

融合多种时间关系的时序图课程推荐算法

张 维, 周旭宸, 曾鑫耀, 朱诗怡

(华中师范大学人工智能教育学部, 湖北 武汉 430079)

摘要: 在学习者学习过程中, 学习记录中的时序特征反映了学习者不断变化的兴趣、学习周期和课程间先后依赖关系等多种重要信息。目前课程推荐只考虑课程顺序关系, 并且大多数图神经网络课程推荐算法完全丢弃了时序特征, 导致性能降低。提出一种融合多种时间关系的时序图模型, 充分利用时序特征提升表征精确度。模型首先将时序特征转换为3种时间关系: 绝对时间、顺序时间、间隔时间, 以获得细粒度的时间信息。其次, 模型依据交互记录构建学习者—课程交互时序图, 通过3种时间关系嵌入和注意力机制为邻居节点分配个性化聚合权重, 再经过残差连接与多层传播得到学习者和课程表征进行最终预测。在MOOCCourse数据集上的大量实验表明, 该模型相比其他推荐模型, 在R@5与NDCG@15两个指标上分别提升了6.58%和2.61%, 并且融合3种时间关系相比仅考虑课程顺序关系在R@5和NDCG@15指标上提升更多。

关键词: 课程推荐; 图神经网络; 时序特征; 推荐系统; 注意力机制

DOI: 10.11907/rjtk.241788

中图分类号: TP391.3

文献标识码: A

文章编号: 1672-7800(2026)001-0054-09

扫描二维码阅读全文:



Time-Series Graph Course Recommendation Algorithm Integrating Multiple Temporal Patterns

ZHANG Wei, ZHOU Xuchen, ZENG Xinyao, ZHU Shiyi

(Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan 430079, China)

Abstract: In the learning process, the temporal features of learning records reflect a variety of important information, such as learners' changing interests, learning cycles, and successively dependent relationships between courses. At present, the course recommendation only considers the course order relationship, and most of the graph neural network course recommendation algorithms completely discard the temporal features, resulting in performance degradation. A time-series graph integrating multiple temporal patterns for course recommendation is proposed to make full use of temporal features to improve the representation accuracy. In order to obtain fine-grained time information, the model first converts the temporal features into three kinds of time patterns: absolute time, sequential time and interval time. Secondly, the model constructs a learner-course interaction time-series graph, assigns individualized aggregate weights to neighbor nodes through three kinds of time patterns embedding and attention mechanisms, and then obtains learner and course representations through residual connection and multi-layer propagation for final prediction. A large number of experiments on the MOOCCourse dataset show that the proposed model outperforms other advanced recommendation models by 6.58% and 2.61% on R@5 and NDCG@15, respectively, and prove that the performance of combining three time patterns is better than considering only the sequential relationship of courses on R@5 and NDCG@15.

Key Words: course recommendation; graph neural network; temporal features; recommender system; attention mechanism

0 引言

自21世纪起, 在线学习平台如雨后春笋般不断涌

现^[1]。截至2024年1月底, 我国上线的MOOC数量超过7.68万门, 学习者面对的问题从“信息匮乏”转变为“信息过载”^[2-3]。研究者尝试使用课程推荐算法改善此问题, 以提升学习效率^[4]。

收稿日期: 2024-10-22

基金项目: 国家自然科学基金项目(62377024)

作者简介: 张维(1975-), 男, 华中师范大学人工智能教育学部教授, 研究方向为计算机应用、教育数据挖掘、智能教育评价; 周旭宸(2000-), 男, 华中师范大学人工智能教育学部硕士研究生, 研究方向为推荐算法; 曾鑫耀(1998-), 男, 华中师范大学人工智能教育学部博士研究生, 研究方向为教育数据挖掘、推荐算法; 朱诗怡(2001-), 女, 华中师范大学人工智能教育学部博士研究生, 研究方向为教育数据挖掘、推荐算法。本文通讯作者: 周旭宸。

近年来研究者将图模型引入课程推荐,构建学习者—课程二部图,获取学习者与课程的高阶表示,进而推荐目标课程。例如Wang等^[5]利用图神经网络(Graph Neural Network, GNN)和两种聚合函数处理学习序列,实现课程推荐。尽管这类方法取得了良好效果,但仍然面临以下挑战:

(1)课程推荐是强时效性的任务,然而大多数基于图模型的课程推荐在构建二部图时丢弃了时序特征,无法捕捉学习者的兴趣演变、学习周期等特征,导致建模的学习

者和课程表征损失了此类信息,从而影响了模型性能。

(2)传统的课程推荐方法通常关注课程之间的顺序关系,但在真实场景中,其他因素如学习时间和周期也对学生选课产生影响。学习者交互数据实例如图1所示。例如,图1中学习者A和B都学习了高等数学,但B的学习时间(t_1)早于A(t_6),可能需要复习材料,而A适合推荐线性代数。此外,学习周期不同也影响知识的吸收和应用,从而影响后续推荐策略。这些信息需要引入更多时间关系来捕捉。

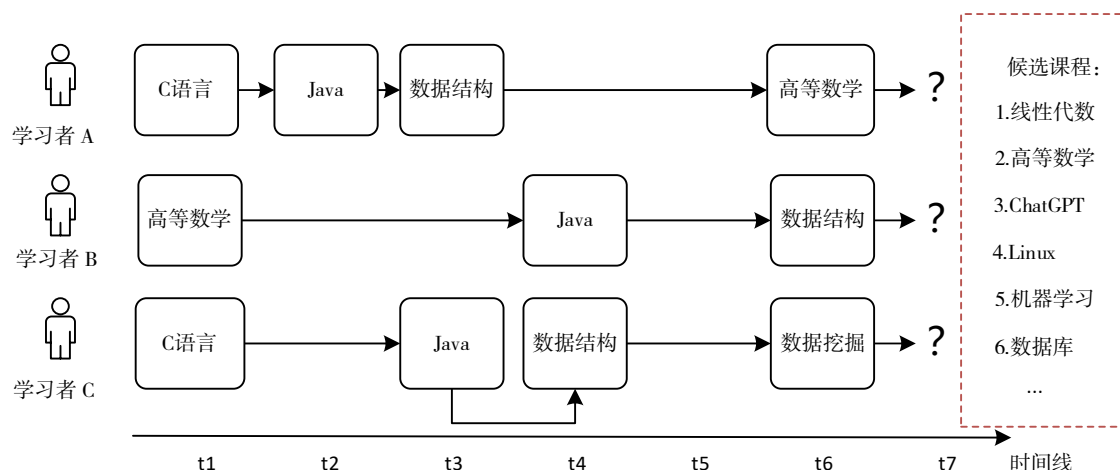


Fig. 1 Examples of learner interaction data

图1 学习者交互数据实例

本文针对课程推荐,提出融合多种时间关系的时序图模型(A Time-Series Graph integrating Multiple Temporal Patterns for Course Recommendation, TMPCR)。模型考虑了绝对时间、顺序时间和间隔时间3种时间关系,通过设计时序图和注意力机制计算不同邻居节点的聚合权重,改善了传统图模型无法捕捉时序特征的缺点。同时,采用层归一化和残差连接操作避免过拟合,并使用负采样策略更新模型参数。本文的主要工作如下:①为了充分利用时序特征,将原来单一的顺序时间扩展为3种有效的时间关系;②为了保证课程推荐的准确性与时效性,设计了一种融合多种时间关系的时序图。在图卷积过程中,依据不同节点和3种时间关系嵌入计算个性化的聚合权重;③在MOOCourse上进行大量实验,结果表明该模型相比已有的先进模型,性能获得了明显提升,且消融实验证明同时利用3种时间关系在召回率和精确率上表现最好。

1 相关工作

1.1 课程推荐算法

课程推荐算法作为教育信息过滤的重要技术,旨为学习者提供个性化的学习资源。传统的课程推荐算法主要利用协同过滤、关联规则和矩阵分解等方法^[6-7]。例如Jing等^[8]提出一种基于协同过滤的课程推荐算法,在课程先决条件关系下学习用户的兴趣表示。其背后的原理直

观,实现快速,为之后的研究打下了良好基础。

近年来,得益于神经网络的强大性能,许多研究者开始探索将深度学习应用于课程推荐领域^[9-10]。在深度课程推荐算法中,序列模型和图神经网络模型成为了主流。例如,Jiang等^[11]利用改进的递归神经网络获取学生的选课信息,从而获取学生的个性化偏好。Zhang等^[12]提出一种学习动机增强的可解释时序课程推荐模型,通过4种学习动机和两种时间关系提升了序列模型的表征能力。随着图模型的表征能力得到认可,研究者将学习者交互序列转换成学习者—课程二部图,通过图卷积机制深入挖掘学习者偏好。例如,Wang等^[13]提出一种基于超边的图神经网络课程推荐算法,学习者之间的相似性关系被表示为超图中超边之间的重叠关系。Zhao等^[14]针对教育场景中的数据稀疏问题,提出一种基于图对比学习的推荐模型,利用学生和课程的静态特征来扩展稀疏的交互数据。Ma等^[15]基于多图对比学习,利用受欢迎程度偏差以及学习者的知识背景提高MOOC教育环境下课程推荐的公平性。

然而,当前图模型在课程推荐系统的应用中仍未能充分利用交互过程中的时序特征。学习者的兴趣动态变化、学习周期以及课程先后依赖关系等关键信息均深藏于交互序列的时序特征中。本文将此类时序特征有效整合至课程推荐模型中,以增强推荐性能。

1.2 图模型推荐算法

如今,图神经网络(Graph Neural Network, GNN)通过

其强大的图表征学习能力,在推荐系统中展现出显著优势^[16]。构建基于GNN的推荐模型通常包括3个步骤:首先,将用户的历史交互数据转化为用户—物品二部图;其次,设计聚合方法从图中节点的邻居学习并更新嵌入表示;最后,利用这些嵌入实现个性化推荐^[17]。

图神经网络的优势主要得益于第二步,通过多层传播,每个节点可以访问高阶邻居信息^[18]。许多研究聚焦于开发不同的聚合技术来优化用户和物品的嵌入学习^[19]。Ying等^[20]开发了一种基于随机游走的推荐算法,该算法在交互图上使用多个图卷积层聚合来自高阶邻居的特征信息。随后,Wang等^[21]通过在用户—项目二部图上传播用户和项目的嵌入,设计了一种神经图协同过滤模型(Neural Graph Collaborative Filtering, NGCF)。He等^[22]进一步去除了NGCF中的冗余模块,降低了模型复杂度。以上模型在表征学习中忽略了时间因素,限制了其在真实世界场景中充分捕捉用户随时间不断变化的动态兴趣的能力^[23]。

Gong等^[24]引入时空图神经网络,使用长短时记忆网络生成时间感知的邻接矩阵,以捕获用户的动态兴趣变化。Chen等^[25]设计了一种序列敏感注意机制,将艾宾浩斯遗忘理论纳入注意力机制,以描述交互过程中的遗忘和记忆过程。

然而,融合类似长短时记忆网络模型的时空图神经网络只能考虑顺序时间关系,用户的兴趣演变也不应只考虑遗忘。本文根据交互记录设计了3种时间关系并将其融入图神经网络,以探索不同时间关系及其组合对推荐性能的影响。

2 问题定义

首先通过一些定义给出本文所研究的问题和形式化描述。

定义 1 交互记录: $U = \{u_1, u_2, \dots, u_m\}$, $C = \{c_1, c_2, \dots, c_n\}$ 代表 m 个学习者、 n 个课程的集合。一条学习记录通常用一个三元组来表示 $(u_j, c_i, t_l) \in U \times I \times T$, 其表示学习者 u_j 和课程 c_i 在 t_l 时刻的交互记录。

定义 2 交互序列: 对于学习者集合中的任何一个学习者 u_j , 其交互序列表示为 $S_{u_j} = \{(c_{u_j1}, t_1), (c_{u_j2}, t_2), \dots, (c_{u_jL}, t_L)\}$, 每一个括号代表一次交互, 包含交互的课程和时间戳。整个交互的长度为 L 。

定义 3 课程推荐任务: 给定一个学习者 $u_j \in U$ 及其历史课程学习序列 S_{u_j} , 任务的目标是预测最符合学习者 u 在 t_{L+1} 时刻偏好的课程 $c_{u_{jL+1}}$ 。

定义 4 时序图: 时序图实例如图 2 所示。原始的学习序列如图 2(a) 所示。将学习者和课程作为节点, 交互作为连接, 构建如图 2(b) 所示的二部图 G 。为了弥补二部图丢失时序特征的问题, 进一步将交互记录中的时间信息融

入其中, 构建如图 2(c) 所示的时序图 G_t 。

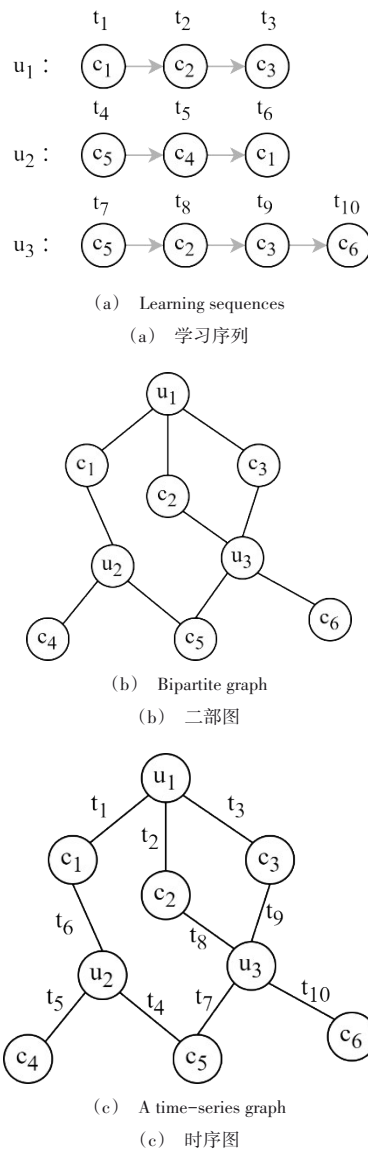


Fig. 2 A time-series graph example

图 2 时序图实例

为了便于表述, 表 1 总结了关键的数学符号及其描述。

3 模型构建

3.1 总体框架

本节将详细介绍所提出的融合多种时间关系的时序图模型。如图 3 所示, 该模型框架主要由 3 部分组成: ①实体与时间关系嵌入模块: 将所有学习者和课程嵌入为低维实值向量, 另外将时序特征分别以绝对时间嵌入、顺序时间嵌入和时间间隔嵌入; ②时序图表示模块: 首先将嵌入模块得到的时间嵌入作为重要特征放入二部图中构建时序图, 再通过图卷积和注意力机制捕捉学习者与课程之间的高阶信息, 并及时更新学习者、课程和时间关系的嵌入

Table 1 Mathematical symbols and their descriptions

表 1 数学符号及其描述

符号	描述
U, C, T	学习者集、课程集、时间集
$U \in \mathbb{R}^{m \times d}, C \in \mathbb{R}^{n \times d}$	学习者嵌入矩阵和课程嵌入矩阵
G	学习者—课程二部图
G_t	学习者—课程时序图
S_{ij}	学习者 u_i 的交互序列
$e_{u_j}, e_{c_i}, e_{t_l}$	学习者 u_j 的嵌入, 课程 c_i 的嵌入, 时间戳 t_l 的嵌入
e_{At}	绝对时间嵌入
e_t^m	月份嵌入
e_t^d	日期嵌入
e_{Pt}	顺序时间嵌入
e_{It}	间隔时间嵌入
d	嵌入维度
W_1, W_2, W_3, W^Q, W^K	模型可训练参数

向量;③预测模块:将上一模块计算的学习者和课程表征结合负采样得到最终的推荐概率,并通过二元交叉熵损失函数迭代更新。

3.2 实体与时间关系嵌入模块

为了便于模型的输入,首先将所有特征进行编码。除学习者和课程两个实体的初始化向量外,本文将详细介绍3种时间关系的编码方式。

3.2.1 学习者与课程嵌入

使用两个嵌入层将学习者和课程id映射成低维向量,得到初始化的嵌入表示,以便于后续的图卷积操作。令 $U \in \mathbb{R}^{m \times d}$ 表示用户嵌入层生成的初始化用户嵌入检索矩

阵, e_{u_j} 为学生 u_j 的嵌入向量。其中, m 为学生个数, d 为嵌入维度。令 $C \in \mathbb{R}^{n \times d}$ 表示课程嵌入层生成的课程嵌入检索矩阵, e_{c_i} 为课程 c_i 的嵌入向量。其中, n 为课程个数, d 为嵌入维度。学习者与课程的向量是随机初始化的,随着损失函数一起迭代更新。

3.2.2 绝对时间嵌入

本文考虑3种有用的时间关系。首先是绝对时间关系,绝对时间是对每次交互当下的时间戳建模。即使两个学习者学习了相同课程并且顺序相同,其交互绝对时间的不同体现了学习者不同的学习周期。在模型中考虑了月和日的时间粒度。时间戳 t_l 的绝对时间向量表示为:

$$e_{At_l} = e_{t_l}^m + e_{t_l}^d \quad (1)$$

其中, $e_{t_l}^m, e_{t_l}^d$ 分别是月从月的嵌入检索矩阵 M^m 和 M^d 中检索时间戳 t_l 对应的嵌入向量, e_{At_l} 是时间戳 t_l 的绝对时间向量。月和日的嵌入检索矩阵是随机初始化的,随着损失函数一起迭代更新。

3.2.3 顺序时间嵌入

顺序时间有助于揭示课程的排期等信息。为了捕获序列中项目的相对顺序,为每一个位置加入一个位置向量,模型采用 Vaswani 等^[26]提出的位置编码方法,不引入额外参数,可以外推到比预定义的固定长度更长的序列长度。该方法定义如下:

$$e_{Pt(pos, 2i)} = \sin\left(\frac{pos}{10000^{2i/d}}\right) \quad (2)$$

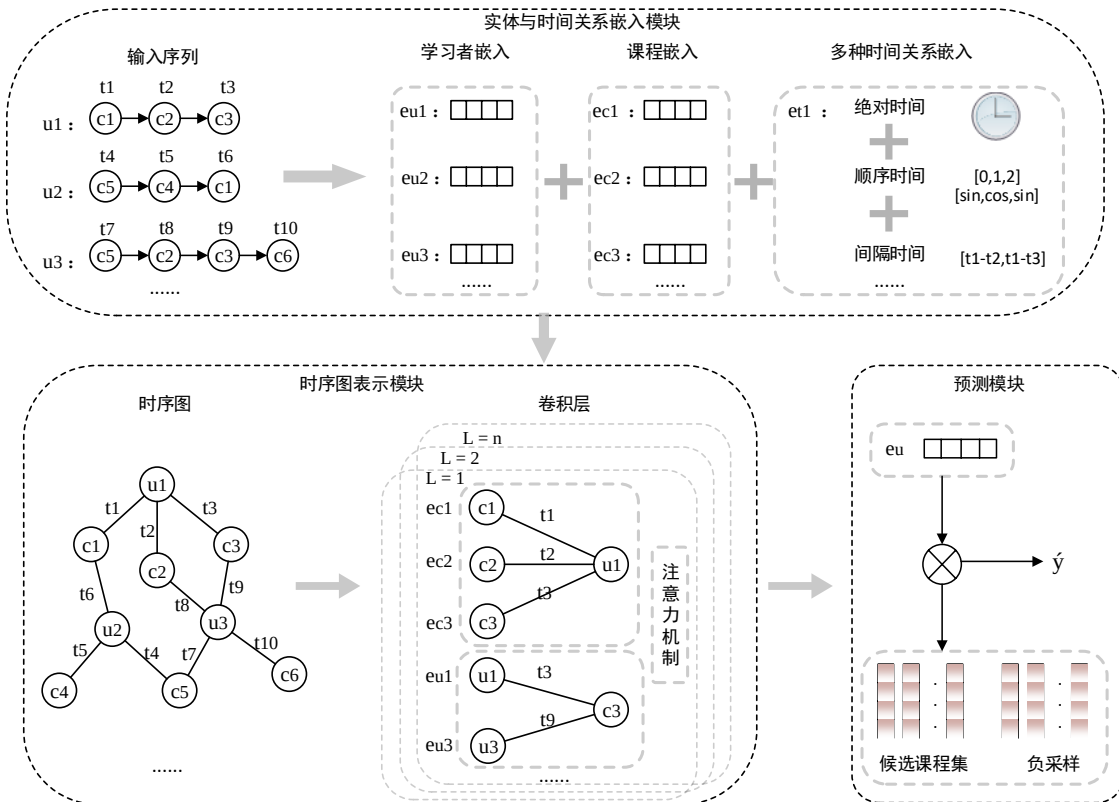


Fig. 3 Model framework

图 3 模型框架

$$e_{pt(pos, 2i+1)} = \cos\left(\frac{pos}{10\,000^{2id}}\right) \quad (3)$$

其中, pos 表示词的位置, i 表示维度, d 表示模型维度大小。也即是说, 位置编码的每个偶数维度对应一个正弦曲线, 奇数维度对应一个余弦曲线。相对时间向量是确定的, 不随着损失函数一起迭代更新。

3.2.4 间隔时间嵌入

学过的课程对下一次推荐的影响与其学习时间戳和当前推荐时间的时间间隔有关。尽管一般认为最近的记录对下一次推荐的影响会更大, 但仍需要通过个性化的建模探索其中的缘由。时间戳每隔一秒都不一样, 对所有不同的时间间隔直接建模计算量太大。本文采用离散化方法将连续数字特征转换为分类特征^[27]。具体来说, 将时间间隔映射为 $[0, 1)$, $[1, 2)$, $[2, 4)$, $\dots$, $[2k, 2k+1)$ 按指数增长的时间段。每一个时间段为一个嵌入向量, 构建随模型更新的时间间隔检索矩阵 INT^{k+1} 。之后, 直接查找分类时间特征, 得到时间间隔嵌入 e_{it} 。

3.3 时序图表示模块

在得到多特征嵌入后, G_t 为了弥补过往学习者一课程二部图丢失时序特征的缺点, 模型将3种时间关系的嵌入融入图中, 构成时序图。图模型可以学习到节点的潜在信息, 得到包含高阶信息的学习者一课程节点嵌入。

传统图神经网络在计算节点嵌入过程中主要包括两个阶段: 聚合和传播。在聚合阶段, 图神经网络聚合每个节点的邻居节点表示, 通常涉及邻居节点表示的加权平均或连接。在传播阶段, 将聚合的邻居节点表示与上一层的节点表示聚合在一起以更新节点表示, 并依次一层一层传播下去。传播可以重复多次, 以便学习到更高阶的信息。上述两个阶段可表述如下:

$$e_{uj}^{(l)} = \text{AGGREGATE}\left(e_{ui}^{(l-1)}, u_i \in N(u_j)\right) \quad (4)$$

$$e_{uj}^{(l)} = \text{PROPAGATE}\left(e_{uj}^{(l-1)}, e_{uj}^{(l)}\right) \quad (5)$$

其中, $N(u_j)$ 为 u_j 的邻居节点。

在传播阶段, 本文采用一个基本的线性传播层来计算第 l 层的节点表示, 公式如下:

$$e_{uj}^{(l)} = \sigma\left(W_l\left(e_{uj}^{(l-1)}\right)\right) \quad (6)$$

其中, W_l 是模型的可训练参数, σ 为激活函数, 本文选取 ReLU 作为激活函数。以学习者节点表示 e_u 为例, 课程向量 e_c 的聚合与传播是一样的, 因此不再赘述。本文模型讨论的重点是聚合过程。

之前的一些研究已经设计了多种聚合器, 例如加权聚合器和平均聚合器。然而, 大多数都忽略了邻居节点交互的时间信息。本文模型在加权聚合器基础上进行改进, 融入3种时间关系。具体来说, 在此过程中, 邻居节点将被有条件地激活。以第 l 层为例:

$$e_{uj}^{(l+1)'} = \sum_{u_i \in N(u_j)} \alpha_{i,j} e_{ui}^{(l)} \quad (7)$$

其中, $\alpha_{i,j}$ 表示课程 c_j 沿着路径传播到学习者 u_j 的可能性。根据前文的分析, $\alpha_{i,j}$ 不仅取决于邻居与中心节点之间的相似度, 而且取决于交互中的时序特征。因此, 计算方法如下:

$$\alpha_{i,j} = \frac{\exp(\alpha'_{i,j})}{\sum_{i \in N(u_j)} \exp(\alpha'_{i,j})} \quad (8)$$

$$\alpha'_{i,j} = \left(W^Q e_{uj}^{(l+1)} + e_{At_i} + e_{Pt_i} + e_{It_i}\right) \cdot \left(W^K e_{ui}^{(l)}\right) \quad (9)$$

其中, $\alpha'_{i,j}$ 是由中心节点嵌入加上3种时间关系嵌入与邻居节点嵌入加上3种时间关系嵌入点积得来的。不同的邻居节点对应不同的时间嵌入, 以此捕捉交互互率中时序特征对推荐结果的影响。 W^Q 和 W^K 为模型训练参数。

在计算自注意力之后, 为了避免过拟合、训练过程不稳定(如梯度消失)以及需要更多训练时间等问题, 与文献[28]一样, 本文也采用层归一化、残差连接和随机丢弃技术来解决这些问题。

$$e_{uj}^{(l+1)} = \text{LayerNorm}\left(\text{Dropout}\left(\text{Relu}\left(W_2 e_{uj}^{(l+1)'}\right) W_3\right) + e_{uj}^{(l+1)'}\right) \quad (10)$$

至此, 得到了 l 阶学习者的表示(课程的表示获取方式相同)。为了获得高阶信息, 以更好地评估学习者的学习偏好是至关重要的, 模型堆叠更多包含时序特征的图神经网络层。

值得注意的是, 当一门课程的学习者过多或一名学生学习过的课程过多时, 聚合会导致学习者和课程表示过度平滑。为了解决该问题, 模型采用抽样方法, 选取固定数量的邻居集来限制邻居的大小。

3.4 模型预测与训练

在堆叠多层卷积后, 获得学习者和课程的嵌入向量。使用一个简单的内积运算来预测学习者对目标课程的偏好:

$$\hat{y}_{c|uj} = e_{uj} \cdot e_{cj} \quad (11)$$

为了优化模型参数, 使用二元交叉熵作为损失函数。公式如下:

$$L = -\left[\sum_{(u, c, t) \in p^+} \hat{y}_{c|uj} + \sum_{(u, c, t) \in p^-} \log(1 - \hat{y}_{c|uj})\right] \quad (12)$$

其中, p^+ 表示正例, p^- 表示负例。对于训练集所有的历史学习者交互, 训练的目的是优化模型参数, 使得每次交互时, 在 t 时刻 u_j 和 c_j 的输出结果接近1, 并且 u_j 和 c_j' 的输出结果尽可能接近0。所有参数及嵌入都通过反向传播以端到端的训练方式学习。

4 实验与结果分析

本文研究的主要问题如下:

RQ1: 与其他先进的基线模型相比, 本文模型性能表现如何?

RQ2:3种时间关系的融合是否能提升推荐模型的性能表现?同时考虑3种关系是否比单独使用一种或两种更好?

RQ3:嵌入维度和传播层数两个超参数对模型性能有什么影响?

4.1 数据集

实验在一个真实的数据集 MOOCCourse 上进行, MOOCCourse 是从学堂在线平台获得的离线数据集。数据时间范围为2016年10月1日—2018年3月31日。本文主要对 MOOC 数据中的学生、学生选择的课程以及选课时间进行分析。其中,每个用户都学习了两门以上的课程。数据总共包含 82 535 名学生、1 302 门课程、458 454 个学生—课程交互行为,用户平均交互数为 5.55。本文根据记录时间从前到后将该数据集按 8:2 的比例分为训练集和测试集。

4.2 基线模型

本文模型将与以下相关的基线模型进行比较,基线模型主要分为3类:

(1)传统的推荐算法。早期的推荐算法思想简单,实现快速,是后续推荐算法的基础,也是大多数先进推荐模型都需要比较的参照模型。

POP(Popularity-based Recommendation):该方法根据课程的受欢迎程度进行推荐,以推荐最受欢迎的课程给学生。该方法简单直观,也是推荐中常用的基线模型。

BPR(Bayesian Personalized Ranking):该方法是一种经典的基于贝叶斯个性化排序的 top-k 项目推荐方法。其通过成对排序损失对候选项目的顺序进行建模,而不考虑序列模式^[29]。

(2)基于序列模型的推荐方法。近年来,深度推荐算法开始成为主流。基于序列的模型根据学习过程的先后顺序输入神经网络中,与本文考虑顺序时间的想法类似。

GRU4REC:该方法使用基于循环神经网络的门控单元从会话序列中识别物品的转移概率,从而生成个性化的推荐列表。该模型首次将循环神经网络运用于会话推荐中^[30]。

LETCTCR (Learning-motivation-boosted Explainable Temporal Point Process Model for Course Recommendation):该方法将交互记录中的时间戳视为绝对时间,交互记录的先后顺序视为相对时间。通过注意力机制结合4种学习动机计算不同历史交互记录对推荐结果的不同贡献,以此提高模型的推荐精准性^[12]。

(3)基于图模型的推荐算法。凭借图卷积可以获取高阶信息的优势,图模型在推荐领域应用广泛。

PinSage:该方法基于有效的随机游走进行卷积,是一个利用图生成用户/项目特征表示的推荐模型。该模型是第一次将图卷积网络应用于工业级推荐系统的图算法^[20]。

TP-GNN (Top-N Personalized Recommendation With

Graph Neural Networks in MOOCs):该方法是一种先进的 MOOC 课程推荐方法,其通过图卷积网络捕获课程之间的高阶语义关系,并使用注意力机制生成最终的课程表示^[5]。

4.3 评价指标

为了便于与之前的工作进行比较,本文采用了两个评价指标,这两个指标在 Top-K 推荐和个性化排名任务中得到了广泛应用。

(1)召回率 R@K。表示前 k 个推荐项目中预测正确的项目占所有正例的百分比。R@K 定义如下:

$$R@k = \frac{|\text{pred}_u@k \cap \text{pos}_u|}{|\text{pos}_u|} \quad (13)$$

其中, $\text{pred}_u@k$ 是推荐结果中前 k 个给学生推荐的课程, pos_u 是学生真实喜欢的正例。

(2)归一化折损累积增益 NDCG@k。也称为对顺序敏感的精确率。首先需要计算折损累积收益 DCG:

$$DCG@k = \sum_i^k \frac{r(i)}{\log_2(i+1)} \quad (14)$$

其中, $r(i)$ 表示第 i 个物品的分数,预测正确为 1,错误为 0。接下来需要用 IDCG@k 进行归一化:

$$IDCG@k = \sum_i^k \frac{1}{\log_2(i+1)} \quad (15)$$

最后得到 NDGC@k:

$$NDCG@k = \frac{DCG@k}{IDCG@k} \quad (16)$$

4.4 实验设置

本文使用 GeForce RTX 4090 24 G GPU 和 Pytorch 框架实现了 TMPCR,对于所有参数(包括用户向量、课程向量、时间向量、权重等)进行随机初始化。另外使用 Adam 作为模型的优化器,并设置权重衰减率为 0.001, batch-size 大小为 128。对于超参数,设置学习率为 0.001,用户向量、课程向量、时间向量维度为 256,传播层数为 2,每层采样数量为 6。在测试时为每一个正例匹配 99 个负例。所有基线模型的参数与其原论文中的设定一致,以确保其获得最佳性能。

4.5 实验结果分析

不同模型在数据集上的实验结果比较如表 2 所示。

由表 2 可见,在 MOOCCourse 数据集上, TMPCR 在不同评价指标上的推荐性能均优于其他基线方法。与最优的基线模型(LETCTCR)相比, R@5、R@10、R@15、R@20 分别提高了 6.58%、2.88%、2.76% 和 2.81%, NDCG@5、NDCG@10、NDCG@15、NDCG@20 分别提高了 2.06%、2.39%、2.61% 和 0.93%。从实验结果中可以得出以下结论:

(1)在传统的推荐算法中,基于流行度的推荐算法(POP)和基于个性化排名的算法(BPR)表现最差。首先, POP 算法主要依据课程出现频率进行推荐,该方法虽然简单易行,但忽略了学生个人的兴趣和偏好,无法满足个性化学习的需求。其次, BPR 算法通过矩阵分解技术来建模

Table 2 Comparison of experimental results of different models on the dataset

表2 不同模型在数据集上实验结果比较 (%)

Models	R@5	R@10	R@15	R@20	NDCG			
					@5	10	15	20
POP	7.28	12.43	19.34	25.48	5.80	7.79	9.86	11.67
BPR	21.67	34.54	39.05	42.46	13.97	18.04	20.33	19.42
GRU4REC	24.87	34.04	38.05	43.56	14.30	18.04	21.01	20.33
LETGR	<u>32.07</u>	<u>52.40</u>	<u>64.57</u>	<u>74.68</u>	<u>32.49</u>	<u>39.32</u>	<u>45.21</u>	<u>48.59</u>
PinSage	24.99	37.83	44.96	49.79	14.70	19.90	22.19	23.58
TP-GNN	26.35	38.82	46.11	50.85	15.51	20.83	22.47	23.91
TMPCR	34.18	53.91	66.35	76.78	33.16	40.26	46.39	49.04
Improve	6.58	2.88	2.76	2.81	2.06	2.39	2.61	0.93

学生的偏好,但学习者的学习路径往往是非线性和动态变化的,BPR算法未能充分考虑这一点。传统算法重点关注学习者与课程之间的交互行为,模型实现非常简单,但在处理复杂数据时的拟合能力较差。相比之下,TMPCR模型采用图模型的方法,应对复杂数据的能力较强。

(2)基于图模型的方法比纯粹使用序列模型的方法效果好。无论是本文提出的TMPCR还是PinSage和TP-GNN性能都优于GRU4Rec。尽管GRU4Rec通过考虑学习者行为序列的时间动态来捕捉学习者的短期兴趣,但无法捕捉不同学习者交互间的潜在信息,学习到的特征太少。TMPCR通过聚合其多层邻居的表示来更新每个学习者或课程的嵌入,使其能够学习到学习者—课程间的高阶信息,从而在数据集学习到更多特征。

(3)融入时序特征的时序图比普通图模型效果更好。相比PinSage和TP-GNN,本文模型TMPCR没有在构建二部图时摒弃交互中的时间戳,反而将其扩充为3种时间关系,以捕捉学习者动态兴趣、学习周期和课程间依赖关系等重要信息。在邻居聚合过程中,依据3种时间关系的嵌入给每个邻居节点分配不同权重,以得到更加精确的节点表示。

(4)本文模型TMPCR同时利用时序特征和学习者—课程高阶信息,取得了最好效果。LETGR也考虑了时序特征,尽管其分析了4种学习者选择课程时的学习动机,一定程度上提高了序列模型的表征能力,但这4种学习动机无法全面涵盖交互中的隐藏信息,也只能捕捉单一层级的关系。本文模型TMPCR在保留时序特征和图模型优势的基础上,进一步将时间关系扩充为3种,以捕捉学习者—课程交互记录中的高阶关系,因而取得了最好的效果。

4.6 消融实验

为了验证所提出的3种时间关系的有效性,本文实验分析了多种时间关系对TMPCR性能的影响。其中既考虑了单一时间关系,又考虑了它们的组合。具体来说,本文选择3种时间关系来建模学习者—课程交互记录中的时序特征,包括绝对时间(Absolute Time, AT)、顺序时间(Position Time, PT)、间隔时间(Interval Time, IT)。为了研究不同时间关系带来的性能提升,使用3个单一时间关系及其

两两组合对二部图进行表征学习。实验结果如图4、图5所示。

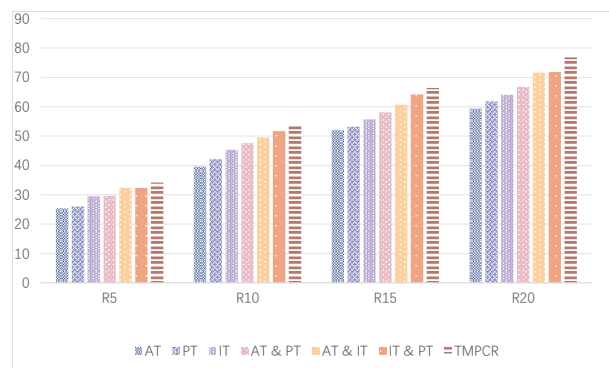


Fig. 4 Results of ablation experiments at R@K

图4 消融实验在R@K指标下的结果

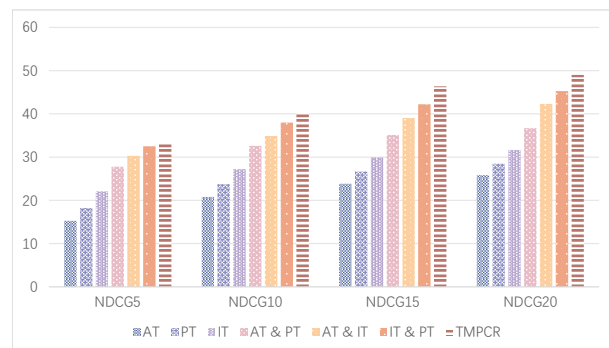


Fig. 5 Results of ablation experiments at NDCG@K

图5 消融实验在NDCG@K指标下的结果

由图可以看出,提出的所有时间关系都对最终TMPCR模型的推荐性能有利,这也充分证明了本文所提出3种时间关系的有效性。从单个时间关系来看,时间间隔最为重要,绝对时间影响最小。这也证明了时间间隔能反映出学习者兴趣的动态变化,进而精确建模学习者与课程嵌入。绝对时间的影响较弱主要归因于MOOCourse数据集比较稀疏,学习者平均交互记录较少,难以体现学习者学习周期等信息,但利用绝对时间仍能提升模型预测表现。从多个时间关系组合来看,不同时间关系结合在一起比单一使用一种时间关系的效果好,证明了融合3种时间关系对于提升模型性能并不冲突,一起使用效果最好。

4.7 超参数分析

由于嵌入维度和传播层数两个超参数能对模型结果产生巨大影响,为了验证本文设定的超参数的合理性,通过控制变量的方法在数据集上进行分析,即只改变分析变量而模型中的其他参数相同。为了使结果展示直观,这里选取R@10、R@20、NDCG@10、NDCG@20作为对比指标。

首先,研究了嵌入维度对TMPCR性能的影响,将嵌入维度设置为64、128、256和512。图6中的结果显示,随着维度的增加,TMPCR的表征能力增强,性能逐渐提升。然而,增加的幅度是递减的,当嵌入维度设置为256时,性能达到最高值。维度低于256时模型的学习能力较弱,高于256时模型会产生过拟合,训练时间也成倍增长。因此,为

了兼顾时间效率和性能,本文将TMPCR的嵌入维数设置为256。

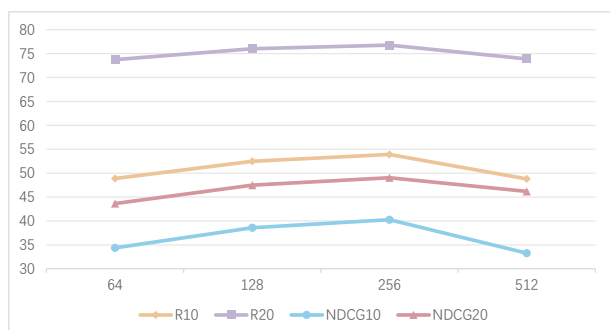


Fig. 6 Experimental results of embedding dimension hyperparameter analysis

图6 嵌入维度超参数分析实验结果

其次,研究了传播层数对TMPCR性能的影响,将传播层数设置为1、2、3。图7中的结果显示了不同层聚合下TMPCR的性能。可以看出,当层数从1层增加到2层时,模型性能有所提高。这是因为1层传播其实相当于模型退化为序列模型,没有获取学习者与课程间的高阶信息。反之,当层数为3时,模型性能略有下降。过多的全局信息会导致表征的过度平滑,即节点间的嵌入差异变小,同时增加了时间复杂度和参数数量。因此,本文将TMPCR的传播层数设置为2。

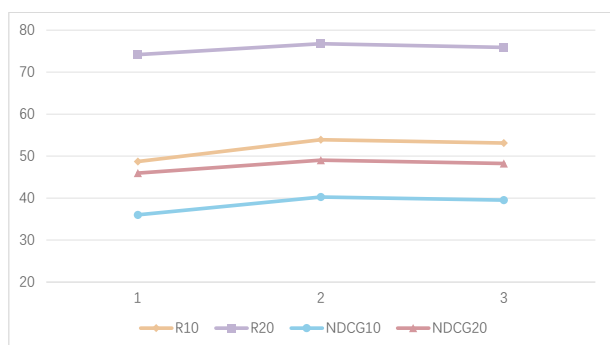


Fig. 7 Experimental results of propagation layer hyperparameter analysis

图7 传播层数超参数分析实验结果

5 结语

本文充分利用图模型和时序特征的优势,提出一种融合多种时间关系的时序图课程推荐算法TMPCR,摒弃了以往大部分课程推荐模型未考虑时序特征或者认为所有历史交互同等重要的思想。TMPCR不仅考虑了学习记录中课程的顺序时间关系,而且提出了绝对时间关系和间隔时间关系。采用不同的嵌入方法将3种时间关系融合到二部图中,构建时序图。模型利用了注意力机制和3种时间关系为历史交互分配不同的聚合权重,以此捕捉学习者兴趣的变化,并能更准确地对学习者和课程表征进行建模。在MOOCourse数据集上的大量实验结果表明,考虑

多种时间关系能显著提升推荐性能。在消融实验中,发现间隔时间对模型性能的提升大于顺序时间和绝对时间。

未来的研究计划引入更多维度的外部特征,例如课程涉及的知识点、课程类别等关键信息。研究将深化对学习者的兴趣随时间变化的分析,从对单一课程的兴趣变化进一步细化为对特定知识点或课程类别的多重兴趣演变。通过这些辅助信息的引入,期望能够更准确地捕捉并反映学习者在真实学习环境下的认知变化过程。

参考文献:

- [1] LIU Y, DONG Y Q, JIA R, et al. Hierarchical and phased attention network model for personalized course recommendation[J]. Journal of Computer Applications, 2023, 43(8): 2358-2363.
刘源,董永权,贾瑞,等.面向个性化课程推荐的分层分期注意力网络模型[J]. 计算机应用, 2023, 43(8): 2358-2363.
- [2] DENG W, ZHU P, CHEN H, et al. Knowledge-aware sequence modeling with deep learning for online course recommendation[J]. Information Processing & Management, 2023, 60(4): 103377.
- [3] VEDAVATHI N, KM A K. E-learning course recommendation based on sentiment analysis using hybrid Elman similarity[J]. Knowledge-based Systems, 2023, 259: 110086.
- [4] NILASHI M, MINAEI-BIDGOLI B, ALGHAMDI A, et al. Knowledge discovery for course choice decision in Massive Open Online Courses using machine learning approaches[J]. Expert Systems with Applications, 2022, 199: 117092.
- [5] WANG J, XIE H, WANG F L, et al. Top-N personalized recommendation with graph neural networks in MOOCs[J]. Computers and Education: Artificial Intelligence, 2021, 2: 100010.
- [6] KHAN Z Y, NIU Z, SANDIWARNO S, et al. Deep learning techniques for rating prediction: a survey of the state-of-the-art[J]. Artificial Intelligence Review, 2021, 54: 95-135.
- [7] SYMEONIDIS P, MALAKOUDIS D. Multi-modal matrix factorization with side information for recommending massive open online courses[J]. Expert Systems with Applications, 2019, 118: 261-271.
- [8] JING X, TANG J. Guess you like: course recommendation in MOOCs[C]//Proceedings of the International Conference on Web Intelligence, 2017: 783-789.
- [9] YUN Y, DAI H, ZHANG Y P, et al. State-of-the-art survey on personalized learning path recommendation[J]. Journal of Software, 2022, 33(12): 4590-4615.
云岳,代欢,张育培,等.个性化学习路径推荐综述[J]. 软件学报, 2022, 33(12): 4590-4615.
- [10] KHALID A, LUNDQVIST K, YATES A. A literature review of implemented recommendation techniques used in massive open online courses[J]. Expert Systems with Applications, 2022, 187: 115926.
- [11] JIANG W, PARDOS Z A. Time slice imputation for personalized goal-based recommendation in higher education[C]//Proceedings of the 13th ACM Conference on Recommender Systems, 2019: 506-510.
- [12] ZHANG W, ZHOU X, ZENG X, et al. Learning-motivation-boosted explainable temporal point process model for course recommendation[J]. IEEE Access, 2024, 12: 93876-93888.
- [13] WANG X, MA W, GUO L, et al. HGNN: hyperedge-based graph neural network for MOOC course recommendation[J]. Information Processing & Management, 2022, 59(3): 102938.

- [14] ZHAO Y, JIANG F, PANG Y, et al. EduLGCL: local-global contrastive learning model for education recommendation[J]. Knowledge-based Systems, 2024, 286: 111357.
- [15] MA W, CHEN W, LU L, et al. Integrating learners' knowledge background to improve course recommendation fairness: a multi-graph recommendation method based on contrastive learning[J]. Information Processing & Management, 2024, 61(4): 103750.
- [16] YUAN H, YU H, GUI S, et al. Explainability in graph neural networks: a taxonomic survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(5): 5782–5799.
- [17] GUO Q, ZHUANG F, QIN C, et al. A survey on knowledge graph-based recommender systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(8): 3549–3568.
- [18] XU Y, LI T, YANG Y, et al. An adaptive category-aware recommender based on dual knowledge graphs[J]. Information Processing & Management, 2024, 61(3): 103636.
- [19] WANG G, YIN K. Adaptive graph convolutional recommendation algorithm integrating user social relationships [J]. Application Research of Computers, 2024, 41(2): 482–487.
王光, 尹凯. 融合用户社交关系的自适应图卷积推荐算法[J]. 计算机应用研究, 2024, 41(2): 482–487.
- [20] YING R, HE R, CHEN K, et al. Graph convolutional neural networks for web-scale recommender systems [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018: 974–983.
- [21] WANG X, HE X, WANG M, et al. Neural graph collaborative filtering [C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019: 165–174.
- [22] HE X, DENG K, WANG X, et al. LightGCN: simplifying and powering graph convolution network for recommendation [C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020: 639–648.
- [23] WANG R, LOU J, JIANG Y. LightGCAN: a lightweight graph convolutional attention network for user preference modeling and personalized recommendation [J]. Expert Systems with Applications, 2023, 232: 120741.
- [24] GONG J, ZHAO Y, ZHAO J, et al. Personalized recommendation via inductive spatiotemporal graph neural network [J]. Pattern Recognition, 2024, 145: 109884.
- [25] CHEN X, LIANG Q, CHEN Y, et al. Cognitive-based knowledge learning framework for recommendation [J]. Knowledge-based Systems, 2024, 287: 111446.
- [26] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6000–6010.
- [27] ZHOU C, BAI J, SONG J, et al. ATRank: an attention-based user behavior modeling framework for recommendation [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2018: 4564–4571.
- [28] LI J, WANG Y, MCAULEY J. Time interval aware self-attention for sequential recommendation [C]//Proceedings of the 13th International Conference on Web Search and Data Mining, 2020: 322–330.
- [29] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback [C]//Proceedings of the Twenty-fifth Conference on Uncertainty in Artificial Intelligence, 2009: 452–461.
- [30] HIDASI B, KARATZOGLOU A, BALTRUNAS L, et al. Session-based recommendations with recurrent neural networks [DB/OL]. <https://arxiv.org/abs/1511.06939>.

(责任编辑:黄健)