MRMLREC: A Two-Stage Approach for Addressing Data Sparsity in MOOC Video Recommendation (Student Abstract)

Ye Zhang[1], **Yanqi Gao**[1], **Yupeng Zhou**[1*], **Jianan Wang**[2], **Minghao Yin**[1*]

[1]College of Information Science and Technology, Northeast Normal University
[2]School of Physics, Northeast Normal University
{zhangy923, gaoyq557, zhouyp605, wangjn, ymh}@nenu.edu.cn

Abstract

With the abundance of learning resources available on massive open online courses (MOOCs) platforms, the issue of interactive data sparsity has emerged as a significant challenge.This paper introduces MRMLREC, an efficient MOOC video recommendation which consists of two main stages: multi-relational representation and multi-level recommendation, aiming to solve the problem of data sparsity. In the multi-relational representation stage, MRMLREC adopts a tripartite approach, constructing relational graphs based on temporal sequences, courses-videos relation, and knowledge concepts-video relation. These graphs are processed by a Graph Convolution Network (GCN) and two variant Graph Attention Networks (GAT) to derive representations. A variant of the Long Short-Term Memory Network (LSTM) then integrates these multi-dimensional data to enhance the overall representation. The multi-level recommendation stage introduces three prediction tasks at varying levels—courses, knowledge concepts, and videos—to mitigate data sparsity and improve the interpretability of video recommendations. Beam search (BS) is employed to identify top-β items at each level, refining the subsequent level's search space and enhancing recommendation efficiency. Additionally, an optional layer offers both personalization and diversification modes, ensuring variety in recommended videos and maintaining learner engagement. Comprehensive experiments demonstrate the effectiveness of MRMLREC on two real-world instances from *Xuetang X*.

Introduction

As an emerging education mode, MOOCs platforms are developing rapidly, which provide universal access to a vast array of MOOC videos from prestigious universities worldwide. The challenge of efficiently assisting students in selecting the appropriate video within a limited timeframe remains a pressing issue that requires resolution.

The existing MOOC video recommendation is typically classified into two categories: (1) learning record-based recommendations, which utilize sequence recommendation methods to emphasize the significance of previous courses (Xia et al. 2021; Dang et al. 2023); (2) graph-based recommendations, which construct a heterogeneous information graph of learners, videos, and courses to obtain the preferences of learners (He et al. 2020; Xu et al. 2023). Although these methods are well-designed, they do not account for the short-term learning needs of learners during the learning process. Furthermore, they do not address the issue of the sparse interaction matrix, making it difficult to recommend the next learning videos.

In this paper, we propose a novel approach called MRMLREC, which utilizes a multi-layer mapping of multiple relationships. The method is divided into two stages: (1) multi-relational learner representation, which not only captures the representations under three relationships - courses, knowledge concepts, and temporal relationships - but also addresses the issue of data sparsity. (2) Multi-level recommendation, which integrates course and knowledge concept prediction tasks into the video recommendation task, performs the beam search considering only the joint probabilities of the top-β items at different levels to reduce the search space, and includes an optional layer balance for personalization and diversification. MRMLREC is an interpretable method with a small search space that can effectively handle data sparsity. Our experiments are conducted on real-world datasets, demonstrating the effectiveness of MRMLREC.

Method

Problem Formulation. Given $|U|$ learners, $|V|$ MOOC videos, $|K|$ knowledge concepts and $|C|$ courses, let $U = \{u_1, u_2, ..., u_{|U|}\}$ be the set of learners, $R = \{R_{u_1}, R_{u_2}, ..., R_{u_{|U|}}\}$ be the set of watching records for all learners. Each learner's watching sequence $R_{u_m} = \{(v_{t_1}, time_{t_1}), (v_{t_2}, time_{t_2}), ..., (v_{t_i}, time_{t_i})\}$ consists of consecutive video learning information, where v_{t_i} is the video watched in time t_i, and $time_{t_i}$ is the timestamp. As each learner's sequence R_{u_m} is divided into several temporal-relational subgraphs $\mathcal{G}^{t_i}$ to capture the short-term learning interests of learners, each video has its own corresponding relationship with knowledge concepts and courses. A course-relational graph is defined as $\mathcal{G}_C = \{\mathcal{V}_C, \mathcal{E}_C\}$, where $\mathcal{V}_C = V$, and $\mathcal{E}_C$ denotes the association of videos containing the same course. A knowledge concept-relational graph is defined as $\mathcal{G}_K = \{\mathcal{V}_K, \mathcal{E}_K\}$, where $\mathcal{V}_K = V$, and $\mathcal{E}_K$ represents whether the knowledge concepts covered by different videos are similar. Each pair of videos (v_i^K and v_j^K)

*These authors contributed equally.

	Computer			*MOOCCube*		
	HR@10	**NDCG@10**	**MRR**	**HR@10**	**NDCG@10**	**MRR**
COTREC	0.0667	0.0304	0.0235	0.0192	0.0120	0.0118
lightGCN	0.0366	0.0178	0.0151	0.0490	0.0210	0.0188
TiCoseRec	0.1306	0.0599	0.0442	0.1426	0.0724	0.0571
IntGNN	0.4950	0.2219	0.1933	0.1171	0.0603	0.0468
MRMLREC	**0.7887**	**0.7261**	**0.7075**	**0.7932**	**0.7440**	**0.7294**

Table 1: Comparison with baselines.

has a link if they have a Jaccard similarity $\frac{|v_i^K \cap v_j^K|}{|v_i^K \cup v_j^K|}$ above 0.6. Our task is to predict the learner's next learning video through his watching records, course-video correlation matrix and knowledge concept-video correlation matrix.

Multi-relational representation. We obtain the multi-relational representation from three graphs: temporal-relational subgraph $\mathcal{G}^{t_i}$, course graph-relational $\mathcal{G}_C$ and knowledge concept-relational graph $\mathcal{G}_K$, by using GCN and two GAT respectively. Then, we use a variant LSTM layer to learn the dependencies between video learning sequences and jointly integrate these three-relational representations, with the embedding $e_v^{t_{i-1}}$ of video as input:

$$i_{t_i}=\sigma(\mathbf{W}_i e_v^{t_i}+\mathbf{U}_i h_{t_{i-1}}+\mathbf{V}_i p_{t_i}^{\mathcal{G}_C}+\mathbf{Z}_i p_{t_i}^{\mathcal{G}_K}+\mathbf{Q}_i \mathbf{H}_{t_i}^{\mathcal{G}_{t_i}}+\mathbf{b}_i) \quad (1)$$

$$f_{t_i}=\sigma(\mathbf{W}_f e_v^{t_i}+\mathbf{U}_f h_{t_{i-1}}+\mathbf{V}_f p_{t_i}^{\mathcal{G}_C}+\mathbf{Z}_f p_{t_i}^{\mathcal{G}_K}+\mathbf{Q}_f \mathbf{H}_{t_i}^{\mathcal{G}_{t_i}}+\mathbf{b}_f) \quad (2)$$

$$o_{t_i}=\sigma(\mathbf{W}_o e_v^{t_i}+\mathbf{U}_o h_{t_{i-1}}+\mathbf{V}_o^{\mathcal{G}_C}+\mathbf{Z}_o p_{t_i}^{\mathcal{G}_K}+\mathbf{Q}_o \mathbf{H}_{t_i}^{\mathcal{G}_{t_i}}+\mathbf{b}_o) \quad (3)$$

$$h_{t_i}=tanh(\mathbf{W}_c e_v^{t_i}+\mathbf{U}_c h_{t_{i-1}}+\mathbf{V}_c p_{t_i}^{\mathcal{G}_s}+\mathbf{Z}_c p_{t_i}^{\mathcal{G}_t}+\mathbf{Q}_c \mathbf{H}_{t_i}^{\mathcal{G}_{t_i}}+\mathbf{b}_c) \quad (4)$$

where $e_v^{t_i}$, $p_{t_i}^{\mathcal{G}_C}$, $p_{t_i}^{\mathcal{G}_K}$ and $\mathbf{H}_{t_i}^{\mathcal{G}_{t_i}}$ are the representations of the video watched at time t_i, $\mathcal{G}_C$, $\mathcal{G}_K$ and $\mathcal{G}^{t_i}$ respectively. $W\bullet, U\bullet, V\bullet, Z\bullet, Q\bullet$ are the weight matrices for $e_v^{t_i}$, $h_{t_{i-1}}$, $p_{t_i}^{\mathcal{G}_C}$, $p_{t_i}^{\mathcal{G}_K}$ and $\mathbf{H}_{t_i}^{\mathcal{G}_{t_i}}$ respectively. $i_\bullet, f_\bullet, o_\bullet$ are the input, forget and output gate respectively. $b\bullet$ is the bias. The current cell state c is applied to compute the output hidden multi-relational representation h_{t_i} of learner at time t_i. Then, softmax normalization is performed to compute the conditional probability of the next video $tk^{Video} = P(v_{t_i}|v_{t_{i-1}}) = softmax(DL_{\mathbf{W}_L}(DO(h_{t_i} \oplus e_{u_m})))$, where DO is the dropout layer, DL is projected onto the $|V|$ videos as a dense linear layer parameterized with $\mathbf{W}_L$, and e_{u_m} is the embedding of u_m.

Multi-level recommendation. Since courses, knowledge concepts and videos are interrelated, we use the previous domain to help predict next domain $L@D, D \in \{C, K\}$ tasks. For all tasks TK, we use h_{t_i} as a common representation to share the feature learning. We apply the task-specific embedding of the previous domain $e_{L@D}^{t_{i-1}}$ to predict the next corresponding $v_{t_{i-1}}^{L@D}$. The task distribution is computed by the joint probabilities $tk^{L@D} = P(v_{t_i}^{L@D}|v_{t_{i-1}}) = softmax(DL_{\mathbf{W}_{L@D}}(DO(h_{t_i} \oplus e_{L@D}^{t_{i-1}})))$, where DO is the dropout layer. Traversing and sorting the sum of the log-probabilities of $\{tk_1^{L@C}, tk_2^{L@K}, tk_3^{Video}\}$ over all items is inefficient. Therefore, we adopt the idea of BS to expand only the top-β promising items of each task distribution during traversal, where β is the beam width. Specifically, given a continuous pair of task distribution from $\{tk_1^{L@C}, tk_2^{L@K}, tk_3^{Video}\}$, tk_i^β is calculated by Eq (5).

$$tk_i^\beta=f(\{\log(tk_{i-1,b}^\beta)+\log(tk_{i,j})|tk_{i,j} \in \mathcal{N}(tk_{i-1,b}^\beta)\}) \quad (5)$$

where tk_{i-1}^β and tk_i^β are the top-β partial solutions for the input task distributions of tk_{i-1} and tk_i respectively. For each top beam $b \in \{1, 2, ..., \beta\}$ of previous input $tk_{i-1,b}^\beta$, we first identify the directed neighborhoods $tk_{i,j} \in \{\mathcal{N}(tk^{\beta_{i-1,b}})\}$ from the next input task distribution tk_i, and compute the sum of the log-probabilities of items connected to j (i.e. $log(tk_{i-1,b}^\beta) + log(tk_{i,j})$) in each level. Eq (5) is used to calculate tk_i^β considering only top-β partial solutions. An optional layer is then incorporated to predict whether the learner has watched the next video. If this prediction is true, the personalization mode is employed, which recommends videos based on the learner's representation. Otherwise, the diversification mode is used, which considers the joint probabilities $P(L@C, L@K, v_{t_i}|v_{t_{i-1}})$ of items at different levels obtained through BS to conform to the logical structure of the learning process.

Experiment

Datasets & baselines. We conduct experiments on two public real-world datasets on *Xuetang X* (https://www.xuetangx.com) MOOCs platform, namely *Computer* and *MOOCCube*. We compare MRMLREC with the following baselines: COTREC (Xia et al. 2021), lightGCN (He et al. 2020), TiCoSeRec (Dang et al. 2023), IntGNN(Xu et al. 2023).

Performance Comparison. Table 1 summarizes the results. MEMLREC outperforms baselines and draws two important conclusions: (1) Time intervals play a crucial role in recommendation with sparse data, and MRMLREC has improved significantly in handling this issue by effectively considering the impact of time intervals on learners' interests. Our multi-relational learner representation extracts the learner's short-term learning interests, which helps address this problem. (2) MRMLREC incorporates course and knowledge concept prediction tasks into the recommendation process to address data sparsity. This approach leads to more interpretable recommendations by leveraging the relationships between courses and knowledge concepts. These are two crucial reasons for the superior performance of MEMLREC over other baselines.

Acknowledgments

This work is supported by NSFC under Grant No.(62206045, 61976050, 61972384, 61806050, 61972063), the Fundamental Research Funds for the Central Universities under Grant No. (2412022QD041, 2412020FZ030, 2412019ZD013, 2412019FZ051, 2412022QD040, 2412022ZD018, 93K172022K10), and the Natural Science Research Foundation of Jilin Province of China under Grant No. YDZJ202201ZYTS423.

References

Dang, Y.; Yang, E.; Guo, G.; Jiang, L.; Wang, X.; Xu, X.; Sun, Q.; and Liu, H. 2023. Uniform Sequence Better: Time Interval Aware Data Augmentation for Sequential Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 4225–4232.

He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; and Wang, M. 2020. 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*, 639–648.

Xia, X.; Yin, H.; Yu, J.; Shao, Y.; and Cui, L. 2021. Self-supervised graph co-training for session-based recommendation. In *Proceedings of the 30th ACM international conference on information & knowledge management*, 2180–2190.

Xu, G.; Yang, J.; Guo, J.; Huang, Z.; and Zhang, B. 2023. Int-GNN: A User Intention Aware Graph Neural Network for Session-Based Recommendation. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.