A MOOC Course Recommendation Method Integrating Reinforcement Learning and Neural-Symbolic Reasoning

Xiaolan Cao
Big Data and Artificial Intelligence College
Anhui Xinhua University
Hefei, China
E-mail: smile_lan27@163.com

Xuechen Xu
Big Data and Artificial Intelligence College
Anhui Xinhua University
Hefei, China
E-mail: 1291631351@qq.com

Jingyi Hu
Big Data and Artificial Intelligence College
Anhui Xinhua University
Hefei, China
E-mail: hujingyi@axhu.edu.cn

Shengquan Wang
Big Data and Artificial Intelligence College
Anhui Xinhua University
Hefei, China
E-mail: 3429952495@qq.com

***Abstract*—To address the issue of overloaded course resources on MOOC platforms and the poor effectiveness of personalized course recommendations for users, this paper proposes a MOOC course recommendation method that integrates reinforcement learning and neural-symbolic reasoning. The method uses neural networks to extract features of user behavior in MOOCs, the neural-symbolic reasoning module to explore paths in the knowledge graph, and reinforcement learning to simulate user-course interactions for recommendation decisions. Experiments conducted on the MOOCCube dataset show that the proposed method integrating reinforcement learning and neural-symbolic reasoning improves the NDCG and HR metrics by 9.96% and 23.5%, respectively.**

Keywords-Knowledge Graph; Reinforcement Learning; Neuro-Symbolic; Explainable Recommendation

I. INTRODUCTION

The rapid growth of online education has led to an explosive increase in MOOC platform courses. However, the difficulty in precisely matching courses with learner needs has resulted in choice overload and high dropout rates, making personalized course recommendation a hot topic. Introducing recommendation algorithms [1] and knowledge graphs [2] can optimize recommendations. Na Z et al. [3] proposed the Multi-Path RNN Encoder (AGRE) knowledge graph recommendation algorithm, which combines path association with RNN encoding to utilize multiple paths. Xu Y et al. [4] constructed a dual knowledge graph to aggregate graph information and capture user preferences for both interacted and non-interacted item categories. Yang Z et al. [5] employed Fourier transform, inverse transform, and convolutional optimization techniques to process knowledge graph triplet information in the frequency domain. By filtering noise through frequency-domain features and enhancing triplet information via positive contrast learning, they accurately identified user preferences while reducing graph noise and strengthening knowledge perception. However, the large scale of knowledge graphs and lack of effective guidance resulted in inefficient recommendation path search. While reinforcement learning alleviates this issue, it still suffers from low efficiency and insufficient guidance, making it more challenging to discover potential paths within large-scale knowledge graphs. To address this, this paper proposes the IRLNSR method, which integrates the strengths of reinforcement learning and neuro-symbolic reasoning. It relies on reinforcement learning to generate target recommendation paths, leverages neuro-symbolic reasoning for efficient and interpretable inference, and utilizes user profiles to guide path reasoning.

To this end, some scholars have considered incorporating reinforcement learning. Reinforcement learning [6] can model problems as

Markov decision processes (MDPs), naturally capturing the temporal dependencies in decision-making. Altaha A M [7] proposed a method combining reinforcement learning with genetic algorithms (RLGA-FER), which enhances the performance of facial expression recognition systems by dynamically optimizing training data selection. Fan Z et al. [8] proposed a memory-based deep reinforcement learning (MDRL) algorithm. By constructing a historical navigation memory space and integrating gated recurrent units, they effectively enhanced the collision avoidance capability of unmanned surface vehicles in unknown environments. Liwei H et al. [9] introduced a long-term recommendation algorithm. It employs reinforcement learning to capture dynamic views and utilizes recurrent neural networks to learn user interaction behaviors, thereby generating long-term recommendations. Huiting L [10] et al. proposed an interactive reinforcement learning model that leverages text reviews combined with pre-trained review representation models to obtain enhanced embedded representations of item reviews. They formalized the recommendation problem as a Markov decision process, enabling differentiated modeling of users' long-term dynamic preferences. While reinforcement learning techniques were employed to address this challenge, issues such as inefficiency and lack of effective guidance persist, particularly in large-scale knowledge graphs where identifying potential paths remains difficult.

By constructing a shallow neuro-symbolic inference engine as the policy network and processing graph reasoning tasks through behavior cloning training, the method ultimately achieves efficient, precise, and interpretable MOOC course recommendations. The main contributions of this paper include:

Construct a knowledge graph based on the MOOCCube dataset, covering various complex relationships and learners' historical records, providing a basis for the interpretability of recommendation results.

Use reinforcement learning to simulate interactions between users and courses to make recommendation decisions and generate explainable recommendations.

Introduce neural-symbolic reasoning techniques, using them as the policy network in reinforcement learning to explore paths in the knowledge graph, simplifying the action space and improving reasoning efficiency.

II. Related Research

In this article, the MOOC knowledge graph GM is used as the foundation. To conduct more effective path reasoning and course recommendations, the definition and description of the problem will be provided next.

User-centric path: In the knowledge graph G_M, starting from a student entity, a path goes through a series of relationships and entities to finally reach a course entity. A user-centric path can be formalized as:

$$L(e_u, e_c) = (e_u, r_{1,} e_1, ..., r_{t-1}, e_{t-1,} r_t, e_c) \quad (1)$$

Here, e_u is a student entity, which belongs to the user entity subset E_U, that is, $e_u \in E_u$; e_c is a course entity, which belongs to the course entity subset E_C, that is, $e_c \in E_C$.Each r_i in the path represents the relationship at step i and each e_i represents the entity reached at step i.

User Center Pattern: The User Center Pattern is a sequence of relationships within the User Center path, represented as:

$$p = \{r_1, r_2, ..., r_t\} \quad (2)$$

This model depicts specific behavioral patterns of users toward a course through a series of actions (i.e., relationships). By analyzing this model, one can capture students' behavioral characteristics and preferences.

User Profile: A user profile T_U is a collection of user-centered patterns for user u, where each pattern is assigned a corresponding weight to measure its importance to user u. $T_u = \{(p1, w1), (p2, w2), \dots, (p|Tu|, w|Tu|)\}$,$w1, w2, \dots, w|Tu| \in \mathbb{N}$ represent the weights of each pattern. User profiles provide

personalized guidance for recommendation systems, helping them better understand students' needs and preferences.

Inference Engine: The inference engine $\emptyset$ consists of a set of neural symbol inference modules, each corresponding to a relation in the knowledge graph. These modules are used toselect the next node during path inference.

Problem Statement: Given a MOOC knowledge graph G_M containing historical interaction data of students watching learning videos on the MOOC platform, construct user profiles using the knowledge graph. Based on these profiles, select appropriate neural-symbolic reasoning modules to perform path reasoning. Generate a set of recommended courses for students along with the resulting recommendation paths.

III. MCNSR0 MODEL

This model comprises five components: a knowledge graph, user profiles, a neural symbolic reasoning module, a layout tree, and an output path. By leveraging user behavior data within the knowledge graph, it captures key behavioral patterns and preference characteristics to construct user profiles. Through merging patterns from these profiles, it builds a layout tree to guide path inference, ultimately generating a recommended course path. This delivers personalized and explainable course recommendations to users.

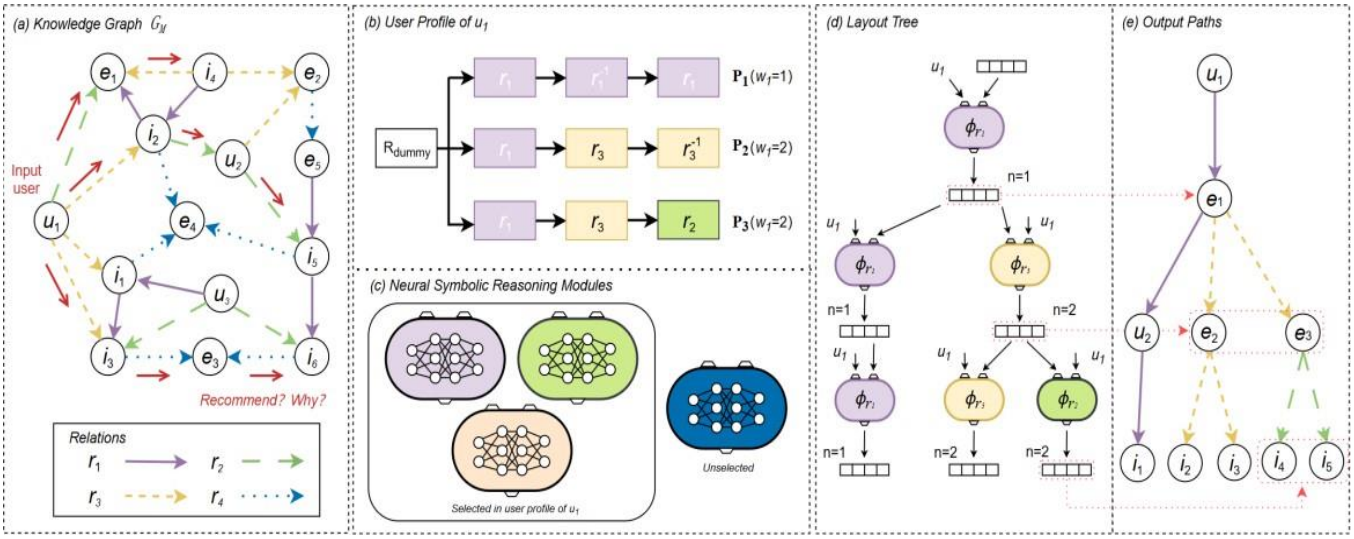

Figure 1. Overall Framework of the MCNSR Algorithm

In Figure 1, a user profile is first constructed using the MOOC knowledge graph G_M to capture significant user behavior patterns. Guided by this profile, path reasoning maintains a neural-symbolic inference machine $\emptyset_r$, where each module corresponds to a relation in the knowledge graph. These modules select the next node during path reasoning. Finally, based on the user profile, corresponding neural-symbolic reasoning modules are selected to construct a layout tree. The structure of this layout tree reflects the relationships and hierarchical levels among different reasoning modules within the user profile, guiding the path inference process to generate recommendations and explanations.

IV. METHOD DESIGN INTEGRATING REINFORCEMENT LEARNING AND NERUAL-SYMBOLIC REASONING

The overall framework is illustrated in Figure 2. The proposed fusion method aims to combine the strengths of reinforcement learning and neuro-symbolic reasoning to enhance the accuracy and diversity of recommendation systems. The specific framework comprises two main stages: path generation and path selection.

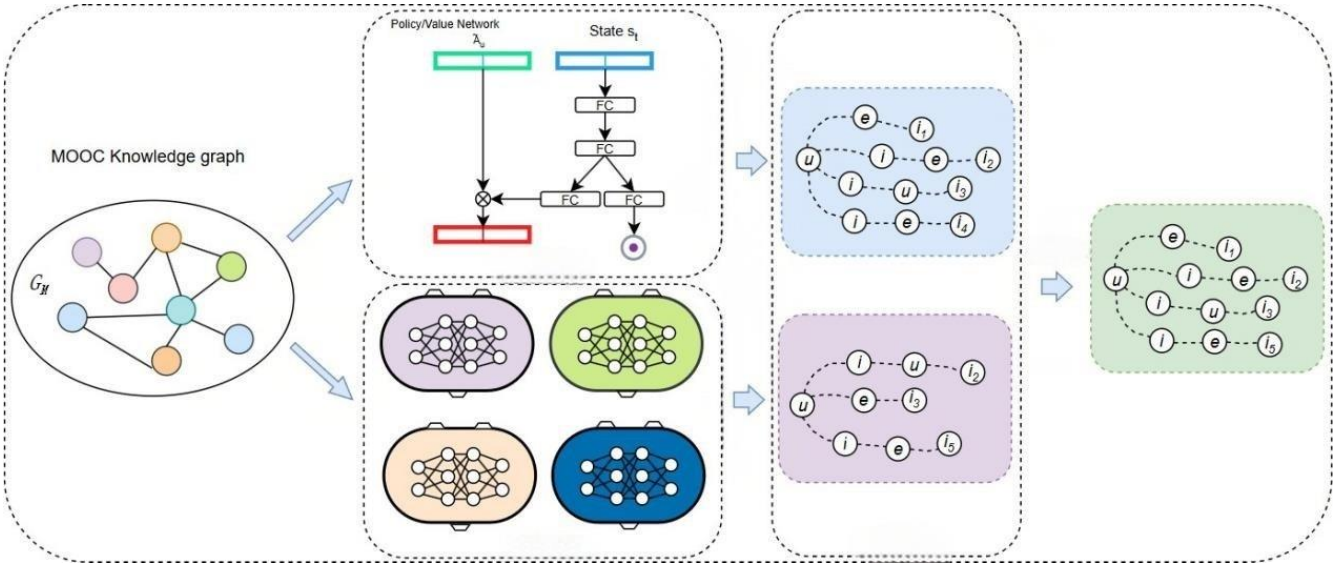

Figure 2. Overall Framework Integrating Reinforcement Learning and Neural Symbolic Reasoning

During the path generation phase, reinforcement learning methods are first employed to dynamically explore paths within the knowledge graph through a policy network. This gradually constructs pathways leading to potential recommended items while calculating the generation probability for each path. Subsequently, user profiles constructed from the knowledge graph guide the neural-symbolic module in path inference, generating diverse paths and assigning generation probabilities to each. These two approaches capture the relationship between user behavior and item attributes from different perspectives, providing a rich pool of candidate paths for selection.

A. Calculation of Path Generation Probability

Path generation is the core step of the fusion method, accomplished through reinforcement learning and neuro-symbolic reasoning. Path selection is achieved via a weighted fusion strategy, with the specific mechanism as follows:

Reinforcement Learning Path Generation: Reinforcement learning methods utilize policy networks to progressively explore the knowledge graph starting from user nodes, discovering paths leading to potential recommended courses. Specifically, the generation probability pRL(p) for each path is calculated by multiplying the probabilities of each action step along the path:

$$P_{RL}(p) = \prod_{t=1}^{T} \pi(a_t \mid s_t, \tilde{A}_u) \tag{3}$$

$\pi(at|st, \tilde{A}u)$ represents the probability of taking action at in state st, as output by the policy network.

Neural-Symbolic Path Generation: The neural-symbolic reasoning method utilizes knowledge graphs to construct user profiles, guiding the neural-symbolic module in path inference to ultimately derive pathways to potentially interesting courses. The generation probability PNSR (L|u) for each path is calculated as follows:

$$P_{NSR}(L \mid u) = \prod_{t=1}^{T} p(r_t, e_t \mid u, h_t) \tag{4}$$

$P(rt, et|u, ht)$ represents the probability of selecting the relationship rt and entity et, as output by the neural symbolic reasoning module, given user u and historical trajectory ht.

Normalization of Path Generation Probabilities: To ensure comparability between the two methods, we applied Min-Max normalization to the probabilities, mapping them to the interval [0,1]:

$$P_{nomalized}(p) = \frac{p(p) - \min(p)}{\max(p) - \min(p)} \tag{5}$$

Where P(p) is the original generation probability of path p, min(p) and max(p) are the minimum and maximum values of all path generation probabilities, respectively.

B. Integration Strategy

For each user, the paths generated by the two methods are weighted averaged based on their generation probabilities. For each path pair (p_i, L_j), the fused generation probability is calculated:

$$P_{combined}(p_u, L_i) = \alpha \cdot P_{RL}(p_i) + (1-\alpha) \cdot P_{NSR}(L_j) \quad (6)$$

The paths p_i and L_j are generated through two distinct methods, where α is a weighting parameter used to balance the importance of paths produced by each method. All fused paths are ultimately sorted in descending order based on their generation probability, and the top K paths with the highest generation probabilities are selected as the final recommended paths.

V. EXPERIMENTS AND ANALYSIS OF RESULTS

A. Baseline Method

BPR: Bayesian Personalized Ranking aims to uncover latent associations between users and items. It primarily analyzes implicit user behaviors such as clicks and saves. Through Bayesian inference, BPR calculates the probability of a user's preference for each item, thereby ranking items to generate a recommendation list.

BPMF: Bayesian Probabilistic Matrix Factorization is a recommendation model that employs Markov Chain Monte Carlo methods. It is derived from the Probabilistic Matrix Factorization (PMF) model by incorporating a Bayesian framework.

DQN: A Deep Reinforcement Learning Method. Combining the concepts of Q-learning and deep neural networks, it utilizes a deep neural network to approximate the Q-function. Through continuous training, it learns optimal recommendation strategies to deliver personalized recommendation results.

CF: A user-based collaborative filtering method. It leverages similarities among users and their behavioral histories to provide item recommendations aligned with their interests.

B. Parameter Settings

The experimental parameters are set as follows: 20% of the dataset is used as the test set, while the remaining 80% serves as the training set. The learning rate (lr) for training the knowledge graph embeddings was set to 0.0001, with 50 training epochs. Each entity's embedding vector has 100 dimensions. For each neural relation module ϕr associated with a specific relation r the parameters are $Wr,1 \in \mathbb{R}200\times256$, $Wr,2 \in \mathbb{R}256\times256$ and $Wr,3 \in \mathbb{R}256\times100$.

Parameters were initialized using the Xavier initialization method. Model training employed the Adam optimizer with a learning rate of 10-4, a batch size of 128, and a total of 50 training epochs. The ranking loss weight parameter λ was set to 10, while the fusion weight parameter α was set to 0.4. The maximum path length was similarly constrained to 3.

C. Experiments and Results Analysis

TABLE I. EXPERIMENTAL COMPARISON

Measures(%)	*HR(k=5)*	*NDCG(k=5)*	*HR(k=10)*	*NDCG(k=10)*
DQN	1.17	0.45	1.24	0.47
BPMF	1.54	0.92	1.71	0.78
BPR	4.32	2.36	8.47	3.89
U-CF	16.87	8.36	21.96	9.98
IRLNSR	**24.67**	**10.41**	**37.78**	**13.84**

The experimental results are shown in Table 1, presenting the HR and NDCG outcomes for K=10 and K=5. The DQN method performed worst across all metrics, primarily due to its difficulty in learning effectively under high-dimensional state spaces and sparse reward signals, resulting in suboptimal recommendations. BPR relies heavily on implicit user feedback, a method that overlooks the diversity and complexity of user behavior, making it challenging to capture intricate

relationships between users and items. Overall, the IRLNSR method outperforms other baseline results in both HR and NDCG. By integrating reinforcement learning with neural symbolic reasoning, this hybrid approach optimizes both path generation and selection phases, fully leveraging the rich information from knowledge graphs. This enables IRLNSR to capture relationships between user behavior and item attributes more comprehensively and flexibly, thereby enhancing the accuracy and diversity of recommendations.

D. Ablation Experiment

To validate the effectiveness of the fusion method, ablation experiments were conducted in this section. The three approaches — IMCRec, MCNSR, and IRLNSR — were compared under identical experimental conditions, with MCNSR implemented using neuro-symbolic reasoning, as shown in Table 2.

TABLE II. ABLATION EXPERIMENTS

Measures(%)	*HR(k=5)*	*NDCG(k=5)*	*HR(k=10)*	*NDCG(k=10)*
IMCRec	23.55	8.66	37.39	10.22
MCNSR	22.73	9.45	33.32	12.23
IRLNSR	**24.67**	**10.41**	**37.78**	**13.84**

As shown in Table 2, the IRLNSR method outperforms both IMCRec and MCNSR across all metrics. This superiority stems primarily from its integration of reinforcement learning's dynamic decision-making capabilities with neural-symbolic reasoning's efficient inference abilities. This synergy enables more comprehensive utilization of knowledge graph information, generating more accurate and diverse recommendation paths. Additionally, IRLNSR enhances the interpretability of recommendation results compared to IMCRec. Since each neuro-symbolic module corresponds to a specific relationship, the generated paths can be clearly explained as the process by which users navigate from one entity to another through these relationships. This explicit relational path makes the interpretation of recommendation results more intuitive and easier to understand.

E. Comparison of Loss Values Between IMCRec and MCNSR

The loss values for training the MCNSR model and the IMCRec model were compared, with the results shown in Figure 3、figure4.

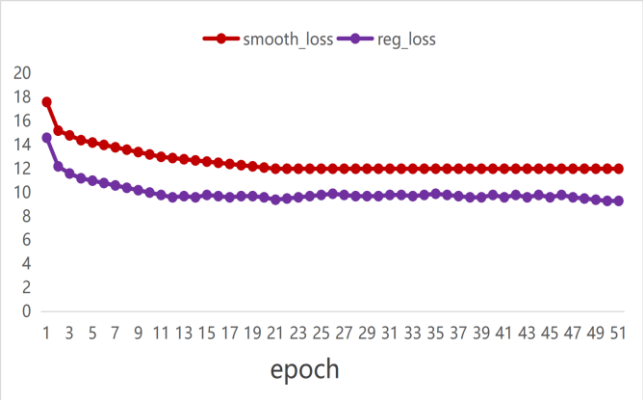

Figure 3. The change in loss value during MCNSR.

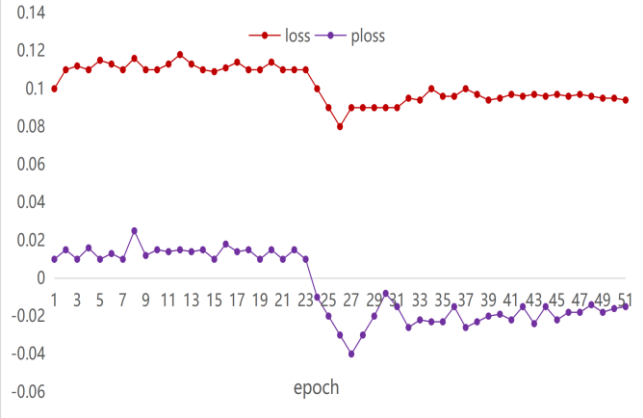

Figure 4. The change in loss value during IMCR.

In the loss value changes of MCNSR, we present the smoothed total loss and regularization loss. During the first 0-5 training rounds, both smoothed loss and regularization loss decrease rapidly, stabilizing around 25 rounds. In contrast, during the training of the IMCRec model, both the total loss and policy loss exhibit significant fluctuations as training iterations increase, demonstrating insufficient stability. This stems from the reinforcement learning model's continuous "interaction-trial-and-error" process without true labels, leading to high randomness. However, the fluctuations in policy loss also

indicate the model's balance between exploration and exploitation.

F. Authors and Affiliations

To determine the optimal weight parameter α, we systematically searched within a predefined weight range using a grid search method. Specifically, we uniformly selected 10 points between 0 and 1 as candidate values for α. Figure 5、Figure 6 illustrates the results of NDCG and HR as α varies when K=10.

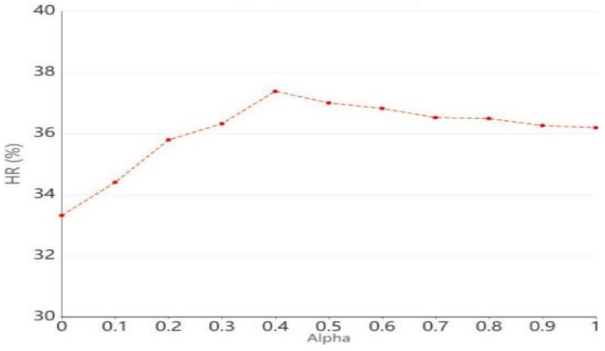

Figure 5. The change in HR value when k=10

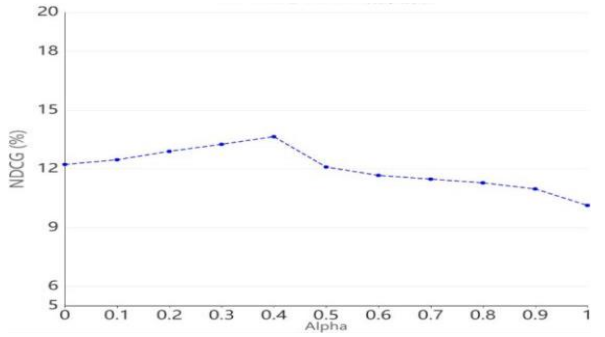

Figure 6. The change in NDCG value when k=10

At α=0.4, HR@10 reaches its maximum value. Prior to this point, the model's hit rate improves as α increases. Once α exceeds 0.4, HR@10 begins to decline, indicating that further increasing α may degrade model performance. NDCG@10 first increases and then decreases with increasing α, peaking at α=0.4. Therefore, at α=0.4, the model not only achieves the best hit rate but also the optimal normalized discounted cumulative gain.

VI. Summary and Future Work

This chapter first introduces a course recommendation framework based on neuro-symbolic reasoning, encompassing user profiling construction, inference module design, and path reasoning processes. User profiles capture key characteristics by analyzing behavioral patterns within knowledge graphs, providing personalized guidance for the recommendation system. The inference module processes relationships in knowledge graphs via shallow neural networks, simplifying the action space and enhancing reasoning efficiency. The path inference process constructs layout trees and executes the inference module to generate paths leading to potentially interesting courses. Subsequently, we propose an integrated framework combining reinforcement learning with neural symbolic reasoning. During path generation, reinforcement learning dynamically explores paths via policy networks, while neural symbolic reasoning utilizes user profiles to guide path inference, generating diverse paths. Finally, normalized processing and weighted fusion strategies select the path with the highest generation probability as the final recommendation path. Experimental results demonstrate that the combined reinforcement learning and neuro-symbolic reasoning approach outperforms other baseline methods on both HR and NDCG metrics.

Acknowledgment

The authors would like to express their gratitude to Anhui Xinhua University (China) for the support the University-level Scientific Research Project of Anhui Xinhua University (2024zr012). Additionally, we appreciate the support from. Provincial Innovation and Entrepreneurship Program for College Students (S202512216136)

References

[1] Huanyu Z, Xiaoxuan S, Baolin Y, et al. KGAN: Knowledge Grouping Aggregation Network for course recommendation in MOOCs[J]. Expert Systems with Applications,2023,211.

[2] Qiu Z, Tao Y, Pan S, et al. Knowledge Graphs and Pretrained Language Models Enhanced Representation Learning for Conversational Recommender Systems. [J]. IEEE transactions on neural networks and learning systems, 2024, PP.

[3] Na Z, hen L, Jian W, et al. AGRE: A knowledge graph recommendation algorithm based on multiple paths embeddings RNN encoder [J]. Knowledge-Based Systems, 2023,259.

[4] Xu Y, Li T, Yang Y, et al. An adaptive category-aware recommender based on dual knowledge graphs [J]. Information Processing and Management, 2024, 61(3): 103636-.

[5] Yang Z, Li L. Knowledge graph-based recommendation with knowledge noise reduction and data augmentation[J]. Applied Intelligence,2024, (prepublish):1-27.

[6] Arta I, Ali M G, Mubbashir A, et al. A reinforcement learning recommender system using bi-clustering and Markov Decision Process [J]. Expert Systems with Applications, 2024,237(PB):

[7] Altaha A M, Jarraya I, Haddad L, et al. RLGA-FER: reinforcement learning based on genetic algorithm for facial expression recognition enhancing [J]. International Journal of Machine Learning and Cybernetics,2024, (prepublish):1-14.

[8] Fan Z, Wu D, Li Y, et al. Memory-based deep reinforcement learning for COLREGs-compliant obstacle avoidance in USV with limited environmental knowledge [J]. Ocean Engineering, 2025,338121978-121978.

[9] Liwei H, Mingsheng F, Fan L, et al. A deep reinforcement learning based long-term recommender system [J]. Knowledge-Based Systems,2021,213.

[10] Huiting L, Kun C, Peipei L, et al. REDRL: A review-enhanced Deep Reinforcement Learning model for interactive recommendation [J]. Expert Systems with Applications, 2023, 213(PA).