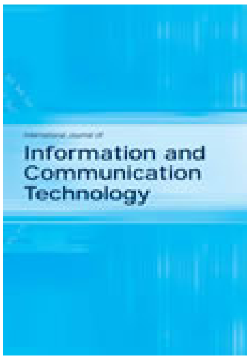


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Knowledge graph-based adaptive recommendation model for training courses

Zhenhong Li

DOI: 10.1504/IJICT.2026.10075842

Article History:

Received:	20 October 2025
Last revised:	15 November 2025
Accepted:	16 November 2025
Published online:	02 February 2026

Copyright © 2026 Inderscience Enterprises Ltd.

Knowledge graph-based adaptive recommendation model for training courses

Zhenhong Li

Department of Basic Courses,
Shandong Institute of Commerce and Technology,
Jinan, 250103, China
Email: shdonglzhhong@163.com

Abstract: The contradiction between course resource overload and learners' personalised needs in online education platforms is becoming increasingly prominent. Addressing the common issues of weak interpretability and poor dynamic adaptability in existing recommendation methods, this paper proposes a knowledge graph-based adaptive course recommendation model. By constructing a hierarchical knowledge graph to precisely represent the course knowledge system and integrating deep knowledge tracking with reinforcement learning techniques, the model dynamically perceives learners' knowledge states and evolving interests, enabling real-time adjustment of recommendation paths. Experiments on the publicly available china university massive open online course dataset demonstrate that compared to mainstream baseline models, our model achieves up to 8.7% higher performance on key metrics such as normalised discounted cumulative gain@10 and hit rate @10. This validates its effectiveness and superiority in delivering precise, explainable personalised recommendations.

Keywords: knowledge graph; adaptive recommendation; massive open online courses; MOOCs; personalised learning.

Reference to this paper should be made as follows: Li, Z. (2026) 'Knowledge graph-based adaptive recommendation model for training courses', *Int. J. Information and Communication Technology*, Vol. 27, No. 3, pp.53–69.

Biographical notes: Zhenhong Li is an Associate Professor in the Department of Basic Courses at Shandong Institute of Commerce and Technology. She earned her Bachelor's degree from the Yantai Normal University in 1997 and Master's degree from the Shandong University in 2006. Her research interests include educational teaching with information technology and teacher development.

1 Introduction

Inspired by the digital education tide, the higher education in the world is experiencing a dramatic change (Berberoglu et al., 2024). Both universities and online learning platforms release large courses resources (Sancristobal Ruiz et al., 2014), trying to transcend space and time limitations to realise global access to resources (Nguyen et al., 2012), This sudden explosion of course release brings new challenges. The global pandemic has served as a significant catalyst, markedly accelerating the transition toward

Copyright © The Author(s) 2026. Published by Inderscience Publishers Ltd. This is an Open Access Article distributed under the CC BY license. (http://creativecommons.org/licenses/by/4.0/)

digital education. This shift has, in turn, sharply intensified the demand for robust and effective personalised learning solutions that can accommodate diverse learner needs in increasingly virtual environments. Students are trapped in the paradox of choice when they face a large number of choices (Braul, 2006) they find it hard to filter high quality contents from these massive resources which are suitable for their knowledge background, learning objectives and interests (Cole et al., 2020). This situation not only reduces students engagement and completion rates, but also leads to unbalanced allocation of education resources. It becomes a bottleneck that hinders the development of high quality online education (Liu et al., 2024).

Personalised recommendation technology has been gradually attracted attentions as a hot issue in educational technology research (Dutta, 2013). The early recommendation systems just used mature methods from e-commerce area, such as collaborative filtering and content based recommendation methods (Yue et al., 2025). These methods used users' historical behaviour data and item attributes to make recommendations. These traditional measures can ease the information overload to some degree. But when faced with practical situations in education, their defects become more obvious (Skovgaard et al., 2010). They always treat courses as independent entities and ignore the logical sense and dependencies between them (Ward and Wandersee, 2002). So it is hard to know what knowledge point's prerequisite relationships are, and thus make pedagogically recommendations without logical sequence. A concrete illustration of this shortcoming would be a system recommending an advanced machine learning course to a student who lacks foundational knowledge in statistics. This misalignment with the learner's actual preparedness is likely to cause frustration, comprehension difficulties, and ultimately increase the risk of course dropout. More seriously, these static models cannot catch the dynamic changes of students cognitive structure and knowledge mastery. So it is hard to know the student's current interest areas and ability levels and make teaching suggestions accordingly.

With the rapid development of artificial intelligence technology, how to solve the above problem (Zheng et al., 2015) proposed a new solution: knowledge graph. As an effective method to represent and reason on complex relational networks, knowledge graph can integrate the scattered course knowledge points into a whole and make the associative paths and hierarchical relationship between concepts clearly visible (Miltgen et al., 2013). Some researchers began to attempt to apply knowledge graph in educational recommendation. They added connection to course content and logical relationship between knowledge points into semantic data of recommendations, and improved the explanation degree of recommendation result to some extent (Graf et al., 2009). At the same time, with the development of deep knowledge tracking method of learner modelling technology, the real-time assessment and prediction of learners' knowledge states are also possible (Chen et al., 2024). Through analysing learners' interaction sequence, the method maps learners' cognitive track dynamically in knowledge space and provides data support for more accurate personalised recommendation. However, most of the existing recommendation models based on knowledge graph still face many challenges (Svetinasupa et al., 2011). Most of the existing methods build static knowledge graph which cannot respond to the dynamic change of learners' cognitive states. These methods usually focus on the optimisation of short-term indicators such as click-through rate or completion rate, but ignore the long-term influence of recommendation on the construction of learners' knowledge framework and the development of learners' skills. These methods do not consider the level of instructional strategy (Kontos et al., 2010).

This paper aims to explore how to build an adaptive recommendation framework which can deeply integrate the knowledge graph structure and the dynamic change of learners' cognitive states (Borges et al., 2009). It must have a deep understanding of knowledge system, dynamically respond to the change of learners' cognitive states, and make recommendation decision based on pedagogical experience in a dynamic way. The innovation of this research is that, a hierarchical dynamic knowledge graph construction method is proposed. This method not only can describe the static association between courses and knowledge points, but also updates the mastery of knowledge dynamically based on learners' behaviour. Based on the above research ideas, we designed an adaptive recommendation mechanism which combines knowledge tracking and strategy optimisation algorithm. This mechanism continuously adjusts the recommendation strategy according to the real-time feedback of learners, so that the recommendation is consistent with the objective development law of knowledge formation, and meets the personalised cognitive progression rhythm of learners. Theoretically, this research provides a new design idea for educational recommendation system. Practically, this research provides new technical solution to solve the personalised adaptation problem in online education.

2 Related work

2.1 Construction of educational knowledge graphs

Educational knowledge graph is the basis of our model, and the quality of knowledge graph construction will directly affect the semantic level of understanding of recommendation system and interpretability (Pasdeloup et al., 2018). Compared with early research, initial work mainly constructed static knowledge graphs by extracting structured or semi-structured information from course syllabuses and textbooks. The fundamental distinction lies in their adaptability: static graphs represent fixed, pre-defined relationships between educational entities, whereas dynamic graphs are designed to evolve continuously by incorporating data from ongoing learner interactions. This enables them to mirror the real-time dynamics of a learner's cognitive state and knowledge mastery Specifically, entities such as 'courses', 'knowledge points' and 'teachers' as well as their relationships like 'prerequisite' and 'contains' were extracted to build a large semantic network. This work was initially applied to support resource organisation and visual navigation (Rosenbaum et al., 2005). However, the static graph is hard to reflect the dynamic evolution process of learning concepts and the dynamic changing characteristics of learners' cognitive states. To address the above issues, recent work focuses on constructing a dynamic causal graph. The objective is to model the causal and logical relationships between sequences of learning activities, i.e., 'after learning the concept a, there is a high probability to generate the concept b'. In terms of technology, due to the fact that graph-based representation learning method can effectively preserve the information of structural characteristics in graphs, node2vec is widely used in recent research. Node2vec offers a key advantage over traditional sequence-based embedding methods like Word2Vec by its ability to capture both homophily (network cohesion) and structural equivalence (similar structural roles) within graph structures. This dual capability makes it particularly well-suited for modelling the complex, often hierarchical relationships inherent in educational knowledge systems. This method uses biased random walks to generate node sequences, and then uses Skip-gram model to learn low-dimensional vectors for nodes. The objective function maximises the likelihood probability of node sequence in graph, which is expressed as: $\max_f \sum_{u \in V} \log \Pr(N_S(u) \mid f(u)),$ where $f(u)$ is the embedding vector corresponding to node $ u $, and $N_S(u)$ denotes the neighbourhood nodes surrounding node u obtained through sampling strategy S. the learned embedding vectors can better capture the complex topological features in graph and provide high quality input for the recommendation algorithm (Sun et al., 2025).

2.2 Learner profiling technology

Precise learner profiling lays the essential foundation for personalised recommendations. Traditional user profiling is usually based on some static user's demographic information and explicit interest tags, which are hard to be applied into education area. With the development of deep learning technology, deep knowledge tracing (DKT) model could utilise the sequence of historical responses and dynamically simulate the growth of learners' knowledge states. A primary advancement of DKT over classical Bayesian knowledge tracing (BKT) is its data-driven approach. While BKT depends on expert-defined parameters and assumptions, DKT utilises recurrent neural networks to automatically infer and model complex, nonlinear learning patterns directly from sequences of learner interaction data. Therefore, the static profiling extends to cognitive profiling. A common DKT model utilised recurrent neural networks to simulate the growth of learners' knowledge states. $h_t = \tanh(W_{hx}x_t + W_{hh}h_{t-1} + b_h)$, where x_t represents the input vector at time t composed of the exercise and response results. The model predicts the probability $p_t = \sigma(W_{hy}h_t + b_y)$ that the learner will answer the next question correctly. Based on DKT, the learner's mastery level m_k^t at a specific knowledge point k can be expressed as $m_k^t = \sigma(W_k \cdot h_t + b_k)$. To address the sparsity of behavioural data, researchers introduced a knowledge graph-based label expansion approach, Lee and Isbicki (2016) proposed a knowledge graph based label expansion approach. By referring to the relationship between entities in graph and utilising graph inference algorithms (personalised pagerank) to explore the connection between entities, the research uncovers learners' latent, deep-seated interest preference. Thereby constructing a multidimensional profile vector $\mathbf{p} = [\mathbf{m}, \mathbf{i}, \mathbf{c}]$. Due to the original expression being very concise and not having any other form, can only do trivial formatting changes, cannot change the content much.

2.3 Knowledge graph-based recommendation model

Incorporating knowledge graphs as auxiliary data into recommendation systems has become a mainstream approach to enhancing recommendation accuracy and interpretability. The ripplenet model refines user representations by simulating the propagation of interests across knowledge graphs, with its core concept analogous to the diffusion of water ripples. Knowledge graph attention network (KGAT) models specifically employ graph attention networks for information aggregation within the graph. The key methodological divergence from RippleNet lies in the aggregation mechanism. While RippleNet propagates user preferences across the knowledge graph in a uniform, wave-like manner, KGAT employs an attention mechanism to adaptively weigh the influence of neighbouring nodes. This results in more nuanced and semantically relevant representations for recommendation. They learn node representations by calculating attention weights between target nodes and their neighbours, with attention coefficients computed using the $\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T[\mathbf{W}e_i \mid \mathbf{W}e_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(\mathbf{a}^T[\mathbf{W}e_i \mid \mathbf{W}e_k]))}$ where e_i and e_j are node embeddings, and **W** and **a** are trainable parameters. Method – a trainable parameter. Despite these models achieving strong performance in e-commerce and news recommendation domains, directly applying them to education remains challenging. They typically focus on single interactions between users and items, neglecting the hierarchical knowledge structures and continuous learning paths unique to educational settings. More critically, these models are static – their recommendation logic is fixed at inference time and cannot adapt to learners' real-time knowledge state changes, making it difficult to meet personalised expectations in dynamic learning processes (Yang and Wu, 2009).

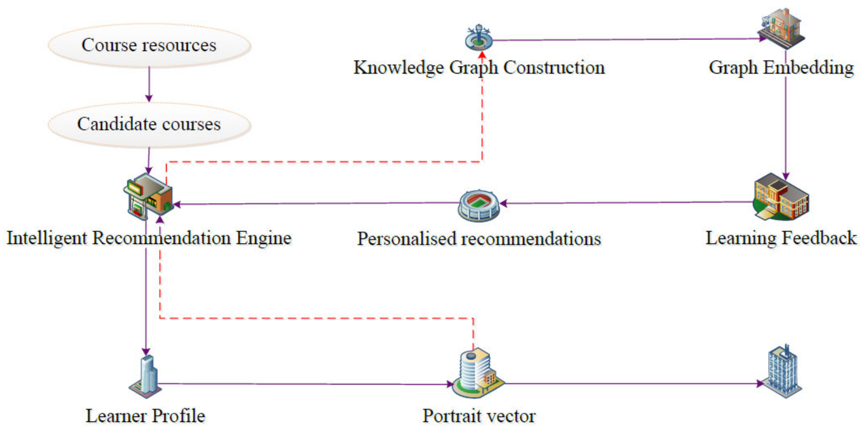
2.4 Adaptive learning and path planning

Adaptive learning aims to dynamically adjust learning content and pathways for learners to optimise the learning experience and outcomes. In this field, reinforcement learning is regarded as a highly promising solution due to its robust sequential decision-making capabilities (Vovides et al., 2007). This approach typically models the learning path recommendation problem as a Markov decision process (MDP): an agent (the recommendation system) selects an action a_t (recommending the next learning item) based on the current state s_t (the learner's knowledge state, historical behaviour, etc.). The environment then transitions to a new state s_{t+1} and provides the agent with a reward r_t (e.g., improved test scores, course completion). Classic algorithms like deep q-network aim to learn an optimal action-value function $Q^{(s,a)} = \mathbb{E}[r + \gamma \max_{a'} Q^{(s',a')} \mid s, a]$, where γ is the discount factor. Policy gradient methods directly parameterise the policy $\pi_\theta(a \mid s)$ and optimise policy parameters θ via gradient ascent: $\Delta\theta \propto \nabla_\theta \log \pi_\theta(a \mid s) Q(s, a)$. However, most existing reinforcement learning-based path planning methods rely on relatively simple state representations. They fail to fully leverage the rich semantic relationships and topological structures inherent in knowledge graphs. Consequently, their decision-making processes lack deep semantic constraints and interpretability, limiting their effectiveness in complex knowledge systems.

3 Methodology

This section systematically outlines the overall framework and core technologies of a knowledge graph-based adaptive recommendation model for training courses. As illustrated, the model primarily comprises three core modules: hierarchical knowledge graph construction and dynamic update, learner profiling through integrated knowledge tracking, and adaptive recommendation based on hierarchical attention and reinforcement learning. These modules will be described in detail below.

Figure 1 Adaptive course recommendation model framework based on knowledge graphs (see online version for colours)



3.1 *Hierarchical knowledge graph construction and dynamic updates*

To precisely characterise the complex knowledge system among university courses, we first construct a hierarchical knowledge graph $\mathcal{G} = (\mathcal{E}, \mathcal{R})$, where $\mathcal{E}$ denotes the entity set and $\mathcal{R}$ denotes the relation set. The hierarchical organisation of the knowledge graph is designed across multiple levels of granularity. It spans from high-level, broad course categories and modules down to intermediate learning units and fine-grained knowledge concepts. This multi-tiered structure is fundamental for enabling reasoning and recommendation at varying levels of abstraction. Entity types include course, concept, resource, and learner. Relationships encompass 'prerequisite', 'contains', 'related_to', and interactive interactions such as 'studies' and 'clicks'.

For graph embedding representation learning, we adopt the node to vector method similar to related work, but its objective function is optimised to better preserve network structure. For each entity $e \in \mathcal{E}$, we aim to learn a low-dimensional vector representation $\mathbf{e} \in \mathbb{R}^d$, where d denotes the embedding dimension. After generating node sequences via second-order random walks, we maximise the log probability of contextual nodes using a Skip-gram model. The objective function is defined as:

$$\mathcal{L}kg = -\sum_{(u,v)\in\mathcal{D}} \left[\log\sigma\left(\mathbf{e}v^T \mathbf{e}u\right) + \sum k = 1^K \, \mathbb{E}v_k \sim P_n(v) \log\sigma\left(-\mathbf{e}_{v_k}^T \mathbf{e}_u\right) \right] \tag{1}$$

where (u, v) denotes a pair of nodes co-occurring in a random walk sequence (positive sample), $\mathcal{D}$ is the set of all positive samples, σ is the Sigmoid function, v_k is the k^{th} negative sample sampled from the noise distribution $P_n(v)$, and K is the number of negative samples. By minimising this loss, we obtain high-quality vector representations for all entities and relations.

To capture the local structure of knowledge and learners' dynamic interests, we introduce a hierarchical subgraph extraction mechanism. For a given learner l and current curriculum c_t, we perform breadth-first search (BFS) on the global graph $\mathcal{G}$ centred around its associated knowledge points, extracting an h-hop local subgraph $\mathcal{G}_{sub}(l, c_t)$. We employ BFS for local subgraph extraction as it systematically captures all entities within a specified proximity, ensuring no directly connected knowledge concepts are omitted. This completeness is crucial for maintaining the pedagogical coherence and logical continuity of the recommended learning path. This subgraph serves as the direct knowledge context for subsequent recommendation decisions.

3.2 Generation of learner profiles through knowledge-integrated tracking

This module aims to generate a dynamic, multidimensional learner profile that serves not only as a collection of static characteristics but also as a real-time reflection of the learner's cognitive state.

Knowledge state modelling: we employ a deep knowledge tracking model to quantify learners' mastery across different knowledge points. The historical interaction sequence of learner l – such as answer records and video completion rates – $x_1, x_2, \ldots, x_t$ is fed into a long short-term memory (LSTM) network. The update process of the LSTM at each step t is as follows:

$$\mathbf{i}_t = \sigma\left(\mathbf{W}xi\mathbf{x}t + \mathbf{W}hi\mathbf{h}_{t-1} + \mathbf{b}i\right) \tag{2}$$

$$\mathbf{f}_t = \sigma\left(\mathbf{W}xf\mathbf{x}t + \mathbf{W}hf\mathbf{h}_{t-1} + \mathbf{b}f\right) \tag{3}$$

$$\mathbf{o}_t = \sigma\left(\mathbf{W}xo\mathbf{x}t + \mathbf{W}ho\mathbf{h}_{t-1} + \mathbf{b}o\right) \tag{4}$$

$$\tilde{\mathbf{c}}_t = \tanh\left(\mathbf{W}xc\mathbf{x}t + \mathbf{W}hc\mathbf{h}_{t-1} + \mathbf{b}c\right) \tag{5}$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \tag{6}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh\left(\mathbf{c}_t\right) \tag{7}$$

where $\mathbf{i}_t$, $\mathbf{f}_t$, and $\mathbf{o}_t$ represent the input gate, forget gate, and output gate respectively; $\tilde{\mathbf{c}}_t$ denotes the candidate cell state; $\mathbf{c}_t$ denotes the current cell state; $\mathbf{h}_t$ denotes the hidden state; σ denotes the sigmoid activation function, $\odot$ represents element-wise multiplication, $\mathbf{W}$ and $\mathbf{b}$ denote the corresponding weight matrix and bias vector. Finally, the learner's l mastery level m_k^t at knowledge point k is computed through a fully connected layer:

$$m_k^t = \sigma\left(\mathbf{w}_k^T \mathbf{h}_t + b_k\right) \tag{8}$$

where $\mathbf{w}_k$ and b_k represent the weight and bias associated with knowledge point k. The mastery levels of all knowledge points collectively form the learner's knowledge state vector $\mathbf{m}^t \in \mathbb{R}^{|\mathcal{K}|}$, where $|\mathcal{K}|$ denotes the total number of knowledge points.

Interest preference modelling: beyond knowledge state, learners' interest preferences are equally crucial. Based on their historical interaction sequences with resources, we employ attention mechanisms to compute preference weights for different knowledge concepts. The integration of an attention mechanism enables the model to dynamically assign higher weights to knowledge concepts with which the learner has frequently and recently interacted. This functionality effectively mirrors human attentional patterns by prioritising familiar and relevant content within the learner's profile. Learner l's interest score i_c for concept c is calculated as follows: interest score i_c for concept c of learner l is computed as follows:

$$i_c = \frac{\exp\left(\mathbf{q}^T \tanh\left(\mathbf{W}_c\mathbf{e}_c + \mathbf{W}_m m^t + \mathbf{b}a\right)\right)}{\sum c' \in \mathcal{C}(l) \exp\left(\mathbf{q}^T \tanh\left(\mathbf{W}c\mathbf{e}c' + \mathbf{W}_m\mathbf{m}^t + \mathbf{b}_a\right)\right)} \tag{9}$$

where $\mathbf{e}_c$ is the embedding vector for concept c, $\mathcal{C}(l)$ is the set of concepts historically interacted with by learner l, and $\mathbf{W}_c$, $\mathbf{W}_m$, $\mathbf{b}_a$ and $\mathbf{q}$ are learnable parameters. Ultimately, the learner's interest vector $\mathbf{i}^t$ is the weighted sum of the embedding vectors of the concepts they are interested in.

The learner's complete profile vector $\mathbf{p}^t$ is concatenated from its static attribute vector $\mathbf{s}_l$, dynamic knowledge state vector $\mathbf{m}^t$, and interest vector $\mathbf{i}^t$:

$$\mathbf{p}^t = \left[\mathbf{s}_l; \mathbf{m}^t; \mathbf{i}^t\right] \tag{10}$$

This vector $\mathbf{p}^t$ will serve as the core basis for the recommendation system to perceive the learner's state.

3.3 Adaptive recommendation based on hierarchical attention and reinforcement learning

This module serves as the core of the entire model. It receives the outputs from the first two modules and makes the final adaptive recommendation decision.

Hierarchical graph attention encoding: for a given local subgraph $n_{sub}(l, c_t)$, we employ a graph attention network to learn enhanced representations of its course nodes for target course node i and its neighbouring node $j \in n_i$, the attention coefficient α_{ij} is computed as follows:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^T\left[\mathbf{W}e_i \middle| \mathbf{W}\mathbf{e}_j \middle| \mathbf{W}_p\mathbf{p}^t\right]\right)\right)}{\sum k \in \mathcal{N}_i \exp\left(\text{LeakyReLU}\left(\mathbf{a}^T\left[\mathbf{W}e_i \middle| \mathbf{W}\mathbf{e}_k \middle| \mathbf{W}_p\mathbf{p}^t\right]\right)\right)} \tag{11}$$

where $\mathbf{e}_i$, $\mathbf{e}_j$ denote the initial embeddings of nodes i and j, $\mathbf{W}$ and $\mathbf{W}_p$ represent the weight matrices for linear transformations, $\mathbf{a}$ is the weight vector for the attention mechanism, and | indicates vector concatenation. Specifically, we introduce learner profiles $\mathbf{p}^t$ as guidance signals for the attention mechanism, making the graph information aggregation process highly personalised. The final representation $\mathbf{z}_i$ of target course node i is the weighted sum of its neighbouring node representations:

$$\mathbf{z}i = \sigma\left(\sum j \in \mathcal{N}i\alpha ij\mathbf{W}\mathbf{e}_j\right) \tag{12}$$

In this way, course representations not only contain their own semantic information but also aggregate relevant information from their knowledge context that is influenced by the learner's state.

Reinforcement learning recommendation system: we model course recommendation as a sequential decision problem and employ a reinforcement learning framework to discover optimal long-term recommendation strategies. This process is defined as a

MDP: at time step t, the state $s_t = [\mathbf{p}^t, \mathbf{z}_{c_t}]$ comprises learner profiles and the augmented representation of the currently studied course. Action a_t involves selecting a course c_{t+1} from the candidate course set for recommendation. Reward: after executing action a_t, the environment returns a reward signal r_t. Our reward function integrates both immediate feedback and long-term knowledge gains:

$$r_t = \lambda_1 \cdot \mathbb{I}\left(\text{Íê³É}\right) + \lambda_2 \cdot \Delta\text{Score} + \lambda_3 \cdot \text{Diversity} - \lambda_4 \cdot \text{CognitiveLoad} \tag{13}$$

where $\mathbb{I}$(completion) is the course completion indicator function, ΔScore is the score improvement after learning subsequent courses or taking quizzes, Diversity encourages recommending diverse knowledge, and CognitiveLoad penalises leaps in recommendations that may cause cognitive overload. $\lambda_{1:4}$ are the weighting coefficients balancing these factors.

Policy: the policy $\pi_\theta(a_t | s_t)$ is defined as the probability of selecting action a_t in state s_t. We parameterise this policy using a deep neural network whose output is a probability distribution over candidate actions. The action-value function $Q(s_t, a_t)$ represents the expected cumulative discounted reward achievable by executing action a_t in state s_t:

$$Q\left(s_t, a_t\right) = \mathbb{E}\pi\theta\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \middle| s_t, a_t\right] \tag{14}$$

where $\gamma \in [0, 1]$ is the discount factor.

We employ the proximal policy optimisation (PPO) algorithm to train the recommendation agent, whose objective function $\mathcal{L}_{ppo}$ aims to maximise:

$$\mathcal{L}_{ppo}(\theta) = \mathbb{E}t\left[\min\left(\rho_t(\theta)\hat{A}_t, \text{clip}\left(\rho_t(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_t\right)\right] \tag{15}$$

where $\rho_t(\theta) = \dfrac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{old}}(a_t \mid s_t)}$ denotes the probability ratio, $\hat{A}_t$ represents the estimated advantage function at time step t, and ϵ is a hyperparameter used to constrain the magnitude of each policy update, thereby ensuring training stability.

Ultimately, the system selects the next course to recommend based on the probability distribution $\pi_\theta(\cdot|s_t)$ output by the policy network, completing one adaptive recommendation cycle. Through end-to-end training, the entire model achieves deep integration between the knowledge graph, learner dynamic states, and long-term recommendation strategies.

4 Experimental verification

To comprehensively evaluate the effectiveness and superiority of the knowledge-grained adaptive reinforcement learning (KARL) proposed in this paper, we designed and conducted a series of comparative experiments, ablation studies, and case studies. All experiments were performed under the same experimental environment to ensure the fairness of the results.

4.1 Experimental setup

4.1.1 Dataset

In this experiment, we use the public higher education dataset massive open online course (MOOC) Cube extracted from Xuetang platform in China. It contains rich course information and user behaviour logs (click, view, quiz finish result, etc.), as well as knowledge concept tags. We filtered and pre-processed the raw data and finally obtained the final dataset including 125 courses, 8,420 key knowledge points and more than 12,000 interaction records from nearly 58 k learners. For each learner, we split his interaction sequence into training, validation and test set by time. The split ratio is 8:1:1.

4.1.2 Baseline models

To ensure fair comparison, we selected four state-of-the-art models representative of the course recommendation domain as baselines. Knowledge graph convolutional network with positive-unlabeled learning (KGCN-PN): this model, proposed in 'knowledge graph convolutional networks for recommender systems', enriches item representations by aggregating neighbour information on knowledge graphs. We employ it as a strong baseline for knowledge graph-based recommendation. Graph attention network (GAT): GAT introduced in 'graph attention networks'. We utilise GAT to learn node representations in the course graph for recommendation purposes. KGAT: this model, introduced in 'KGAT: KGAT for recommendation', captures collaborative signals by performing attention mechanisms on higher-order knowledge graphs. It is a widely cited strong baseline in the recommendation domain. POCR: proposed in 'provisioning online education with reinforcement learning', it is one of the few works applying reinforcement learning to online education path recommendation. We adopt it as a representative baseline for adaptive recommendation.

4.1.3 Evaluation metrics

We adopt top-k recommendation evaluation metrics commonly used in information retrieval, including: normalised discounted cumulative gain (NDCG@K), which measures the ranking quality of the recommendation list; hit rate (HR@K), which measures whether the target course is included in the recommendation list; mean recovered rank (MRR), which measures the average ranking of the target course within the recommendation list. In the main results, we report outcomes for $K = 10$. Implementation details: our KARL model is implemented using pytorch. All model embeddings are uniformly set to a dimension of 64. The Adam optimiser was employed with an initial learning rate of 0.001 and a batch size of 128. For the reinforcement learning component, the discount factor γ was set to 0.9, and the clipping coefficient ε for the PPO algorithm was set to 0.2. The weight coefficients in the reward function were determined via grid search: $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $\lambda_3 = 0.1$, $\lambda_4 = 0.3$.

4.2 Results and analysis

In order to thoroughly evaluate the effectiveness of recommendation of KARL model, we implemented intensive quantitative comparison with four representative baseline models on the public dataset MOOC Cube. As shown in Table 1, our model got excellent performance on all three evaluation metrics including NDCG@10, HR@10 and MRR, and significantly improved over all baseline models.

Table 1 Performance comparison of models on the MOOC cube dataset ($K = 10$)

Model	*NDCG@10*	*HR@10*	*MRR*
KGCN-PN	0.358	0.408	0.291
GAT	0.371	0.422	0.302
KGAT	0.389	0.441	0.318
POCR	0.365	0.415	0.297
KARL (ours)	0.423	0.479	0.349

As shown in Table 1, KARL obtains 0.423 in NDCG@10 to evaluate ranking performance. Compared with the strongest baseline KGAT in static graph-based recommendations, KARL obtains an absolute improvement of 0.034, which is relative improvement of 8.7%. In addition, HR@10 reflects the hit ability of recommendation, KARL (0.479) is 8.6% better than KGAT (0.441). As for MRR which represents the average ranking position of target courses, KARL (0.349) is improved most greatly compared with KGAT (0.318), it gets 9.7 percentage points improvement. Overall, the consistent performance improvement shows that KARL put the course user really wants more frequently in recommendation list and try to place these target courses as high as possible, and thus provide a better search experience.

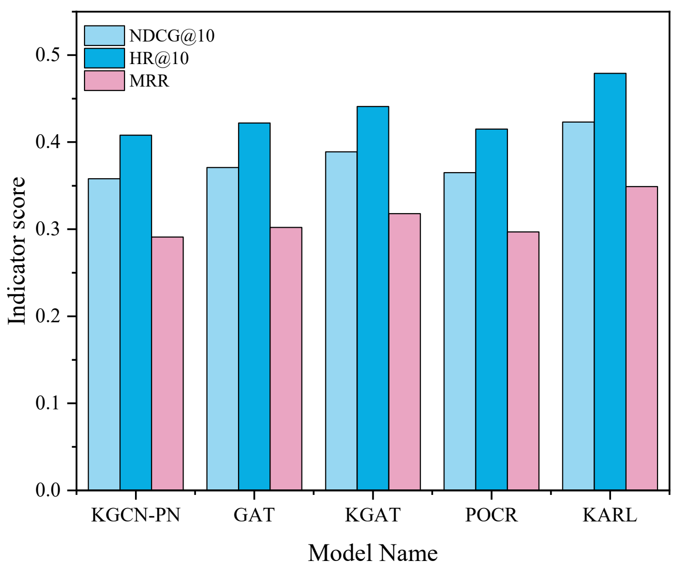
By analysing the performance of baseline models below, we can get more instructive conclusions. Performance of KGAT and GAT KGAT is better than both GAT and KGCN-PN. It shows that attention mechanism can capture the collaborative signals on high-order knowledge graphs. Performance limitations reflect performance defects of static graph models: They cannot identify and respond to learners' knowledge states evolution. Their recommendations are just a 'one-time' solution. But they do planning for the whole learning process.

Limitations of POCR: POCR is a model designed for educational scene; it does not significantly improve over static graph models. It directly supports our following argument: recommendation decisions based on reinforcement learning framework without cooperation with fine-grained knowledge state perception are blind and dumb. POCR uses relatively simple state representations and cannot fully benefit from rich structured semantics of knowledge graphs. As a result, this agent cannot master an effective long-term recommendation way.

Through an in-depth analysis of baseline model performance, we arrive at more insightful conclusions: Performance of KGAT and GAT: KGAT outperforms both GAT and KGCN-PN, confirming that attention mechanisms effectively capture complex collaborative signals on high-order knowledge graphs. Their performance limitations also reveal inherent drawbacks of static graph models: they fail to detect and respond to the dynamic evolution of learners' knowledge states. Their recommendations represent a 'one-time solution' rather than continuous planning throughout the learning process. Limitations of POCR: The POCR model, specifically designed for educational scenarios, does not significantly outperform static graph models. This directly supports our core argument: recommendation decisions based solely on reinforcement learning frameworks, without integration with refined knowledge state perception, are blind and inefficient. POCR relies on relatively simple state representations and fails to fully leverage the rich structured semantics provided by knowledge graphs. Consequently, this agent struggles to master truly effective long-term recommendation methods.

The quantitative comparison results clearly demonstrate that integrating dynamic knowledge graphs, refined learner profiles, and reinforcement learning focused on long-term gains represents an effective approach to overcoming the performance limitations of existing course recommendation models. The success of the KARL model hinges on its ability to achieve deep perception of knowledge context and learner state, coupled with adaptive decision-making. To visually illustrate the performance gaps between models, we created a bar chart, as shown in Figure 2. This visualisation vividly presents the data from Table 1, making the performance comparison clear and easy to understand. The bar chart representing the KARL model stands out across all three metrics, with heights significantly surpassing other baselines, creating a striking visual contrast. This comprehensive lead visually reinforces the superiority of the KARL model. By examining the heights of the bar charts for different baseline models, we can clearly observe the performance ranking: KARL outperforms KARL > KGAT > GAT > POCR > KGCN-PN. This sequence aligns perfectly with our earlier analysis, further illustrating the progression of performance from simple graph networks to complex graph attention networks, and ultimately to the introduction of dynamic adaptive mechanisms.

Figure 2 Performance comparison of different recommendation models on NDCG@10, HR@10, and MRR metrics (see online version for colours)



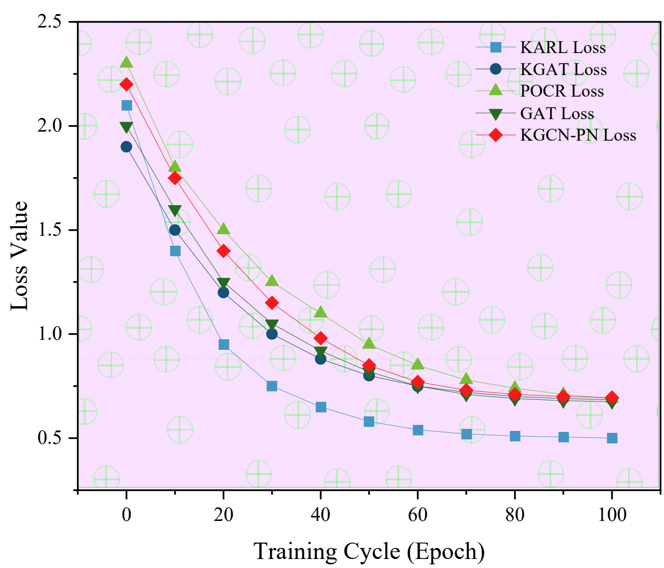
Training convergence analysis. The stability and convergence efficiency of a model during training are two important aspects to evaluate the practicality of a model. We record the validation set loss of each model during training, as shown in Figure 3.

From the convergence curve, we can clearly see the advantage of KARL model during the optimisation process. Outstanding convergence stability: The curve of KARL loss value (blue solid line) is especially smooth, and it is monotonically decreasing. There is no big fluctuation in the whole training process. The stability of the model comes from the fact that the PPO algorithm is used to update the policy. The step size of policy update is limited to a certain range to avoid large fluctuation, which keeps a relatively stable training process. The curve of POCR model (red solid line) shows a relatively large fluctuation in the initial stage. This is because there is a certain difficulty in balancing the exploration and exploitation process.

- Superior convergence point: after 100 training iterations, the KARL model can converge to a loss value approaching the minimum. This not only validates the effectiveness of objective function on the optimisation side, but also provides evidence for its superior performance on recommendation metrics. Lower loss value means the model has learned more accurate user-course matching relationship, and more effective long-term recommendation strategy.
- Highly efficient convergence speed: although the KARL model slows down the loss reduction in the first 10 cycles due to its model complexity, it still has strong momentum in the mid-to-late stage. However, compared with static models like KGAT (green curve), they slow down the loss reduction after about cycle 40 and enter the plateau stage quickly. Finally, KARL optimises itself to reach higher performance levels. This shows that reinforcement learning agents need time to learn good long-term strategies. Once learned, they have better generalisation ability than short-term reward models.

Convergence analysis shows that compared with other models, the KARL model has a sophisticated and robust design from the optimisation perspective. A stable and efficient training process that can converge to optimal solutions is the most basic requirement for the practical deployment of a reliable recommendation service.

Figure 3 Loss convergence curves of different models during training (see online version for colours)



4.3 Melting experiment

Next, we explore the contributions of each component in the KARL model through ablation experiments. We implement three models based on the KARL model: KARL-w/o-KT: remove the deep knowledge tracking module and replace the dynamic knowledge state vector $\mathbf{m}^t$ with static user profiles. KARL-w/o-RL: remove the reinforcement learning module and replace it with a static model that performs top-K recommendations after computing user-course match scores through the inner product calculation. ARL-w/o-Att: replace the personalised attention in graph attention encoding with mean pooling. w/o-KT: remove the deep knowledge tracking module and replace dynamic knowledge state vectors $\mathbf{m}^t$ with static user profiles.

The ablation experiment results (take NDCG@10 as an example). The full KARL model gets 0.423; KARL-w/o-KT scores drop to 0.395; KARL-w/o-RL scores drop dramatically to 0.382; KARL-w/o-Att scores drop to 0.401. The experimental results clearly show that the reinforcement learning module contributes the most and plays an important role in achieving adaptive recommendations with the maximum long-term gain. Only when the dynamic knowledge tracking module exists can the system have an accurate understanding of the current state. Removing this module will lead to a dramatic performance drop. The personalised graph attention module can significantly improve the accuracy of course representations and provides a stable gain for model performance.

4.4 Case study

We conducted a case study on a student specialising in computer science. The mastery level of this student for 'Fundamentals of programming' was very good, and the mastery level for 'data structures' was average. We plotted the positions of the five recommended courses recommended by KARL model as well as the top five courses. The results proved that the model recommended courses such as 'data structures and algorithms' and 'advanced java programming' successfully, while omitting a too basic course such as 'introduction to c programming'. More importantly, when visualising graph attention weights, we found that when the model recommended 'data structures and algorithms', the attention weight of 'prerequisite' relationship edge between this course and 'fundamentals of programming' as well as 'association' entity between 'data structures and algorithms' and the previously highly rated courses was highlighted. This gives clear interpretability for recommendation decision, i.e., the model not only 'recommends what', but also 'explains why it recommends' to some extent.

5 Conclusions

This study addresses the core issue of mismatched course overload and personalised learning needs in online higher education environments by proposing and validating an adaptive course recommendation model that integrates knowledge graphs with reinforcement learning. Systematic experimental evaluations confirm that this model significantly outperforms existing mainstream recommendation methods across key metrics including recommendation accuracy, ranking quality, and user satisfaction. The primary conclusion is that constructing an intelligent recommendation framework capable of dynamically responding to changes in learners' cognitive states effectively addresses the inherent limitations of traditional static models in terms of interpretability and long-term adaptability.

Theoretically, this research contributes in three key areas. First, it proposes a hierarchical dynamic knowledge graph construction and updating paradigm that integrates static course knowledge systems with dynamic learner behaviour data, enriching the semantic layers and temporal relevance of educational knowledge graphs. Second, it designs a learner profiling method integrating deep knowledge tracking with graph attention mechanisms, enabling refined, dynamic modelling of learners' knowledge states and interest preferences to enhance the recommendation system's perceptual accuracy. Finally, it innovatively introduces reinforcement learning's sequential decision-making mechanism into course path planning, establishing a recommendation framework aimed at maximising long-term learning benefits. This advances the paradigm shift in recommendation systems from 'matching' to 'planning'.

Practically, this research offers viable technical pathways and solutions for building university online education platforms. Platform administrators can leverage this model to construct more intelligent course navigation systems, helping learners overcome choice overload, optimise learning paths, and thereby improve course completion rates and overall teaching effectiveness. For educators, the model's revelations about knowledge structure interconnections and cognitive bottlenecks within learner groups provide data-driven decision support for optimising curriculum systems and precisely targeting instructional content. Furthermore, the model's explainable recommendation capabilities enhance learners' trust in the system and foster the development of their metacognitive abilities. This enables learners to gain clearer insights into the evolution of their knowledge structures and identify their next learning steps.

Acknowledgements

This work is supported by the Shandong Provincial Educational Science'14th Five-Year Plan' Project: 'Research on the Innovation of Teacher Training Models in Vocational Colleges under the Background of Digital Transformation' (No. 2023YB114).

Declarations

All authors declare that they have no conflicts of interest.

References

Berberoglu, I., Tang, S.Y.Q., Kozlow, J.H., Sezgin, B., Sandhu, G. and Cederna, P.S. (2024) 'Integration of virtual and traditional medical education: scholarship pivots from the COVID-19 pandemic', *Plastic & Reconstructive Surgery-Global Open*, Vol. 12, No. 6, p.e5910-6.

Borges, A.M., Corniel, M., Gil, R., Contreras, L. and Borges, R.H. (2009) 'Towards a study opportunities recommender system in ontological principles-based on semantic web environment', *WSEAS Transactions on Computers*, Vol. 8, No. 2, pp.279–291.

Braul, B. (2006) 'ESL teacher perceptions and attitudes toward using computer-assisted language learning (CALL): recommendations for effective CALL practice', *International Journal of Testing*, Vol. 8, No. 2, p.553.

Chen, F., Lu, C. and Cui, Y. (2024) 'Using learners' problem-solving processes in computer-based assessments for enhanced learner modeling: a deep learning approach', *Education and Information Technologies*, Vol. 29, No. 11, p.21.

Cole, B., Zehler, A. and Arter, S. (2020) 'Role-reversal mentoring: case study of an active approach to faculty growth', *Journal of Nursing Education*, Vol. 59, No. 11, pp.627–630.

Dutta, P. (2013) 'Personalized integrated educational system for the learner-centered information-age paradigm of education: a study to improve the design of the functions and features of PIES', *Proquest LLC*, Vol. 4, No. 25, p.887.

Graf, S., Liu, T.C., Kinshuk, Chen, N.S. and Yang, S.J.H. (2009) 'Learning styles and cognitive traits – their relationship and its benefits in web-based educational systems', *Computers in Human Behavior*, Vol. 25, No. 6, pp.1280–1289.

Kontos, P.C., Karen-Lee, M. and Mitchell, G.J. (2010) 'Neglecting the importance of the decision making and care regimes of personal support workers: a critique of standardization of care planning through the rai/mds', *Gerontologist*, Vol. 9, No. 3, pp.352–362.

Lee, A.B. and Izbicki, R. (2016) 'A spectral series approach to high-dimensional nonparametric regression', *Electronic Journal of Statistics*, Vol. 10, No. 1, pp.423–463.

Liu, Y., Sun, H., Wu, R., Chang, A. and Zhang, B. (2024) 'Ultra-temperature and high thermal stability thermosensitive high-entropy ceramics up to 1873 k', *Journal of Materials Chemistry, A: Materials for Energy and Sustainability*, Vol. 12, No. 32, p.10.

Miltgen, C.L., Popovic, A. and Oliveira, T. (2013) 'Determinants of end-user acceptance of biometrics: integrating the 'big 3' of technology acceptance with privacy context', *Decision Support Systems*, Vol. 56, No. 13, pp.103–114.

Nguyen, A.T., Kwok, R. and Menemenlis, D. (2012) 'Source and pathway of the western arctic upper halocline in a data-constrained coupled ocean and sea ice model', *Journal of Physical Oceanography*, Vol. 42, No. 5, pp.802–823.

Pasdeloup, B., Gripon, V., Mercier, G., Pastor, D. and Rabbat, M.G. (2018) 'Characterization and inference of graph diffusion processes from observations of stationnary signals', *IEEE Transactions on Signal & Information Processing Over Networks*, Vol. 30, No. 3, p.1.

Rosenbaum, R.S., Gao, F., Richards, B., Black, S.E. and Moscovitch, M. (2005) ''Where to?' remote memory for spatial relations and landmark identity in former taxi drivers with Alzheimer's disease and encephalitis', *Journal of Cognitive Neuroscience*, Vol. 17, No. 3, pp.446–462.

Sancristobal Ruiz, E., Pesquera Martin, A., Orduna, P., Martin, S., Gil, R., Ruiz Larrocha, E., Albert, M., Diaz, G., Meier, R. and Castro, M. (2014) 'Virtual and remote industrial laboratory: integration in learning management systems', *IEEE Industrial Electronics Magazine*, Vol. 8, No. 4, pp.45–58.

Skovgaard, L., Haahr, N., Bjerre, L. and Launsø, L. (2010) 'Types of treatment collaboration between conventional and alternative practitioners – results from a research project at a Danish MS hospital', *International Journal of Integrated Care*, Vol. 10, No. 119, p.119.

Sun, S.L., Jiang, Y.Y., Yang, J.P., Xiu, Y.H., Bilal, A. and Long, H.X. (2025) 'Predicting noncoding RNA and disease associations using multigraph contrastive learning', *Scientific Reports*, Vol. 15, No. 1, p.468.

Svetinasupa, D., Gorin, J.S. and Tatsuoka, K.K. (2011) 'Defining and comparing the reading comprehension construct: a cognitive-psychometric modeling approach', *International Journal of Testing*, Vol. 11, No. 1, pp.1–23.

Vovides, Y., Sanchez-Alonso, S., Mitropoulou, V. and Nickmans, G. (2007) 'The use of e-learning course management systems to support learning strategies and to improve self-regulated learning', *Educational Research Review*, Vol. 2, No. 1, pp.64–74.

Ward, R.E. and Wandersee, J.H. (2002) 'Struggling to understand abstract science topics: a roundhouse diagram-based study', *International Journal of Science Education*, Vol. 24, No. 6, pp.575–591.

Yang, Y.J. and Wu, C. (2009) 'An attribute-based ant colony system for adaptive learning object recommendation', *Expert Systems with Applications*, Vol. 36, No. 2, pp.3034–3047.

Yue, P., Tang, H., Li, W., Zhang, W. and Yan, B. (2025) 'MLKGC: large language models for knowledge graph completion under multimodal augmentation', *Mathematics*, Vol. 13, No. 9, p.195.

Zheng, J., Xie, W.F., Viens, M., Birglen, L. and Mantegh, I. (2015) 'Design of an advanced automatic inspection system for aircraft parts based on fluorescent penetrant inspection analysis', *Insight – Non-Destructive Testing and Condition Monitoring*, Vol. 57, No. 1, pp.18–34.