner and remains relatively stable within each learning stage, leading to shifts in learning patterns over time. Motivated by this observation, we segment a learning path according to the semantic drift of learning patterns, so as to achieve a more fine-grained understanding of the entire path.

Given a learning path $\mathcal{P}_u = [r_1^u, r_2^u, \cdots, r_t^u]$ of u, we first quantify the instantaneous semantic change between two consecutive resources as:

$$\delta_i = 1 - \cos(\mathbf{e}_{r_i}^{LLM}, \mathbf{e}_{r_{i-1}}^{LLM}), \quad (9)$$

where $\mathbf{e}_{r_i}^{LLM}$ and $\mathbf{e}_{r_{i-1}}^{LLM}$ are the LLM-derived semantic representations of r_i and r_{i-1}, and $\cos(\cdot,\cdot)$ denotes the cosine similarity. The change $\delta_i \in [0,1]$ measures local semantic fluctuation caused by short-term exploration or noise.

To suppress sensitivity to such local fluctuations and capture the inertia of pattern evolution, we apply moving average to smooth the instantaneous semantic changes:

$$\tilde{\delta}_i = \eta \cdot \delta_i + (1-\eta) \cdot \tilde{\delta}_{i-1}, \quad (10)$$

where $\tilde{\delta}_i$ is the smoothed semantic change at step i and $\eta \in [0,1]$ controls the smoothing strength. We further calculate the cumulative semantic drift using a sliding window of size ω as:

$$\Delta_i = \sum_{j=i-\omega+1}^{i} \tilde{\delta}_j, \quad (11)$$

where Δ_i reflects the semantic drift over the past ω steps until i. When Δ_i exceeds a predefined threshold γ, a learning pattern drift is detected and a new segment is initiated. The process is repeated until the end of $\mathcal{P}_u$.

Hierarchical Pattern Aggregation

To model fine-grained learning patterns in a structured and progressive manner, we design a hierarchical aggregation framework with three levels: resource, segment, and path.

> Assume you are an experienced teacher skilled in analyzing learning behaviors of learners and inferring their underlying learning patterns. Below is a learning behavior sequence within a learning stage, where each behavior is recorded as: [Resource ID] Resource Title, Behavior Type (e.g., video-watching or question-answering), Timestamp, Engagement Indicator (e.g., duration or correctness), Belong Course, Related Concepts.
>
> **Learner Behaviors**: ...
>
> Please summarize the learning pattern in this stage, including the learning topic, the behavioral pattern, and the temporal continuity. The summary should be no more than 80 words.

Figure 3: Learning pattern summarization prompt for each segment.

At the resource level, we characterize the local pattern within a segment k by extracting two semantic signals: behavior semantics and knowledge semantics, which together answer *what* and *how* the learner studies at this stage. For behavior semantics, we prompt an LLM to summarize the behaviors within segment k and obtain a concise textual summary $s_{u,k}$, as shown in Figure 3. It is then encoded and projected into the recommendation space as:

$$\mathbf{e}_{u,k}^{BE} = \text{ADPT}_u(\text{LLM}_{emb}(s_{u,k})), \tag{12}$$

where ADPT_u is a lightweight adapter, and $\mathbf{e}_{u,k}^{BE}$ denotes the behavior semantic representation of learner u in segment k. For knowledge semantics, we focus on *what* the learner studies in segment k. Let $[r_{k,j}^u]_{j=1}^{L_k}$ denote the resources interacted by u within segment k, where L_k is the segment length, using the behavior semantic representation as a query, we aggregate the corresponding resource knowledge semantics as:

$$\mathbf{e}_{u,k}^{KE} = \text{R\_AGG}(\mathbf{e}_{u,k}^{BE}, [\mathbf{e}_{r_{k,j}^u}^{SE}]_{j=1}^{L_k}), \tag{13}$$

where $\text{R\_AGG}(\cdot)$ denotes a query-conditioned aggregation function, which can be instantiated as an attention-based function weighting over resource semantics. The resulting $\mathbf{e}_{u,k}^{KE}$ summarizes the knowledge learned by u in segment k.

At the segment level, we fuse the behavior and knowledge semantics to form a segment-level representation:

$$\mathbf{e}_{u,k}^{SE} = \text{S\_AGG}(\mathbf{e}_{u,k}^{BE}, \mathbf{e}_{u,k}^{KE}), \tag{14}$$

where $\text{S\_AGG}(\cdot)$ can be any fusion function, which is instantiated as element-wise addition in our implementation. The resulting representation $\mathbf{e}_{u,k}^{SE}$ jointly characterizes *what* the learner studies and *how* the learner engages with resources during a coherent learning stage.

At the path level, we model the temporal dependencies across the segments to capture pattern evolution over time. Given the segment-level pattern sequence $[\mathbf{e}_{u,k}^{SE}]_{k=1}^{T_u}$ for u, where T_u denotes the number of segments in the learning path, we perform sequential aggregation as:

$$\mathbf{e}_{u}^{SE} = \text{P\_AGG}([\mathbf{e}_{u,k}^{SE}]_{k=1}^{T_u}), \tag{15}$$

where $\text{P\_AGG}(\cdot)$ is a temporal aggregation function, which is instantiated using the gated recurrent unit (GRU) [Cho *et al.*, 2014] in our implementation. The resulting representation $\mathbf{e}_{u}^{SE}$ summarizes the learner's overall learning pattern by jointly considering the knowledge, the learning intention, and the temporal progression in the learning path.

4.3 Model Training and Inference

LLM-based semantic extraction is computationally expensive while the extracted semantics are relatively stable over time. We therefore decouple semantic extraction from model training and inference, following a three-stage pipeline: (1) offline semantic extraction, where LLMs are used offline to extract resource knowledge semantics and segment-level behavior semantics, and the obtained semantic representations are frozen; (2) model training, where the recommender model is trained using the interaction data and the pre-extracted semantic representations; (3) online inference, where for newly observed learning behaviors, the path segmentation and temporal aggregation are incrementally performed to update learners' representations, i.e., LLMs are invoked only for newly formed segments.

5 Experiment

5.1 Datasets

We conduct experiments on two real-world educational datasets, **MOOCCubeX** [Yu *et al.*, 2021] and **MOOCCube** [Yu *et al.*, 2020]. Both datasets contain rich and well-structured learning resources (e.g., videos and questions) with informative textual content. Following the settings in [Wei *et al.*, 2024; Ren *et al.*, 2024], we filter out learners whose learning paths are shorter than 10. For each learner, the path is chronologically split into training, validation, and testing sets with a ratio of 8:1:1. Following the item-ranking setting, we adopt Recall@K, NDCG@K, and MRR@K as evaluation metrics, where K is set to 5 and 10. To ensure comprehensive evaluation and mitigate bias, we employ the all-ranking protocol [He *et al.*, 2020; Yang *et al.*, 2023], which ranks positive items against all unobserved items. The statistics of the datasets are reported in Table 1.

Dataset	MOOCCubeX	MOOCCube
Students	9,506	9,831
Resources	7,219	11,885
Graph Edges	659,714	935,943
Graph Density	0.96%	0.80%
Avg. Learning Path Length	69.40	95.20

Table 1: Summary statistics of the datasets.

Model	MOOCCubeX					MOOCCube				
	Recall@5	Recall@10	NDCG@5	NDCG@10	MRR@10	Recall@5	Recall@10	NDCG@5	NDCG@10	MRR@10
LightGCN [2020]	0.4479	0.5969	0.4711	0.5698	0.3355	0.1746	0.2784	0.1619	0.2182	0.1920
KGRec [2023]	0.5049	0.6548	0.5450	0.6457	0.4030	0.1971	0.2861	0.1895	0.2365	0.2245
RecDCL [2024]	0.3761	0.5071	0.3968	0.4817	0.3049	0.1766	0.2673	0.1646	0.2144	0.1820
LLMRec [2024]	<u>0.5500</u>	0.7225	<u>0.5657</u>	<u>0.6782</u>	0.3999	<u>0.2038</u>	<u>0.3025</u>	<u>0.1994</u>	<u>0.2500</u>	0.2236
RLMRec [2024]	0.5452	<u>0.7249</u>	0.5636	0.6773	<u>0.4218</u>	0.1899	0.2886	0.1745	0.2282	0.1995
CIKGRec [2025]	0.5125	0.6678	0.5423	0.6549	0.4096	0.1932	0.2847	0.1936	0.2472	<u>0.2254</u>
ACKRec [2020]	0.5255	0.6691	0.5590	0.6521	0.3831	0.1902	0.2653	0.1866	0.2290	0.2187
CL-KCRec [2024]	0.5269	0.6715	0.5574	0.6657	0.3833	0.1866	0.2892	0.1896	0.2456	0.2128
KCD [2025]	0.5079	0.6515	0.5436	0.6437	0.3819	0.1852	0.2821	0.1782	0.2219	0.2072
KLU4EduRec	**0.6001** (↑ 0.0501)	**0.7765** (↑ 0.0516)	**0.6196** (↑ 0.0539)	**0.7335** (↑ 0.0553)	**0.4388** (↑ 0.0170)	**0.2232** (↑ 0.0194)	**0.3247** (↑ 0.0222)	**0.2129** (↑ 0.0135)	**0.2691** (↑ 0.0191)	**0.2456** (↑ 0.0202)

Table 2: KLU4EduRec Vs. existing models. The best results are bold and the second-best results are underlined.

5.2 Baselines

We compare KLU4EduRec with 9 representative recommendation models: LightGCN [He *et al.*, 2020], KGRec [Yang *et al.*, 2023], RecDCL [Zhang *et al.*, 2024], LLMRec [Wei *et al.*, 2024], RLMRec [Ren *et al.*, 2024], CIKGRec [Hu *et al.*, 2025], ACKRec [Gong *et al.*, 2020], CL-KCRec [Gu *et al.*, 2024], KCD [Dong *et al.*, 2025]. The first three are general GNN-based recommenders, the next three are LLM-enhanced GNN-based recommenders, the following two are graph-based educational recommenders, and the last one is an LLM-enhanced model for cognitive diagnosis from learning paths adapted to educational recommendation[1][2]. For the LLM-based models, LLMRec and CIKGRec use LLMs to enhance user/item semantics or construct graphs, while RLMRec and KCD align LLM-derived representations with traditional models. We use Qwen3-235B-A22B-Instruct-2507[3] to generate textual summaries for both learning resources and learning paths. The generated summaries are then encoded using the Qwen3-Embedding-8B[4] model to obtain the semantic representations.

5.3 Experimental Setting

We implement KLU4EduRec in PyTorch and optimize it using Adam. In the offline semantic extraction stage, learning paths are segmented based on pre-extracted resource knowledge semantics, where the smoothing factor η is set to 0.3, and the sliding window size ω is set to 5. The semantic drift threshold γ is determined as 75% of the semantic drift distribution, which is set to 2.02 for MOOCCubeX and 2.42 for MOOCCube. The learning rate is set to $5\mathrm{e}^{-3}$, and the L2 regularization coefficient λ is set to $5\mathrm{e}^{-4}$ for MOOCCubeX and $5\mathrm{e}^{-5}$ for MOOCCube. The embedding size of the recommender d is set to 64. For graph propagation, we use 2 layers for MOOCCubeX and 1 layer for MOOCCube. The loss weight α is set to $5\mathrm{e}^{-5}$ and β is set to $5\mathrm{e}^{-3}$ for MOOCCubeX and $1\mathrm{e}^{-4}$ for MOOCCube. All experiments are conducted on an NVIDIA A800 GPU with 80GB memory.

5.4 Overall Performance

Table 2 reports the overall performance of KLU4EduRec against baselines. All results are averaged over 4 runs with random seeds. We derive the following observations: (1) KLU4EduRec consistently achieves the best performance across all metrics on both datasets. Compared to the second-best model for each metric (underlined), it yields a clear improvement. (2) Compared to the first three GNN-based methods, KLU4EduRec benefits from the semantic understanding of learning resources and learning paths. (3) Compared to the next three LLM-enhanced recommendation models, KLU4EduRec outperforms by modeling learning paths in a stage-aware manner. While these models treat a learning path as a whole, KLU4EduRec captures learning pattern drift by path segmentation, leading to more precise stage-aligned recommendations. (4) The following two educational baselines without semantic modeling, i.e., ACKRec and CL-KCRec, perform worse than the LLM-enhanced general recommenders, indicating that structural signals are insufficient for understanding learning paths. (5) The last LLM-enhanced educational model KCD performs worse because it directly summarizes the entire path with LLMs.

5.5 Hyperparameter Analysis

We analyze the sensitivity of KLU4EduRec to four key hyperparameters: the embedding size d, the number of graph propagation layers L, the learner semantic alignment weight α, and the resource semantic alignment weight β. As shown in Figure 4, the performance first improves and then slightly degrades as d increases, indicating that moderate embedding sizes strike a good balance between representation capacity and overfitting. In particular, $d = 64$ yields the best results

[1]KCD is originally designed for cognitive diagnosis from learning paths and is adapted to recommendation by replacing the diagnosis model with a GNN-based recommender and retaining its alignment strategy.

[2]We don't compare with SKarREC because it is a sequential model and does not release source code, and we don't compare with ExRec because it cannot be directly adapted to recommending videos since it needs to construct a knowledge tracing model.

[3]https://huggingface.co/Qwen/Qwen3-235B-A22B-Instruct-2507

[4]https://huggingface.co/Qwen/Qwen3-Embedding-8B

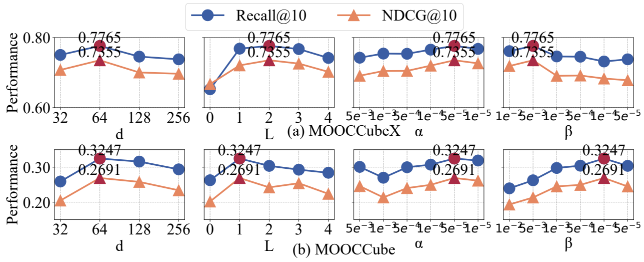

Figure 4: Parameter sensitivity analysis.

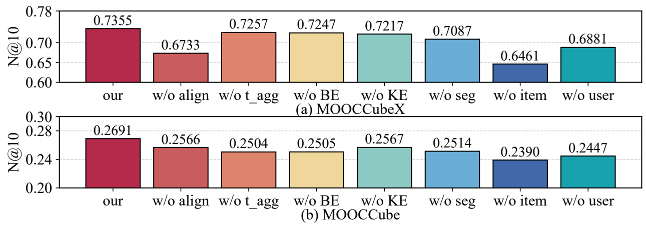

Figure 5: Ablation study.

on both datasets. Second, we observe the best performance at $L = 2$ on MOOCCubeX and $L = 1$ on MOOCCube. The performance drops for deeper propagation suggest potential over-smoothing effects on sparse educational graphs. Third, KLU4EduRec performs the best with $\alpha = 5\mathrm{e}^{-3}$ and $\beta = 5\mathrm{e}^{-5}$ on MOOCCubeX, and with $\alpha = 1\mathrm{e}^{-4}$ and $\beta = 5\mathrm{e}^{-5}$ on MOOCCube. The results highlight the importance of careful balancing between the semantic alignment objectives and the recommendation task.

5.6 Ablation Study

We design seven ablated variants: *w/o user* and *w/o item*, which remove learner and resource semantics, respectively; *w/o seg*, which applies global summarization over the entire learning path; *w/o KE* and *w/o BE*, which remove knowledge and behavior semantics in each segment, respectively; *w/o t_agg*, which removes temporal aggregation and averages segment-level representations; and *w/o align*, which discards the alignment losses in Equation 7. The results are reported in Figure 5. KLU4EduRec consistently outperforms all ablated variants on both datasets, demonstrating the effectiveness of each component. The performance drops of *w/o user*, *w/o item* and *w/o align* validate the importance of jointly learning learner and resource semantics and explicitly aligning semantic and structural representations. Moreover, removing behavior semantics (*w/o BE*) or knowledge semantics (*w/o KE*) also degrades performance, indicating that both semantic signals are important for learning path understanding. Finally, removing path segmentation (*w/o seg*) or temporal aggregation (*w/o t_agg*) further degrades performance, demonstrating the necessity of modeling stage-wise learning patterns.

5.7 Case Study

To demonstrate the effectiveness of capturing the learning pattern drift, we present a case study in Figure 6. The learner shows a learning path in the DATA STRUCTURE course,

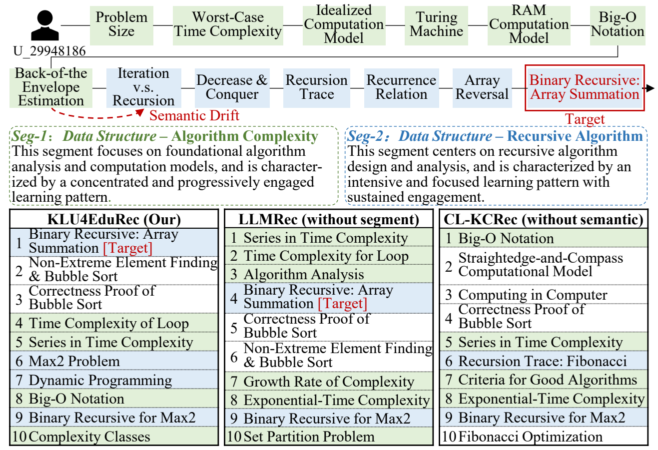

Figure 6: Case study: importance of capturing learning pattern drift.

where a clear learning pattern drift can be observed, i.e., shifting from learning algorithm complexity (green) to recursion (blue). We compare KLU4EduRec with two methods: the best-performing traditional ID-based model CL-KCRec, and the best-performing LLM-enhanced model without learning path segmentation, i.e., LLMRec. We observe that CL-KCRec produces recommendations that are less related to the learner's current learning stage, as it fails to capture semantic relations among learning resources. LLMRec mixes recommendations across two stages, as it models the learning path as a whole without capturing the learning pattern drift. In contrast, KLU4EduRec detects the drift and performs path segmentation, resulting in recommendations that are both semantically coherent and well aligned with the current stage.

6 Collaboration with Stakeholders

The work stems from collaboration with a nation-wide online learning platform in China. Over the past two years, we have worked with its managers, developers, and educators to build personalized learning services, where resource recommendation is a key service. In this study, we aim to improve recommendation accuracy and interpretability with LLMs. Educators help annotate and verify textual information of resources, while managers and developers support model deployment and testing on the platform. Our vision is that through educational recommendation on the platform, we can enable guided learning with minimal efforts of teachers, thus scaling online learning to as many students as possible, especially in rural areas that lack good educational resources and teachers.

7 Conclusion

In this paper, we propose KLU4EduRec, an LLM-enhanced method for graph-based educational recommendation that jointly models resource knowledge semantics and learning path semantics. KLU4EduRec leverages LLMs for both resource understanding and stage-wise learning pattern inference, and integrates semantic and structural representations of resources and learners through a dual-branch architecture. Extensive experiments demonstrate its effectiveness and robustness on real-world educational datasets.

Acknowledgments

This work is supported by the National Key R&D program of China (2023YFC3341200) and the National Natural Science Foundation of China (Grant No. 62277017, 62137001, and 62507015).

References

[Abdelrahman *et al.*, 2023] Ghodai Abdelrahman, Qing Wang, and Bernardo Nunes. Knowledge tracing: A survey. *ACM Computing Surveys*, 55(11):1–37, 2023.

[Bao *et al.*, 2023] Keqin Bao, Jizhi Zhang, Yang Zhang, Wenjie Wang, Fuli Feng, and Xiangnan He. Tallrec: An effective and efficient tuning framework to align large language model with recommendation. In *Proceedings of the 17th ACM conference on recommender systems*, pages 1007–1014, 2023.

[Cho *et al.*, 2014] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar, 2014. Association for Computational Linguistics.

[Deng *et al.*, 2023] Weiwei Deng, Peihu Zhu, Han Chen, Tao Yuan, and Ji Wu. Knowledge-aware sequence modelling with deep learning for online course recommendation. *Information Processing & Management*, 60(4):103377, 2023.

[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.

[Gao *et al.*, 2025] Chongming Gao, Ruijun Chen, Shuai Yuan, Kexin Huang, Yuanqing Yu, and Xiangnan He. Sprec: Self-play to debias llm-based recommendation. In *Proceedings of the ACM on Web Conference 2025*, pages 5075–5084, 2025.

[Gong *et al.*, 2020] Jibing Gong, Shen Wang, Jinlong Wang, Wenzheng Feng, Hao Peng, Jie Tang, and Philip S Yu. Attentional graph convolutional networks for knowledge concept recommendation in moocs in a heterogeneous view. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, pages 79–88, 2020.

[Gong *et al.*, 2023] Jibing Gong, Yao Wan, Ye Liu, Xuewen Li, Yi Zhao, Cheng Wang, Yuting Lin, Xiaohan Fang, Wenzheng Feng, Jingyi Zhang, et al. Reinforced moocs concept recommendation in heterogeneous information networks. *ACM Transactions on the Web*, 17(3):1–27, 2023.

[Gu *et al.*, 2024] Hengnian Gu, Zhiyi Duan, Pan Xie, and Dongdai Zhou. Modeling balanced explicit and implicit relations with contrastive learning for knowledge concept recommendation in moocs. In *Proceedings of the ACM on Web Conference 2024*, pages 3712–3722, 2024.

[He *et al.*, 2020] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, pages 639–648, 2020.

[He *et al.*, 2025] Yingzhi He, Xiaohao Liu, An Zhang, Yunshan Ma, and Tat-Seng Chua. Llm2rec: Large language models are powerful embedding models for sequential recommendation. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 896–907, 2025.

[Hu *et al.*, 2025] Zheng Hu, Zhe Li, Ziyun Jiao, Satoshi Nakagawa, Jiawen Deng, Shimin Cai, Tao Zhou, and Fuji Ren. Bridging the user-side knowledge gap in knowledge-aware recommendations with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 11799–11807, 2025.

[Huang *et al.*, 2019] Zhenya Huang, Qi Liu, Chengxiang Zhai, Yu Yin, Enhong Chen, Weibo Gao, and Guoping Hu. Exploring multi-objective exercise recommendations in online education systems. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1261–1270, 2019.

[Jiang *et al.*, 2023] Lu Jiang, Kunpeng Liu, Yibin Wang, Dongjie Wang, Pengyang Wang, Yanjie Fu, and Minghao Yin. Reinforced explainable knowledge concept recommendation in moocs. *ACM Transactions on Intelligent Systems and Technology*, 14(3):1–20, 2023.

[Jing and Tang, 2017] Xia Jing and Jie Tang. Guess you like: course recommendation in moocs. In *Proceedings of the international conference on web intelligence*, pages 783–789, 2017.

[Kim *et al.*, 2025] Jieyong Kim, Hyunseo Kim, Hyunjin Cho, SeongKu Kang, Buru Chang, Jinyoung Yeo, and Dongha Lee. Review-driven personalized preference reasoning with large language models for recommendation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2025, Padua, Italy, July 13-18, 2025*, pages 1697–1706. ACM, 2025.

[Li *et al.*, 2024] Qingyao Li, Wei Xia, Kounianhua Du, Qiji Zhang, Weinan Zhang, Ruiming Tang, and Yong Yu. Learning structure and knowledge aware representation with large language models for concept recommendation. *arXiv preprint arXiv:2405.12442*, 2024.

[Liang *et al.*, 2023] Zibo Liang, Lan Mu, Jie Chen, and Qing Xie. Graph path fusion and reinforcement reasoning for recommendation in moocs. *Education and Information Technologies*, 28(1):525–545, 2023.

[Liang *et al.*, 2025] Qingqing Liang, Xuesong Lu, Chunyang Wang, Weining Qian, and Aoying Zhou. Dynamic integration of preference and knowledge status for knowledge concept recommendation. *Neurocomputing*, page 131786, 2025.

[Liu *et al.*, 2023] Fei Liu, Xuegang Hu, Shuochen Liu, Chenyang Bu, and Le Wu. Meta multi-agent exercise recommendation: A game application perspective. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1441–1452, 2023.

[Liu *et al.*, 2024] Qidong Liu, Xian Wu, Yejing Wang, Zijian Zhang, Feng Tian, Yefeng Zheng, and Xiangyu Zhao. Llm-esr: Large language models enhancement for long-tailed sequential recommendation. *Advances in Neural Information Processing Systems*, 37:26701–26727, 2024.

[Ozyurt *et al.*, 2025] Yilmazcan Ozyurt, Tunaberk Almaci, Stefan Feuerriegel, and Mrinmaya Sachan. Personalized exercise recommendation with semantically-grounded knowledge tracing. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[Ren *et al.*, 2024] Xubin Ren, Wei Wei, Lianghao Xia, Lixin Su, Suqi Cheng, Junfeng Wang, Dawei Yin, and Chao Huang. Representation learning with large language models for recommendation. In *Proceedings of the ACM web conference 2024*, pages 3464–3475, 2024.

[Schunk *et al.*, 2014] Dale H. Schunk, Judith L. Meece, and Paul R. Pintrich. *Motivation in Education: Theory, Research, and Applications*. Pearson, Boston, MA, 4th edition, 2014.

[Shan *et al.*, 2025] Rong Shan, Jianghao Lin, Chenxu Zhu, Bo Chen, Menghui Zhu, Kangning Zhang, Jieming Zhu, Ruiming Tang, Yong Yu, and Weinan Zhang. An automatic graph construction framework based on large language models for recommendation. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 4806–4817, 2025.

[Wang *et al.*, 2023] Hangyu Wang, Ting Long, Liang Yin, Weinan Zhang, Wei Xia, Qichen Hong, Dingyin Xia, Ruiming Tang, and Yong Yu. Gmocat: A graph-enhanced multi-objective method for computerized adaptive testing. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2279–2289, 2023.

[Wei *et al.*, 2024] Wei Wei, Xubin Ren, Jiabin Tang, Qinyong Wang, Lixin Su, Suqi Cheng, Junfeng Wang, Dawei Yin, and Chao Huang. Llmrec: Large language models with graph augmentation for recommendation. In *Proceedings of the 17th ACM international conference on web search and data mining*, pages 806–815, 2024.

[Yang *et al.*, 2023] Yuhao Yang, Chao Huang, Lianghao Xia, and Chunzhen Huang. Knowledge graph self-supervised rationalization for recommendation. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, pages 3046–3056, 2023.

[Yu *et al.*, 2020] Jifan Yu, Gan Luo, Tong Xiao, Qingyang Zhong, Yuquan Wang, Wenzheng Feng, Junyi Luo, Chenyu Wang, Lei Hou, Juanzi Li, et al. Mooccube: A large-scale data repository for nlp applications in moocs. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 3135–3142, 2020.

[Yu *et al.*, 2021] Jifan Yu, Yuquan Wang, Qingyang Zhong, Gan Luo, Yiming Mao, Kai Sun, Wenzheng Feng, Wei Xu, Shulin Cao, Kaisheng Zeng, et al. Mooccubex: a large knowledge-centered repository for adaptive learning in moocs. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 4643–4652, 2021.

[Zhang *et al.*, 2019] Jing Zhang, Bowen Hao, Bo Chen, Cuiping Li, Hong Chen, and Jimeng Sun. Hierarchical reinforcement learning for course recommendation in moocs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 435–442, 2019.

[Zhang *et al.*, 2023] Huanyu Zhang, Xiaoxuan Shen, Baolin Yi, Wei Wang, and Yong Feng. Kgan: Knowledge grouping aggregation network for course recommendation in moocs. *Expert Systems with Applications*, 211:118344, 2023.

[Zhang *et al.*, 2024] Dan Zhang, Yangliao Geng, Wenwen Gong, Zhongang Qi, Zhiyu Chen, Xing Tang, Ying Shan, Yuxiao Dong, and Jie Tang. Recdcl: Dual contrastive learning for recommendation. In *Proceedings of the ACM Web Conference 2024*, pages 3655–3666, 2024.

[Zhao *et al.*, 2020] Zhongying Zhao, Yonghao Yang, Chao Li, and Liqiang Nie. Guessuneed: Recommending courses via neural attention network and course prerequisite relation embeddings. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 16(4):1–17, 2020.