

知识图与行为图混合嵌入的学习者偏好预测

李晓光¹, 宫磊¹, 李晓莉¹, 张昕¹, 于戈²

(1. 辽宁大学信息学院, 辽宁 沈阳 110036; 2. 东北大学计算机科学与工程学院, 辽宁 沈阳 110169)

摘要: 为了解决知识推荐模型中学习者偏好预测不准确、结构信息利用不充分等问题, 针对学习者偏好预测模型中知识结构与学习者行为结构, 提出了知识图与行为图混合嵌入的学习者偏好预测模型。首先, 考虑使用图卷积的方式拟合结构信息, 将图卷积扩展至知识图以及行为图, 获取学习者整体性学习规律与个性化学习规律。然后, 模型利用知识结构与行为结构的差异性拟合学习者个性化偏好, 采用循环神经网络将学习者的过程性偏好进行编码与解码, 获得学习者偏好分布。在真实数据集上的实验结果表明, 所提模型对拟合学习者偏好效果良好。

关键词: 知识图; 行为图; 图卷积网络; 知识推荐

中图分类号: TN92

文献标识码: A

DOI: 10.11959/j.issn.1000-436x.2021125

Learner preferences prediction with mixture embedding of knowledge and behavior graph

LI Xiaoguang¹, GONG Lei¹, LI Xiaoli¹, ZHANG Xin¹, YU Ge²

1. School of Information, Liaoning University, Shenyang 110036, China

2. School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China

Abstract: To solve the problems of inaccurate prediction of learner preference and insufficient utilization of structural information in the knowledge recommendation model, for the knowledge structure and learner behavior structure in the learner's preference prediction model, the model of learner preferences prediction with mixture embedding of knowledge and behavior graph was proposed. First, considering using graph convolution network (GCN) to fit structural information, GCN was extended to knowledge graph and behavior graph, the purpose of which was to obtain learners' overall learning pattern and individual learning pattern. Then, the difference between knowledge structure and behavior structure was used to fit learners' individual preferences, and recurrent neural network was used to encode and decode learners' preferences to obtain the distribution of learners' preference distribution. The experimental results on the real datasets demonstrate that the proposed model has a good effect on predicting learner preferences.

Keywords: knowledge graph, behavior graph, GCN, knowledge recommendation

1 引言

大数据时代促进了个性化教育的发展, 能准确、高效地为学习者提供定制化的教育资源, 改善学习体验, 提高学习效果, 是当前智慧教育迫切需要解决的问题之一。传统意义上的个性化推荐, 如

基于内容的推荐技术^[1]、协作过滤技术^[2-3]、基于深度学习的推荐技术^[4-5]等, 已经广泛用于商业、新闻等领域, 并取得了大量研究成果。与这些领域不同, 教育领域的知识资源推荐具有很强的结构性, 即推荐知识的结构性和学习行为的结构性。推荐知识的结构性是指知识之间并非独立的, 而是存在一定

收稿日期: 2021-01-06; 修回日期: 2021-03-08

通信作者: 李晓莉, xlli7396@sohu.com

基金项目: 国家自然科学基金资助项目 (No.U1811261); 辽宁省教育厅高等学校基本科研基金资助项目 (No.LWF201705)

Foundation Items: The National Natural Science Foundation of China (No.U1811261), Basic Scientific Research Project of Higher Education Institutions of Education Department of Liaoning Province (No.LWF201705)

的结构关系, 这种结构关系在一定程度上约束了知识推荐的结构。例如, 数据结构中表包含线性表和非线性表, 而线性表包含顺序表和链表。那么在学习链表的结构时, 按照这种自上而下的顺序学习更加有效。从本质上说, 学习者偏好是在不同的知识点上转移扩散的, 而这种转移的关系可以从学习者的历史学习行为中得到, 其通常是多向的和非独立的, 表达应为行为图形式, 而非行为序列形式。

目前, 对于这种具有结构性特点的推荐, 主要方法是基于图神经网络的推荐技术。由于图注意力网络 (GAT, graph attention network)^[6]和图卷积网络 (GCN, graph convolution network)^[7]可以学习和表达图特征, 研究人员已尝试在推荐系统中融合使用图神经网络 (GNN, graph neural network)^[8-10]。例如, 将学习者与资源的交互数据转换成图数据, 通过 GNN 学习图中丰富的特征信息, 并取得好的效果。文献[8]为每一个会话构建子图, 利用 GNN 学习所有项目的嵌入表示, 将得到的特征向量输入循环神经网络 (RNN, recurrent neural network) 中生成与会话对应的嵌入向量, 进而得到项目的预测概率。文献[9]则将学习者、资源以及数据之间的交互转换为二部图模式, 利用 GNN 进行偏好传递, 进而对学习者和资源进行图自动编码, 以获取相应的得分。文献[10]从学习者与资源交互的评分矩阵中抽取封闭的子图, 使用关系图卷积神经网络学习混合图模式下图的结构信息, 利用图的结构信息获得具有归纳性的预测网络。

尽管上述方法取得了较好的结果, 但对于知识推荐来说, 仍然存在局限性。首先, 传统的 GNN 仅考虑了图结构而没有真正利用边信息, 它们都忽略了图中边信息对学习者和资源偏好的影响。特别是在知识推荐中, 知识点之间的不同关系对学习者和资源偏好的影响有着重要的影响。例如, 数据结构中的哈希表、B 树以及 B+树 3 个知识点中, 哈希表与 B 树为相关关系, B+树与 B 树为递进关系。显然学习者在学习 B 树后, B+树的推荐概率要高于哈希表。其次, 知识的推荐是与行为结构和知识结构二者相关的, 但目前大多数方法仅将知识结构和行为结构信息进行简单的融合处理, 如连接、平均池化等, 忽略了 2 种结构对学习者和资源偏好的不同影响, 不能充分且准确地获得学习者嵌入表达。一般来说, 知识结构通常决定了学习者的整体学习规律, 而行为结构通常体现了学习者的个体学习规律。尽管个体学习规律因人

而异, 但对于学习者整体来说, 其呈现出整体学习规律。例如, 当学习者学习基础运算知识时, 如果从知识结构上来看, 整体上学习应按照加、减、乘、除顺序进行。但对于学习者个体来说, 如果该学习者处于学习新知识的初级阶段, 则应该适当忽略学习者的行为信息, 而根据知识结构为其推荐更多基础性的试题。反之, 如果该学习者已经熟练掌握了基础运算知识, 则应该根据该学习者的行为结构为其推荐更多的进阶性试题。

针对上述问题, 本文提出了一种基于知识图与行为图混合嵌入的知识推荐技术, 即知识图与行为图混合嵌入的学习者偏好 (KGBR, learner preference with mixture embedding of knowledge graph and behavior graph)。KGBR 将知识图作为知识结构信息, 在图卷积中引进知识图的边信息, 进而学习知识结构嵌入。KGBR 根据学习者偏好的多向转移特点, 构建学习者历史行为关系图, 结合相邻行为特征, 采用图注意力神经网络学习行为结构嵌入; 根据 2 种结构嵌入的差异性学习学习者偏好嵌入。KGBR 采用 RNN 对历史学习信息进行编码和解码, 得到学习者关于知识点的偏好分布, 用以推断出学习者在各知识点上的偏好程度, 从而为学习者推荐更具针对性的学习资源。

2 相关工作

矩阵分解^[11-13]是推荐系统研究中的经典方法。基本目标是将学习者项目评分矩阵分解为学习者矩阵和项矩阵, 分别代表基于潜在因素的学习者偏好或项目分布。矩阵分解也广泛用于知识推荐中。例如, 文献[14]基于矩阵分解, 提出了一种多层存储桶推荐 (MLBR, multi-layer bucketing recommendation), 用于 MOOC (massive open online course) 推荐。MLBR 将学习者向量更改为相同的长度尺寸, 并将它们分散到包含相似学习者的存储桶中, 这些学习者具有更多共同的课程。MLBR 在一定程度上提高了协作过滤的效率与准确率。针对 MOOC 上低满意度和孤独感导致的放弃这一问题, 文献[15]提出了一种针对 MOOC 的自适应推荐 (ARM, adaptive recommendation) 的解决方案。利用协作过滤提供类似学习者的明确信息, 并支持协作学习以减少孤独感。ARM 创造性地将协作过滤和时间序列相结合, 以提高建议的准确性。但是, 由于矩阵分解技术本身很难

拟合结构性，导致其在教育领域应用性较差。概率矩阵分解技术^[16-18]采用概率图形式来描述推荐模型，可以描述模型中各因素的结构关系，但其要求这种结构关系必须以先验知识的形式明确给出，限制了其适用性。另外，概率矩阵分解技术对于教育领域这种具有潜在变量维度高、不确定性程度大等特点的推荐来说，效果并不理想。

在知识推荐中，一种典型的学习者模型为贝叶斯知识追踪 (BKT, Bayesian knowledge tracing) 模型。BKT 模型是对学生知识点的变化进行追踪，进而用于教育资源推荐。例如，文献[19]通过添加一个学生节点来个性化原始 BKT 模型的先前学习参数。文献[20]提出了具有 3 种学习状态的贝叶斯知识追踪 (TLS-BKT, three learning states Bayesian knowledge tracing) 模型，在原有模型的基础上扩展了学习状态，使模型更好地模拟学习者实际使用的状态。文献[21]捕获学生的学习能力，并按固定的时间间隔将学生动态地分为具有相似能力的不同群体，将该信息与 RNN 进行结合，从而实现知识追踪及推荐。但总体来说，现有模型在构建基于知识点的学习者模型时，均没有综合考虑行为结构与知识结构的融合基础上的构建。

近年来，随着机器学习技术的快速发展，深度神经网络技术由于其泛化能力强、学习误差低、适应性广等特点，被广泛用于各领域的推荐系统。教育推荐的特点在于知识间结构以及学习者偏好的复杂多变性。知识图谱的出现很好地解决了知识点结构表达问题，而图神经网络则可以很好地拟合结构性特征并用于分类和推理。目前，研究者已提出大量的基于图神经网络的推荐系统。例如，文献[8]将学习者会话转化成图数据进行嵌入表示，接着将得到的会话嵌入表示输入 RNN 中得到长期偏好表示，结合当前偏好得到混合嵌入以生成推荐。文献[22]利用图卷积将学习者项目交互图中的高阶连接性嵌入以实现推荐。文献[9]提出了一种基于图结构数据的图卷积矩阵补全框架，将矩阵补全问题转化为链路预测问题。该框架设计了图卷积编码器，有效利用图结构中各个位置之间的权重分配，学习了学习者和资源的嵌入表示，通过双线性解码器提取学习者对各资源的概率分布情况。文献[10]在学习者项目评分矩阵中提取封闭子图，并将这些子图映射到其对应的评分，利用关系图卷积神经网络综合考虑

混合图模式中大量的结构信息，获得学习者与预测项目的嵌入表示从而实现评分预测。但是，目前基于图神经网络的推荐技术在融合知识图和行为图、考虑关系类型等方面，使其更适用于知识推荐，仍有待进一步研究。

3 知识图与行为图混合嵌入模型

3.1 基本概念与定义

本文定义知识资源为学习过程中的文章、试题、视频等。知识资源集合为 S ，其中 $s \in S$ 为资源项。知识点集合为 $K = \{k_1, k_2, k_3, \dots, k_m\}$ ，其中 $k_i \in \mathcal{R}^m$ 表示 K 中第 i 个知识点的特征向量。知识点关系类别集合为 $R = \{r_1, r_2, r_3, \dots, r_n\}$ ，其中 $r_i \in \mathcal{R}^l$ 表示 R 中第 i 种类别的知识点关系的特征向量。知识资源集合 S 所包含的知识图为 $KG = (K, R)$ 。知识资源集合 S 包含知识点集合 K 以及知识点关系类别集合 R 。给定资源项 $s \in S$ ，则定义顶点标记图 $KG_s = (K_s, R_s)$ 为其资源知识图，其中 $K_s \subseteq K$ ， $R_s \subseteq R$ 。学习者的学习行为序列 U 为学习者学习过程中的资源项序列，即 $U = (s_i | s_i \in S, i = 1, 2, \dots, l)$ ，其中 $s_i \in \mathcal{R}^m$ 。对于学习行为序列 U ，定义其行为图 $BR_U = (S_U, E_U)$ ，其中 $S_U \subseteq U$ ；对于 $\forall e \in E_U$ ， $\exists (s_i, s_{i+1}) = e$ ， $s_i \in U$ 且 $s_{i+1} \in U$ 。例如，如果学习行为序列为 $U = s_1 \rightarrow s_2 \rightarrow s_4 \rightarrow s_2 \rightarrow s_5 \rightarrow s_7 \rightarrow s_1 \rightarrow s_2$ ，则对应的行为图 BR_U 如图 1 所示，其中， d_{ji} 表示节点 s_j 到节点 s_i 的边数， d_i^in 表示节点 s_i 的入度数。学习者偏好为在知识集合 K 上的概率分布 $I_P = \{P(k_i) | \sum_{i=1}^m P(k_i) = 1\}$ 。

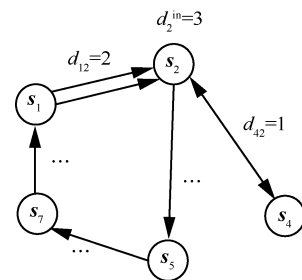


图 1 BR_U 的构建

3.2 KGBR 模型结构

KGBR 模型结构如图 2 所示。KGBR 模型的输入为学习者的行为图 BR_U 和行为图中资源项蕴含的资源知识图 KG_s 。KGBR 模型包括知识图卷积网络 (KGCN, knowledge graph convolution network) 层、行为图注意力网络 (RGAT, relation graph at-

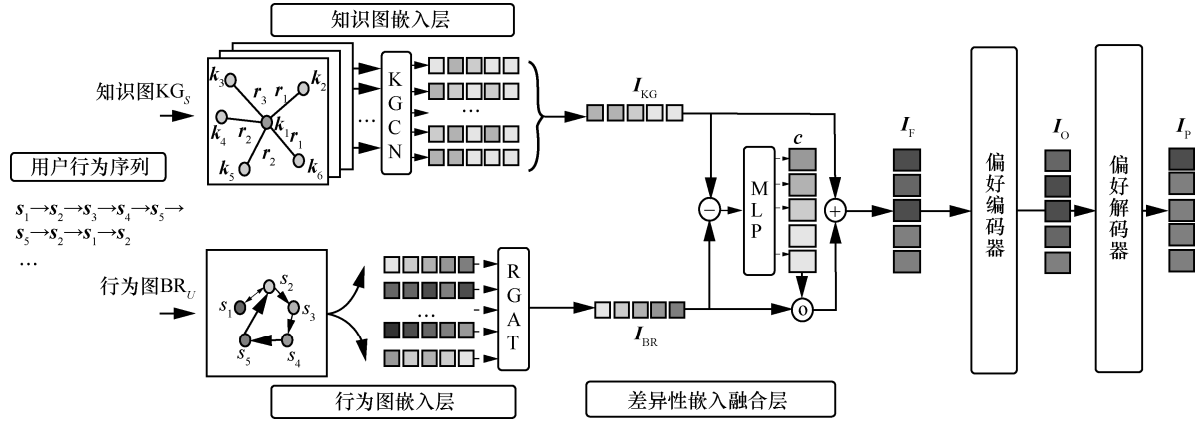


图2 KGBR 模型结构

tention network) 层和差异性嵌入融合 (DEF, differential embedding fusion) 层。KGCN 层和 RGAT 层为图神经网络, 其主要作用是获得 KG_S 和 BR_U 的嵌入表示 I_{KG} 和 I_{BR} 。其中, I_{KG} 是对知识结构的拟合, 表示整体性学习规律; I_{BR} 是对行为结构的拟合, 表示个体性学习规律, 强调了学习者个体性偏好所在。KGCN 层在嵌入时考虑到不同的关系对于学习特征表示具有不同的影响, 为各知识点间不同种类的关系引入了独立的特征嵌入矩阵以更好地学习各知识点之间结构关系。RGAT 层根据行为图中资源项相似性以及行为图结构信息, 利用注意力机制, 拟合了学习行为对知识偏好的影响, 进而表达基于学习行为的学习者个体性学习规律。在对中间特征 I_{KG} 和 I_{BR} 进行融合时, DEF 层考虑了知识结构和行为结构之间的差异性, 利用差异性融合矩阵 W_d 对二者的差值进行拟合以获得知识权重控制向量 c , 并对 2 种结构向量进行差异性融合。KGBR 层采用 RNN 作为学习者嵌入 I_F 编码器, 得到学习者偏好中间编码 I_O , 并采用 RNN 解码器获得学习者关于知识点集合 K 的概率分布 I_P 。

3.3 KGBR 模型描述

3.3.1 知识图嵌入

关于知识图的嵌入, 本文提出了一种基于图卷积神经网络的聚合器 KGCN。KGCN 综合考虑了来自邻居节点的特征信息以及知识图的边信息。也就是说, 利用图卷积将来自邻居节点及其对应的边信息进行了聚合。对于知识图 KG_S 中的节点 k_i , 其嵌入为

$$k_i^{(l+1)} = \frac{1}{\sqrt{|N_i|}} \sum_{k_j \in N_i} (k_j^{(l)} + W_{r_j} r_j) \quad (1)$$

其中, N_i 为知识图 KG_S 中节点 k_i 的邻居节点集合, $|N_i|$ 为邻居节点的个数; $k_i^{(l+1)}$ 和 $k_j^{(l)}$ 分别为知识点 k_i 迭代 $l+1$ 次的嵌入表示以及邻居节点 k_j 迭代 l 次的嵌入表示; $W_{r_j} \in \mathbb{R}^{m \times f}$ 为第 j 种关系 r_j 的特征嵌入矩阵。由于资源项与知识点是多对多的关系, 因此, 本文采用平均池化的方法得到关于资源项 s 的嵌入表示为

$$I_{KG}^s = \text{avg} \left(\sum_{k_i \in s} k_i^{(L)} \right) \quad (2)$$

其中, $k_i^{(L)}$ 为资源项 s 中包含的第 i 个知识点迭代 L 次的嵌入表示。实验结果表明, $L=2$ 时模型性能最好。本文将知识图 KG_S 中包含的所有资源项的嵌入表示进行求和, 得到学习者基于知识图的嵌入 I_{KG} 为

$$I_{KG} = \sum_{s \in U} I_{KG}^s \quad (3)$$

3.3.2 行为图嵌入

关于行为图的嵌入, 本文采用了一种基于 GAT^[6] 的行为图注意力网络 RGAT。RGAT 比较不同资源项间的相似度, 利用行为图发掘学习者行为序列中的各行为之间的联系, 以便为不同的边(关系)赋予不同的权重。对于行为图 BR_U 中的资源项 s_i , 注意力机制为

$$\alpha_{ij} = \frac{\exp(\gamma(\mathbf{a}^T [s_i || s_j]))}{\sum_{s_j \in N_i} \exp(\gamma(\mathbf{a}^T [s_i || s_j]))} \quad (4)$$

其中, γ 表示 LeakReLU 函数, $\mathbf{a} \in \mathbb{R}^{2m}$ 表示权重向量, $||$ 表示连接, α_{ij} 表示节点 s_j 对节点 s_i 的内容重要性。由于多个资源项会重复出现在行为图中,

因此本文为每条边分配归一化的权重,即用节点 s_j 到节点 s_i 的边数除以节点 s_i 的入度数表示不同邻居对节点 s_i 的结构重要性,如式(5)所示。

$$\beta_{ij} = \frac{d_{ji}}{d_i^{\text{in}}} \quad (5)$$

本文使用 α_{ij} 与 β_{ij} 共同组成注意力系数为

$$e_{ij} = \beta_{ij} \alpha_{ij} \quad (6)$$

则基于学习者行为图的嵌入 I_{BR} 为

$$I_{\text{BR}} = \sigma \left(\sum_{s_i \in \text{BR}_U} \sum_{s_j \in M_i} e_{ij} s_j \right) \quad (7)$$

其中, σ 为 sigmoid 函数, M_i 表示节点 s_i 的邻居集合。

3.3.3 差异性嵌入融合层

差异性嵌入融合层 DEF 融合 I_{KG} 与 I_{BR} , 以拟合学习者基于知识结构的整体性偏好和基于行为结构的个体性偏好, 具体如式(8)~式(10)所示。

$$X = I_{\text{BR}} - I_{\text{KG}} \quad (8)$$

$$c = \sigma(W_d X + b) \quad (9)$$

$$I_{\text{F}} = c \circ I_{\text{BR}} + I_{\text{KG}} \quad (10)$$

其中, X 表示学习者个体性学习规律与整体性学习规律的差异; c 表示知识权重控制向量, 控制个体性学习规律对偏好的影响; $W_d \in \mathcal{R}^{m \times m}$ 表示差异性融合矩阵; b 表示偏置; \circ 表示哈达玛乘积; I_{F} 表示学习者融合嵌入向量。本文也考虑了其他融合的方式。例如, 使用内积计算个体性学习规律与整体性学习规律的相似度, 然后使用门控机制控制 2 种表达对学习者偏好不同的重视程度。这种方式从整体上考虑了 2 种规律对学习者的影响, 但难以实现对个体性学习规律每一维的精细控制, 因此本文采用式(10)的方式进行融合。

3.3.4 偏好编码与解码

I_{F} 中蕴含学习者整体性学习规律与个体性学习规律的特征信息, 为了将其转化成学习者偏好分布向量, 本文首先使用 RNN 作为偏好编码器, 将学习者从知识图与行为图中得到的特征信息 I_{F} 进行编码, 获得学习者偏好中间编码 I_{O} , 如式(11)所示。

$$I_{\text{O}} = \sigma(\text{RNN}(I_{\text{F}}, h_e^{(t-1)})) \quad (11)$$

其中, $h_e^{(t-1)}$ 为编码器的隐藏层输出。然后, 同样采

用 RNN 作为偏好解码器, 将 I_{O} 映射到知识空间并解码成学习者偏好分布 I_{P} , 如式(12)所示。

$$I_{\text{P}} = \text{soft max}(\text{RNN}(I_{\text{O}}, h_d^{(t-1)})) \quad (12)$$

其中, I_{P} 为学习者对各知识点的偏好程度, $h_d^{(t-1)}$ 为解码器的隐藏层输出。需要说明的是, 编码器与解码器采用相同的结构但并不共享参数, 这是因为二者的功能相似但并不相同。编码器将相关信息进行编码以表示学习者的偏好信息; 解码器则是将被编码后的中间信息进行解码, 以获取在知识空间中的学习者偏好分布情况。

3.3.5 损失函数

本文采用交叉熵损失函数^[23]优化模型。设 D^+ 表示学习者与资源项之间真实存在的交互行为序列, 为从真实数据集中得到的训练集; D^- 表示学习者与资源项之间并不真实存在的交互行为序列, 为从数据集中通过随机采样生成该学习者未被观察到的交互行为序列的负训练集。

$$J = \left(- \sum_{U \in D^+} \log I_{\text{P}} - \left(1 - \sum_{U \in D^-} \log I_{\text{P}} \right) \right) \quad (13)$$

4 实验分析

本节首先描述实验中使用的数据集和评估指标, 然后在不同的实验设置下对 KGBR 进行了详细的分析。

4.1 实验设置

4.1.1 数据集描述

本文在 2 个真实的数据集上评估所提出的方法, 即 Algebra1/20 和 Assitment1/10。Algebra1/20 是 KDD Cup2010 所提供的公共数据集 algebra_2006_2007。Algebra1/20 采用了学生与计算机辅助教学系统之间的交互记录的形式, 记录了共 53 427 条来自不同学习者的学习记录。Assitment1/10 来自一个公共数据集 Assitments 2009-2010, 为在线导师记录的学生数学练习日志。Assitment1/10 数据集中记录了 46 298 条不同学生练习试题的学习记录。对于 2 个数据集, 本文设置 80% 的数据为训练集, 20% 的数据为测试集, 并进行如下数据预处理, 将所有的学习记录按照学生编号进行汇总, 以便得到学习者的行为序列, 并按照各试题编号对其包含的知识点进行了汇总。表 1 展示了数据集的基本信息。

表 1 数据集的基本信息

数据集	知识点数/个	试题数/个	记录数/个	学生数/个
Algebra1/20	240	1 843	53 427	575
Assitment1/10	113	13 409	46 298	287

4.1.2 基线方法

本节实验将本文模型 KGBR 与以下代表性基线模型进行了比较。

MF (recommendation from implicit feedback like matrix factorization) [24]。该模型利用学习者与资源项的交互作用,仅学习学习者和资源项的嵌入,没有利用 GNN。

NFM (neural factorization machine) [25]。该模型利用多个隐藏层来捕获学习者、资源项和上下文特征之间的非线性和高阶交互。

NGCF (neural graph collaborative filtering) [22]。模型将学习者资源项交互图中的高阶连接性嵌入学习者与资源项特征中。

RippleNet [26]。该模型利用知识图谱向外扩散学习者偏好,将外部信息与学习者和资源项的交互信息进行了融合。

SR-GNN (session-based recommendation with GNN) [8]。该模型将学习者行为构成具有多向转移关系的会话子图,利用子图中丰富的交互信息获得所有资源项的嵌入表示。

4.1.3 参数设置

本文的实验环境如下:操作系统为 Windows10, CPU 采用 Inter@CoreTMi7-9700K 8 GB, 频率为 3.70 GHz; GPU 采用 NVIDIA GeForce GTX1060 6G; IDE 环境为 JetBrains PyCharm 2019.1.1 x64; 编程语言为 Python3.6.2; 实验在 Pytorch1.4.0 中, 利用 dgl0.4.2 框架完成模型训练与测试。不同模型参数如下。

MF。对于 Algebra1/20 数据集,模型嵌入层大小为 240×240 ; 对于 Assitment1/10 数据集,模型嵌入层大小为 113×113 。使用 Adam 优化器,学习率设置为 0.001。

NFM。对于 Algebra1/20 数据集,模型嵌入层

大小为 240×480 , NFM 层大小为 480×240 ; 对于 Assitment1/10 数据集,模型嵌入层大小为 113×226 , NFM 层大小为 226×113 , 批量处理大小固定为 20, 使用 Adam 优化器,学习率设置为 0.001。

NGCF。对于 Algebra1/20 数据集,模型所有嵌入层大小为 240×240 ; 对于 Assitment1/10 数据集,模型的所有嵌入层大小为 113×113 。批量处理大小固定为 20, 使用 Adam 优化器,学习率设置为 0.001。

RippleNet。对于 Algebra1/20 数据集,模型知识图嵌入维数 $d=240$, 知识图阶数 $H=3$; 对于 Assitment1/10 数据集,模型知识图嵌入维数 $d=113$, 知识图阶数 $H=2$ 。

SR-GNN。按照原文方法分别对 2 个数据集设置了潜在向量维数 $d=100$, 此外,使用平均值为 0 且标准偏差为 0.1 的高斯分布初始化所有参数。批量处理大小为 100, 使用 Adam 优化器,学习率设置为 0.001。

KGBR。采用 Adam 优化器来优化模型,批处理大小固定为 30, 学习率设置为 0.001。KGBR 的网络参数如表 2 所示。

4.2 实验结果

4.2.1 基线方法性能比较

由于学习者偏好是存在于隐空间的特征向量,数据集中不存在能够指导模型训练的真实标签。因此,本文采用间接评价的方法测试模型性能。具体来说,将学习者偏好视为学习者特征向量,与各资源项的特征向量做计算得到学习者对各资源项的评分,根据推荐结果评估模型性能。本文采用 AUC 作为评估指标。AUC 被定义为 ROC 与横坐标轴围成的面积,取值范围为 $[0.5, 1]$ 。AUC 越接近 1, 表示模型真实性越高, AUC=1 时,真实性最高; AUC=0.5 时,真实性最低。

KGBR 与基线方法比较结果如表 3 所示。通过表 3 可以看出, MF、NMF 等传统算法的 AUC 相对较低。MF 模型虽然考虑了学习者与试题间的关系,但其本质上是一种线性模型,模型的代表能力有限。NMF 利用了学习者与试题间的非线性特征交

表 2

KGBRD 的网络参数

数据集	KGCN	RGAT	DEF	编码器	解码器
Algebra 1/20	1×240	1×480	480×1	240×240	240×240
Assitment 1/10	1×113	1×226	226×1	113×113	113×113

互信息，在 MF 的基础上有所加强。与 MF、NMF 相比，NGCF 利用图卷积拟合学习者与资源项之间的交互信息，通过堆叠多个嵌入传播层捕获知识的高阶连通信息，拟合了知识的结构性。但以上 3 种方法都没有考虑知识具有结构性的特点，因而整体表现较差。

表 3 KGBR 与基线方法比较结果

方法	AUC	
	Algebra1/20	Assitment1/10
MF	0.62	0.63
NFM	0.64	0.65
NGCF	0.69	0.70
RippleNet	0.80	0.82
SR-GNN	0.81	0.82
KGBR	0.83	0.85
KGBR _(withoutRGAT)	0.76	0.77
KGBR _(withoutKGCN)	0.80	0.82

RippleNet 和 SR-GNN 相比于前 3 种方法，AUC 值提升比较明显。这是因为 RippleNet 将试题与知识图上的实体对应，考虑了知识点间的结构性特点，符合学习者的整体性学习规律。缺点是对学习者个体性学习规律的考虑不够充分且该模型没利用知识图的边信息。SR-GNN 利用学习者历史行为构建学习者行为表示，该行为表示可以自动选择最重要的试题转换并忽略历史信息中影响较弱的甚至是无效的学习者行为。但是知识点之间具有很强的结构性，而该方法仅强调了学习者个性化的学习规律，忽略了学习者的整体性学习规律。与 RippleNet 和 SR-GNN 相比，KGBR 结合二者的优点，采用知识图与行为图混合嵌入的方式对学习者建模并采用差异性嵌入融合方式获取学习者偏好，因而 KGBR 在 2 个数据集上的所有基线方法中表现是最优的，在 Algebra1/20 数据集上的 AUC 为 0.83，在 Assitment1/10 数据集上的 AUC 为 0.85。

除此之外，为了了解 KGBR 中各部分对于模型的贡献，本文进行了针对不同部分的消融实验，其中 KGBR_(withoutRGAT) 表示去除行为图嵌入层的模型，KGBR_(withoutKGCN) 表示去除知识图嵌入层的模型。KGBR_(withoutRGAT) 在 2 个数据集上的 AUC 分别为 0.76 与 0.77，该模型没有考虑学习者自身的学习行为，忽略了学习者的个体性偏好导致模型性能下降比较明显。KGBR_(withoutKGCN) 在 2 个数据集的 AUC

分别为 0.80 与 0.82，这是因为该模型没有考虑学习者的整体性学习规律，忽略了知识结构行特点，所以 AUC 相对于 KGBR 有所下降，但不同学习者偏好估计显然应该更加偏重学习者的个体性偏好，因此 AUC 的变化较小。

综上所述，该实验结果验证了知识图嵌入层对拟合学习者整体性学习规律的重要性以及个体性学习规律与整体性学习规律间的差异性对学习者优先偏好的影响。

4.2.2 不同融合方式比较

本节实验比较 KGBR 模型不同融合方式的模型性能，结果如表 4 所示。KGBR-avg 在 Algebra1/20 和 Assitment1/10 数据集上的 AUC 分别为 0.80 和 0.81，性能较弱。这是因为平均池化方式将个体性学习规律和整体性学习规律对学习者优先偏好的影响视为同等重要，不能根据学习者特点具有针对性地放大或缩小个体性学习规律的影响。KGBR-inner 在 2 个数据集上的 AUC 都为 0.81。其采用向量内积的方式计算 2 种特征向量的相似度，利用该相似度获取个体性学习规律与整体性学习规律的差异性，但缺点是对个体性特征向量的控制不够精细，因而在 2 个数据集上的 AUC 提升不明显。KGBR-concat 在 2 个数据集上的 AUC 分别为 0.82 和 0.83。其采用串联方式虽然最大限度地保留了特征向量的内部信息，但是也因此无法区分 2 种学习规律中各自独有的特征信息，对二者的差异性不敏感。本文采用 KGBR-DEF 方式，用哈达玛积实现了对个体性特征向量每一维的控制，在进行特征向量融合时可以准确地捕获 2 种学习规律的差异。其在 Algebra1/20 数据集上的 AUC 为 0.83，在 Assitment1/10 数据集上的 AUC 为 0.85，均获得了最佳结果。这证实了针对不同的学习者特点进行个性化的融合对获取准确的学习者偏好分布至关重要。

表 4 不同融合方式比较结果

方法	AUC	
	Algebra1/20	Assitment1/10
KGBR-avg	0.80	0.81
KGBR-inner	0.81	0.81
KGBR-concat	0.82	0.83
KGBR-DEF	0.83	0.85

4.2.3 隐藏层结构影响比较

由于 KGBR-DEF 中使用神经网络获得知识权

重控制向量，因此本节讨论了具有不同数量隐藏层的 MLP 对模型性能的影响，结果如表 5 所示。其中，MLP-0 表示无隐藏层，MLP-1 表示具有一个隐藏层的 MLP，其余符号类似。从表 5 可以看出，对于 Algebra1/20 和 Assitment1/10 数据集，当不使用神经网络时，模型的性能非常差，这是因为仅利用学习者的个体性学习规律和整体性学习规律的差值不足以建模二者之间的差异性，因此有必要使用隐藏层对其进行特征转换。MLP-1 和 MLP-2 的 AUC 相对于 MLP-0 提升比较显著，在 2 个数据集上的 AUC 分别为 0.83 和 0.85，但上升趋于平缓。这说明隐藏层能够对 2 种学习规律的差异性进行比较准确的建模。MLP-3 的 AUC 发生了下降，说明对于 KGBR 来说，2 层隐藏层已经足够建模 2 种学习规律的差异性，而过多的隐藏层带来的高度非线性特征使模型出现过拟合现象。因此，对于差异性嵌入融合层来说，当隐藏层的数量等于 2 时，模型性能最佳。

表 5 隐藏层深度的影响

隐藏层	AUC	
	Algebra1/20	Assitment1/10
MLP-0	0.63	0.66
MLP-1	0.82	0.83
MLP-2	0.83	0.85
MLP-3	0.81	0.83

4.2.4 知识图迭代次数比较

由于知识图嵌入部分对于拟合学习者整体性的学习规律十分重要，因此本节对知识图的迭代次数进行了比较，结果如图 3 所示。

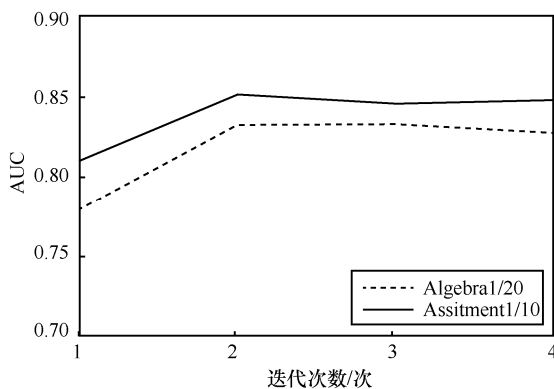


图 3 知识图迭代次数比较

从图 3 可以看出，AUC 随着 L 的增大而升高，并在 $L=2$ 到顶峰，随后趋于平缓。当 $L=1$ 时 AUC 较

低，知识图嵌入并不能很好地拟合学习者的整体性学习规律。当 $L=2、3、4$ 时，AUC 值变化不大且趋于平缓，这可能是由于 2 个数据集中学习者涉及的知识点比较集中且数量不大，不需要过多地学习相关信息。综上所述，本文将知识图迭代次数设置为 $L=2$ 。

4.2.5 模型收敛性分析

本节针对模型的收敛性进行了实验，分别在不同数据集上验证迭代次数与 AUC 的关系。实验结果如图 4 和图 5 所示。KGBR-train 表示模型训练时的 AUC，可以看出，在 2 个数据集上模型迭代 10 次之前 AUC 提升较明显，迭代 20 次时趋于平缓，之后 AUC 随着迭代次数的增加而缓慢增加，这表明模型收敛速度良好且随着迭代次数的增加模型性能也在逐渐提高。KGBR-test 表示模型测试时的 AUC，可以看出，经过训练后模型在 2 个数据上的性能比较稳定且较优秀。

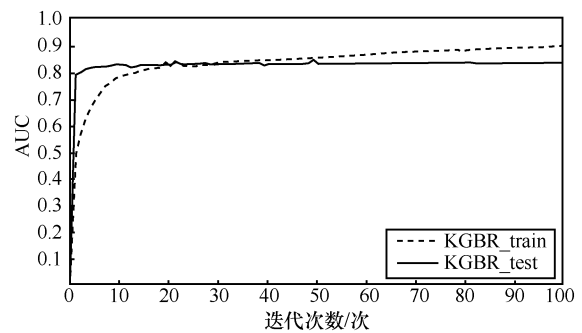


图 4 Algebra1/20 的 AUC 与迭代次数的关系

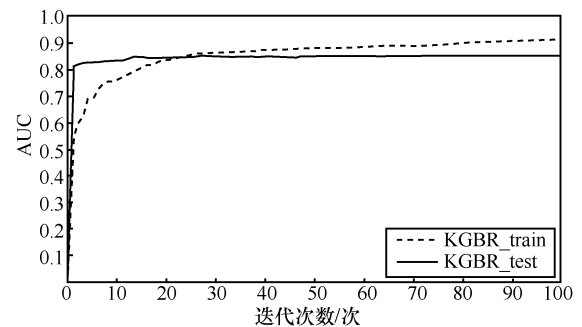


图 5 Assitment1/10 的 AUC 与迭代次数的关系

5 结束语

针对知识推荐中知识结构信息与行为结构信息的利用问题，本文提出一种融合知识图和行为图的知识推荐模型 KGBR。利用图卷积技术对知识图和行为图进行了融合嵌入。根据知识的关系类型，KGBR 在结构嵌入时考虑图边的信息，同时考虑了学习者整体

性偏好与个体性偏好间的差异性, 引入独特的融合机制, 获得更准确的学习者偏好预测模型。本文在 2 个真实数据集上进行了实验分析, 实验结果表明, 与传统的知识推荐方法相比, KGBR 的 AUC 平均提高了 2.49%, 这证明了 KGBR 的有效性。

参考文献:

- [1] TANISKIDOU E K, PAPADAKIS G, GIANNAKOPOULOS G, et al. Comparative analysis of content-based personalized microblog recommendations experiments and analysis[J]. arXiv Preprint, arXiv: 1901.05497, 2019.
- [2] ZHOU T Q, CHEN L N, SHEN J. Collaborative filtering-based recommendation system for big data[J]. International Journal of Computational Science and Engineering, 2020, 21(2): 219.
- [3] CHEN J, WANG X S, ZHAO S, et al. Deep attention user-based collaborative filtering for recommendation[J]. Neurocomputing, 2020, 383: 57-68.
- [4] HE X N, LIAO L Z, ZHANG H W, et al. Neural collaborative filtering[C]//Proceedings of the 26th International Conference on World Wide Web. New York: ACM Press, 2017: 173-182.
- [5] BOBADILLA J, ORTEGA F, GUTIÉRREZ A, et al. Classification-based deep neural network architecture for collaborative filtering recommender systems[J]. International Journal of Interactive Multimedia and Artificial Intelligence, 2020, 6(1): 68.
- [6] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. arXiv Preprint, arXiv: 1710.10903, 2017.
- [7] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. arXiv Preprint, arXiv: 1609.02907, 2016.
- [8] WU S, TANG Y Y, ZHU Y Q, et al. Session-based recommendation with graph neural networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 346-353.
- [9] BERG R V D, KIPF T N, WELING M. Graph convolutional matrix completion[J]. arXiv Preprint, arXiv: 1706.02263, 2017.
- [10] ZHANG M H, CHEN Y X. Inductive matrix completion based on graph neural networks[J]. arXiv Preprint, arXiv: 1904.12058, 2019.
- [11] MNH A, SALAKHUTDINOV R. Probabilistic matrix factorization[C]// Advances in neural information processing systems. Massachusetts: MIT Press, 2007: 1257-1264.
- [12] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems[J]. Computer, 2009, 42(8): 30-37.
- [13] KOREN Y, BELL R. Advances in collaborative filtering[M]. Berlin: Springer, 2011.
- [14] PANG Y, JIN Y, ZHANG Y, et al. Collaborative filtering recommendation for MOOC application[J]. Computer Applications in Engineering Education, 2017, 25(1): 120-128.
- [15] PANG Y X, LIU W Y, JIN Y Y, et al. Adaptive recommendation for MOOC with collaborative filtering and time series[J]. Computer Applications in Engineering Education, 2018, 26(6): 2071-2083.
- [16] PUJAHARI A, SISODIA D S. Pair-wise preference relation based probabilistic matrix factorization for collaborative filtering in recommender system[J]. Knowledge-Based Systems, 2020, 196: 105798.
- [17] LI K K, ZHOU X Z, LIN F, et al. Deep probabilistic matrix factorization framework for online collaborative filtering[J]. IEEE Access, 2019, 7: 56117-56128.
- [18] XIAO Y Y, WANG G W, HSU C H, et al. A time-sensitive personalized recommendation method based on probabilistic matrix factorization technique[J]. Soft Computing, 2018, 22(20): 6785-6796.
- [19] PARDOS Z A, HEFFERNAN N T. Modeling individualization in a Bayesian networks implementation of knowledge tracing[C]//User Modeling, Adaptation, and Personalization. Berlin: Springer, 2010: 255-266.
- [20] ZHANG K, YAO Y Y. A three learning states Bayesian knowledge tracing model[J]. Knowledge-Based Systems, 2018, 148: 189-201.
- [21] MINN S, YU Y, DESMARAI S M C, et al. Deep knowledge tracing and dynamic student classification for knowledge tracing[C]//2018 IEEE International Conference on Data Mining. Piscataway: IEEE Press, 2018: 1182-1187.
- [22] WANG X, HE X N, WANG M, et al. Neural graph collaborative filtering[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2019: 165-174.
- [23] LIAN J X, ZHOU X H, ZHANG F Z, et al. Combining explicit and implicit feature interactions for recommender systems[J]. arXiv Preprint, arXiv: 1803.05170, 2018.
- [24] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback[J]. arXiv Preprint, arXiv: 1205.2618, 2012.
- [25] HE X N, CHUA T S. Neural factorization machines for sparse predictive analytics[C]//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2017: 355-364.
- [26] WANG H W, ZHANG F Z, WANG J L, et al. RippleNet: propagating user preferences on the knowledge graph for recommender systems[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2018: 417-426.

[作者简介]



李晓光 (1973-), 男, 辽宁沈阳人, 博士, 辽宁大学教授、硕士生导师, 主要研究方向为机器学习等。

宫磊 (1996-), 男, 辽宁大连人, 辽宁大学硕士生, 主要研究方向为推荐系统等。

李晓莉 (1973-), 女, 辽宁沈阳人, 辽宁大学副教授, 主要研究方向为大数据等。

张昕 (1979-), 男, 辽宁沈阳人, 博士, 辽宁大学副教授、硕士生导师, 主要研究方向为复杂系统等。

于戈 (1962-), 男, 辽宁沈阳人, 博士, 东北大学教授, 主要研究方向为数据库等。