

基于知识图谱与协同过滤混合策略的在线编程评测系统题目推荐模型

刘泽京, 邬楠, 黄抚群, 宋友

引用本文

刘泽京, 邬楠, 黄抚群, 宋友. [基于知识图谱与协同过滤混合策略的在线编程评测系统题目推荐模型](#) [J]. 计算机科学, 2023, 50(2): 106-114.

LIU Zejing, WU Nan, HUANG Fuqun, SONG You. [Hybrid Programming Task Recommendation Model Based on Knowledge Graph and Collaborative Filtering for Online Judge](#) [J]. Computer Science, 2023, 50(2): 106-114.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于特征融合的小样本目标检测](#)

Few-shot Object Detection Based on Feature Fusion

计算机科学, 2023, 50(2): 209-213. <https://doi.org/10.11896/jsjcx.220500153>

[一种基于多模态深度特征融合的视觉问答模型](#)

Visual Question Answering Model Based on Multi-modal Deep Feature Fusion

计算机科学, 2023, 50(2): 123-129. <https://doi.org/10.11896/jsjcx.211200303>

[基于图注意力的神经协同过滤社会推荐算法](#)

Neural Collaborative Filtering for Social Recommendation Algorithm Based on Graph Attention

计算机科学, 2023, 50(2): 115-122. <https://doi.org/10.11896/jsjcx.211200019>

[基于先验知识图谱的多代理被遮挡目标类别推理模型](#)

Novel Class Reasoning Model Towards Covered Area in Given Image Based on Informed Knowledge Graph Reasoning and Multi-agent Collaboration

计算机科学, 2023, 50(1): 243-252. <https://doi.org/10.11896/jsjcx.220700112>

[残差注意力与多特征融合的图像去模糊](#)

Image Deblurring Based on Residual Attention and Multi-feature Fusion

计算机科学, 2023, 50(1): 147-155. <https://doi.org/10.11896/jsjcx.211100161>

基于知识图谱与协同过滤混合策略的在线编程评测系统题目推荐模型

刘泽京¹ 邬楠¹ 黄抚群² 宋友¹

¹ 北京航空航天大学软件学院 北京 100191

² 葡萄牙科英布拉大学信息学和系统中心 科英布拉 3000-115

(sy1921105@buaa.edu.cn)

摘要 在线编程评测系统(Online Judge,OJ)是一种被广泛应用于计算机编程教学与竞赛的代码测评系统。用户在规模庞大的题库中寻找适合当前学习阶段的题目时,往往会感到迷茫。如何为用户推荐合适的题目和规划学习路径,是在线编程评测系统研发中的一个重要研究课题。传统推荐算法存在可解释性和准确性难以兼顾的问题。文中提出了基于知识图谱与协同过滤混合策略的在线评测系统推荐模型(A Hybrid Programming Task Recommendation Model Based on Knowledge Graph and Collaborative Filtering,HKGCF)。该模型通过推荐与用户当前知识和技能掌握程度相匹配的题目,来帮助用户提升学习效果。文中设计和实现了该模型,并将其集成到了北京航空航天大学在线编程评测系统中,以适应OJ平台特有的交互形式。线上测试和离线测试实验的结果表明,提出的HKGCF模型在准确率和可解释性方面均优于典型传统算法。

关键词:编程教育;在线编程评测系统;个性化推荐;知识图谱;协同过滤;特征融合

中图法分类号 TP391

Hybrid Programming Task Recommendation Model Based on Knowledge Graph and Collaborative Filtering for Online Judge

LIU Zejing¹, WU Nan¹, HUANG Fuqun² and SONG You¹

¹ School of Software, Beihang University, Beijing 100191, China

² Centre for Informatics and Systems, University of Coimbra, Coimbra 3000-115, Portugal

Abstract The online judge (OJ) is a widely used system for programming education, learning and contests. Users often get lost in searching for tasks of interest in the massive database. How to recommend suitable programming tasks to the users and plan the learning path is a significant research topic in the development of online programming evaluation system. Existing traditional recommendation methods have the limitation of making a trade-off between interpretability and effectiveness. This paper proposes a task-recommending model for the OJ platform - hybrid programming task recommendation model based on knowledge graph and collaborative filtering for online judge (HKGCF). The HKGCF model can help users improve their learning effect by recommending questions that match their current knowledge levels and skills. The model is designed based on a hybrid strategy that integrates the knowledge graph representation learning with an improved collaborative filtering algorithm. The model is implemented and integrated into the OJ platform of Beihang University, and meet the specific interaction formats of the OJ platform. We conducted two experiments, an online and an offline test, to validate the proposed model and its implementations. The results show that the proposed model outperforms the representative conventional recommendation algorithm in terms of interpretability and accuracy.

Keywords Programming education, Online judge systems, Personalized recommendations, Knowledge graph, Collaborative filtering, Feature fusion

1 引言

近年来,在线学习平台的兴起加强了大数据与知识之间的联系,改变了人们的学习方式和部分生活方式,同时也颠覆

了教育系统的常规模式。随着计算机硬件和网络的快速发展,用户在在线学习平台上的活动愈加频繁,收集到的用户行为数据日益增多,为智能化、个性化和定制化服务的诞生和发展提供了契机,同时也催化了学术界和工业界对个性化推荐、

到稿日期:2021-12-09 返修日期:2022-07-13

基金项目:国家自然科学基金面上项目(61977002);国家自然科学基金青年科学基金项目(62107002);河北省重点研发计划(21310101D)

This work was supported by the National Natural Science Foundation of China General Program(61977002), National Natural Science Foundation of China for Youths(62107002) and key Research and Development Program of Hebei Province(21310101D).

通信作者:宋友(songyou@buaa.edu.cn)

点击率预估等技术的广泛关注。通过分析用户的学习行为、建模用户的行为特点,能够挖掘出用户的潜在兴趣偏好,更好地为用户提供贴近其实际需求的产品或服务。

编程教育领域中,以 Codeforces^[1]及 Peking University Online Judge (POJ)^[2]为代表的在线评测系统为用户提供了程序编写的平台与技术支持,成为了用户在线编程学习过程中不可或缺的一环。在线评测系统起源于 ACM 国际大学生程序设计竞赛(International Collegiate Programming Contest, ICPC),打破了对编程学习的时间和空间的限制,满足了学习者多样化的学习需求,非常适合应用于 C, C++, Java 等各种编程语言的教学和培训。

随着在线编程教育的普及,平台中的题目数量也迅速增加,用户寻找到符合自身兴趣且适合自己当前学习阶段的题目也愈发困难。为了满足用户对题目的多样化需求,本文设计了一种混合策略的推荐模型,即基于知识图谱和协同过滤的题目推荐模型——HKGCF。该模型利用在线评测系统中积累的大量零散、无意识的学习行为数据,通过图嵌入技术将学习行为数据分解为可获取、易操作的数据指标,利用知识图谱表征学习发现用户的兴趣点,融合改进的协同过滤算法增强结果的准确率和可解释性,进而向用户推荐合适的题目以帮助用户提高编程实践能力。

2 相关工作

2.1 协同过滤推荐算法

在编程教育领域内,已有很多研究将传统推荐算法应用在编程在线评测平台上,其中最为经典和应用最为广泛的是依据群体相似特性进行推荐的协同过滤算法^[3-4],该算法主要包括 3 种类型:基于用户、基于物品和基于模型的协同过滤推荐算法^[5]。Wu 等提出了一种改进的基于用户的协同过滤算法,该算法结合了用户相似性和信任度,以获得更合适的最近邻居集,其结果准确率有一定的提升^[6]。Yu 等参考设计了基于“用户-兴趣-物品”的三层模式协同过滤推荐方法,并针对编程在线评测平台的特定交互模式进行了重新设计,获得了更好的推荐效果^[7]。协同过滤算法具有推荐准确率高、可解释性强、部署性好的特点,具备应用于推荐编程题目的潜力。然而,协同过滤算法本身泛化能力较弱,且与题目内容相关性不强,新用户和新题目加入后需要解决冷启动问题,无法加入用户、物品属性、上下文特征等信息。虽然有研究尝试通过融合用户及项目特征来提高传统协同算法的表现^[8-9],但无法有效捕捉在线评测平台中用户的潜在做题兴趣,因此在实际应用中遇到了一些瓶颈。

2.2 基于知识图谱表征学习的推荐算法

近年来,深度学习技术(Deep Learning, DL)快速发展,在用户行为建模和预测领域得到了广泛应用。一些比较成熟的深度学习模型如循环神经网络(Recurrent Neural Network, RNN)、图神经网络(Graph Neural Networks, GNN)等也被引入到学习题目推荐和学习行为建模中。与决策树(Decision Tree, DT)、逻辑回归、朴素贝叶斯模型(Naive Bayesian Model, NBM)等传统模型对比,深度学习模型呈现出了较优的性能^[10-11]。同时,注意力模型(Attention Model,

AM)被引入推荐领域,解决了推荐结果可解释性不高、无法捕捉用户的长期兴趣与短期偏好等瓶颈。AFM^[12], DIN 等^[13]基于注意力机制的推荐模型在学习平台推荐、学生画像生成等方向大放异彩,例如 Chounta 等利用注意力机制,自动学习学生表现和学生行为之间的重要性权重,更精准地模拟学生的行为数据^[14]。然而,随着越来越多的学者将深度学习应用在推荐系统中,深度学习网络的过拟合、无法对结果进行解释等缺点也逐渐暴露出来。

作为传统数据表征形式的补充,知识图谱可以通过对知识图谱中的实体和关系进行表征学习来计算实体间的语义相关性,从而得到解释性更强的结果^[15]。以 TransE^[16], TransH^[17]及 STTransH^[18]为代表的知识图谱翻译表征模型,通过将异构信息投影到统一的低维嵌入空间,来将实体和关系转化为向量表征自动学习语义嵌入。该做法不仅提高了算法效率,并且解决了传统推荐算法的数据稀疏性问题,因此被成功应用到推荐系统中,如 MKR^[19], entity2rec^[20]等。

2.3 结合协同过滤与知识图谱的混合推荐算法

混合推荐模型可以将多个不同类型的推荐算法整合起来,融合并放大各种方法的优势,从而得到更好的推荐效果。目前已有国内外学者尝试集成知识图谱表征学习与传统协同过滤算法,以提高传统算法的准确率,同时增强可解释性和解决数据稀疏性的问题。例如,Zhang 等提出的 CKE (Collaborative Knowledge Base Embedding for Recommender Systems) 算法使用结构化知识表征方法将物品的嵌入向量融合到协同过滤算法中,有效提升了推荐准确率^[21]。然而,该方法直接叠加两种算法向量的数据融合方式,损失了部分上下文信息。近年出现的融合知识图谱与协同过滤的混合推荐模型(HCKDC model),基于深度学习网络的知识图谱模型并结合协同过滤算法,按预测评分的准确度将两个模型的结果相互融合,提升了准确率^[22],然而其结果的可解释性有待提高。

综上所述,现有方法在推荐结果准确度和可解释性的权衡中难以达到平衡,本文将知识图谱表征学习与协同过滤算法借鉴多路召回策略的思路有效结合,在保证准确度的前提下,最大限度地提升了可解释性。

3 HKGCF 算法的设计

本文设计的 HKGCF 算法包含协同过滤单元、知识图谱表征学习单元和精准排序单元 3 个部分,模型架构如图 1 所示。

协同过滤单元负责进行用户交互信息向量建模。本单元吸收协同过滤算法的推荐原理,将用户与题目的交互数据转换为用户-题目交互矩阵 \mathbf{W} ,并通过梯度下降(Gradient Descent)的方法进行向量分解,从而获得稠密且包含交互信息的用户隐向量 \mathbf{u}_{cf} 及题目隐向量 \mathbf{t}_{cf} 。

知识图谱表征学习单元通过构建两个蕴含不同维度关系的知识图谱,对用户与题目的近似度进行计算。由于在线评测系统中题目和题目属性具有明显的从属关系,因此可以将数据集中的题目 $\{t_1, t_2, \dots, t_n\}$ 及题目属性分别看作实体,通过关联信息联结构成一个题目知识图谱 \mathcal{G}_t ,继而通过改进的知识图谱表征学习方法将知识图谱中的点与边表示成数值化

的低维稠密向量,从而计算得出题目的知识特征向量 t_{kg} 。与此同时,在计算用户的知识特征向量时,本文模型将题目知识图谱 \mathcal{G}_t 与用户-题目交互二部图结合得到用户-题目知识图谱 \mathcal{G}_{ut} ,在知识图谱中引入了用户交互信息,将用户与题目中其他属性实体间接联系起来,从而通过知识表征学习得到用户的知识特征向量 u_{kg} 。

精准排序单元将前两个单元得到的用户维度的两个向量 u_{cf} 、 u_{kg} 及题目维度的两个向量 t_{cf} 、 t_{kg} 分别进行聚合。聚合过程采用加权聚合的方式,其中4个向量两两计算元素积后输入到注意力单元中计算各自的权重 a 。最终计算得到用户对题目的兴趣分数 \hat{y}_{ut} ,并由高到低排序生成用户的 Top-K 推荐集。

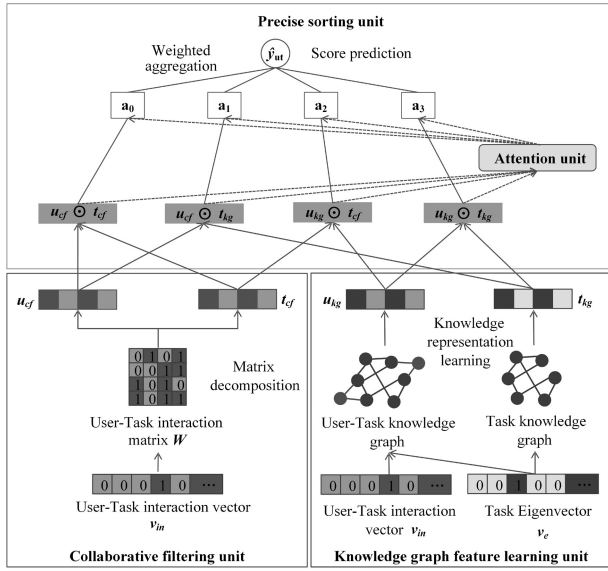


图1 HKGCF 模型架构图

Fig. 1 HKGCF model architecture

3.1 问题描述

推荐算法设计的第一步,需要对所推荐问题进行形式化描述。对于给定的 m 个用户组成的集合 $U = \{u_1, u_2, \dots, u_m\}$ 以及 n 道题目组成的集合 $T = \{t_1, t_2, \dots, t_n\}$,本文模型构建用户-题目交互矩阵 W 、题目知识图谱 \mathcal{G}_t 和用户-题目知识图谱 \mathcal{G}_{ut} ,HKGCF 模型通过加权聚合后的特征向量融合结果 \hat{y}_{ut} 来量化用户 u 对题目 t 的偏好值,从而完成用户对题目潜在兴趣的预测过程。

本文使用的符号集合如表 1 所列。

表1 文中使用的符号集合

Table 1 Set of symbols used in this paper

Notations	Descriptions
U, T	the set of users and tasks
m, n	the number of users and tasks
v	the initial vectors
u, t	the vector of users and tasks
h, r, l	the vector of entities and relations
W	the interaction matrix
N	the weighting matrix
M	the conversion matrix
\mathcal{G}	the knowledge graph
a	the weighting factors

协同过滤单元主要关注用户与题目的交互过程,通过数据集中所有用户与题目的交互关系计算用户和题目间的相似度。对于用户 u ,其提交过的题目及提交结果可以组成用户交互向量 v_m ,多个用户交互向量便可形成用户-题目交互矩阵 $W \in \mathbb{R}^{m \times n}$ 。在线评测系统中大部分用户仅仅会与一部分题目产生交互,这样形成的矩阵往往较为稀疏。因此,需要使用矩阵分解的方法,在不损失向量特征的情况下,将稀疏向量转变为低维稠密向量。矩阵分解方法可以将用户和题目都映射到一个 K 维空间中,生成的用户和题目隐向量都蕴藏着用户与题目间深层次的联系。相比传统协同过滤算法,矩阵分解方法有效加强了模型处理稀疏矩阵的能力,针对性地解决了协同过滤算法泛化能力弱、头部效应明显的问题。同时,与传统的商品推荐或电影推荐等推荐场景不同的是,编程教学在线评测系统的用户不对题目进行评分。用户对于一个题目或一个知识点的掌握程度是通过其题目提交后的评测结果进行量化而来的。该评测结果等效于用户对该题目的需求分数。OJ 中常见的题目提交评测结果类型如表 2 所列。

表2 题目提交评测结果分类

Table 2 Classification of submit assessment results for common tasks

Returned results	Descriptions
AC(Accepted)	Pass all test points of the task
WA(Wrong Answer)	The program returned an incorrect result at one of the test points
CE(Compile Error)	The program did not pass the system's compilation
RE(Runtime Error)	An error occurred while the program was running that caused it to crash
PE(Presentation Error)	The program outputs the correct result, but the output is incorrectly formatted
TLE(Time Limit Exceeded)	The program runs beyond the runtime limit

知识图谱表征学习单元着眼于挖掘用户与题目间以及题目与题目属性间的关联关系,使用知识图谱对深层特征进行向量表示。由于知识图谱是一种特殊网络,其中每个节点代表现实世界中的实体,而节点间的边表示实体之间的关系。因此,一般用三元组形式表示内涵知识,构建三元组集合知识库 $S = \{(h, r, l)\}$,其中 h 代表头实体, l 代表尾实体, r 代表从头实体到尾实体的关系。在线评测系统主要考虑 5 种实体,即用户、题目、知识点、难度和比赛类型,详细描述如表 3 所列。这 5 类实体之间的关系如表 4 所列。

通过构建题目和题目间属性的关系,将多个三元组相互连接,最终形成如图 2 所示的题目知识图谱。此外,加入用户-题目交互二部图信息,进而形成如图 3 所示的用户-题目知识图谱。与协同过滤单元中不同的是,在构建用户与题目的交互关系时,仅将某用户正确提交即返回结果为 AC 的题目纳入该用户的尾实体列表中。该设计的原因是,提交结果反馈为“正确”的题目表示用户已经掌握这个题目及其所包含的知识点,该题目才可作为用户的知识特征建模的一部分,来提高向量的表示精度,从而提高推荐结果的准确性。

表3 HKGCF模型中知识图谱实体列表

Table 3 Knowledge graph entities in HKGCF model

Entity Names	Descriptions
Task	Tasks in OJ system
Knowledge point	Knowledge points of the task
Difficulty	Difficulty of the task
Contest	Contests in OJ system
User	Users in OJ system

表4 HKGCF模型中知识图谱实体关系列表

Table 4 Knowledge graph entity relationships in HKGCF model

Entity A	Entity Relationships	Entity B
Task	Contains	Knowledge point
Task	Has property	Difficulty
Task	Belongs to	Contest
User	Submits correctly	Task
User	Participates in	Contest

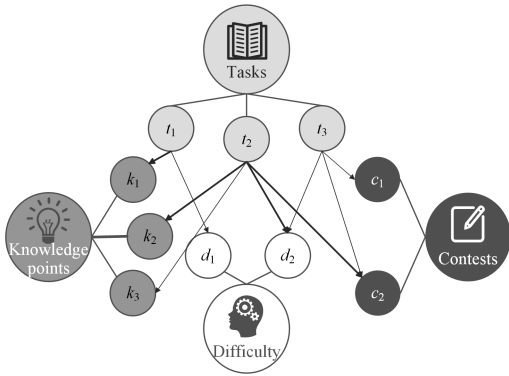


图2 题目知识图谱

Fig. 2 Tasks knowledge graph

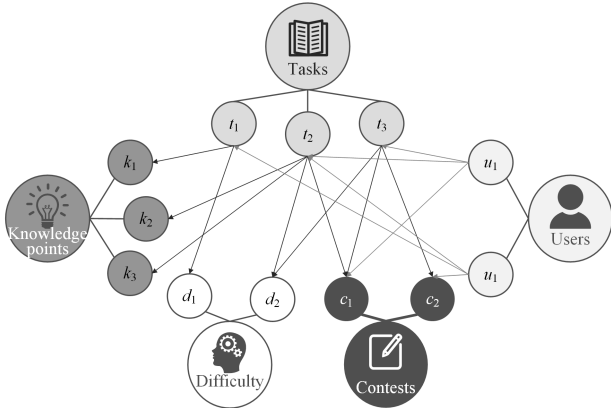


图3 用户-题目知识图谱

Fig. 3 Users-Tasks knowledge graph

精准排序单元的目标是在特征信息损失尽量少的前提下,将前两个单元得出的特征向量进行融合,然后计算用户对题目的偏好值 \hat{y}_{ut} ,并对题目集中的题目进行偏好值排序,最终得出推荐结果。

3.2 基于知识图谱与协同过滤的混合推荐模型

3.2.1 基于矩阵分解的协同过滤算法

HKGCF通过协同过滤中的矩阵分解算法计算得出用户和题目的隐向量。对于题目推荐系统来说,如果用户对一个题目的掌握度不高,那么该用户对相似题目的需求就越大,系统则应优先推荐与此题相关的题目。因此,如果将用户的

评测结果作为题目的需求权重 ω_{ut} ,那么评测结果越高,用户对该题目的权重越低。由表3可知,提交结果AC和PE表示用户已经基本掌握题目所考察的知识, ω_{ut} 应设为较小数;反之对于其他提交结果, ω_{ut} 应设为较大数。模型通过式(1)计算用户对每道题目的需求权重,其中, N_u 代表用户 u 交互过的题目合集, s_u 代表用户提交题目的次数, δ_t 代表每次提交评测结果的量化系数,该量化系数的具体取值如式(2)所示:

$$\omega_{ut} = \left(\frac{\sum_t \delta_t}{s_u} \right)^{-\frac{1}{2}} \quad (1)$$

$$\delta_t = \begin{cases} 1, & \text{提交结果为AC及PE} \\ e^{-1}, & \text{提交结果为其他结果} \end{cases} \quad (2)$$

用户的交互向量 \mathbf{v}_u 由计算得到的每道题目的 ω_{ut} 值组成,将数据集中 m 个用户的交互向量作为矩阵的行向量便可获得用户-题目权重矩阵 $\mathbf{W}_{m \times n}$ 。将该矩阵通过梯度下降方法后即可完成隐向量维度为 k 的矩阵分解过程。其目标函数需要使原始权重 ω_{ut} 与分解后的用户向量和物品向量之间的乘积之差尽量小,以保证最大限度地保留原始权重信息,如式(3)所示:

$$\min_{\mathbf{q}^*, \mathbf{p}^*} \sum_{(u,t) \in K} (\omega_{ut} - \mathbf{q}_t^T \mathbf{p}_u)^2 \quad (3)$$

其中, \mathbf{p}_u 是用户 u 在用户矩阵 \mathbf{U} 中的对应行向量, \mathbf{q}_t 是题目 t 在题目矩阵 \mathbf{T} 中的对应列向量, K 是所有用户需求权重的集合。由于用户对题目的需求权重体现的是客观的知识掌握程度,与用户个人的主观意愿无关,因此不需要消除用户对题目的打分偏差。同时需要加入正则化项来控制模型的方差,修改后的目标函数如式(4)所示,其中 λ 为超参数,用于控制正则化项。

$$\min_{\mathbf{q}^*, \mathbf{p}^*} \sum_{(u,t) \in K} (\omega_{ut} - \mathbf{q}_t^T \mathbf{p}_u)^2 + \lambda (\|\mathbf{q}_t\| + \|\mathbf{p}_u\|)^2 \quad (4)$$

考虑到真实的在线评测系统中用户除提交题目外也会有点击题目、长时间浏览题目等侧面反映了用户喜好的隐式反馈情况,因此需要在算法中加入此类额外的信息源,以帮助系统更好地预测用户兴趣,同时解决显式评分行为较少导致的冷启动问题。因此,在式(4)的基础上,考虑用户有过隐式交互行为的题目合集,并和权重矩阵中的题目一样增加隐因子参数,将所有的隐因子向量相加,从而获得最终的用户和题目隐向量。则预测权重 $\hat{\omega}_{ut}$ 的计算式更新为式(5),目标函数如式(6)所示:

$$\hat{\omega}_{ut} = \mathbf{q}_t^T (\mathbf{p}_u + |N(u)|^{-0.5} \sum_{j \in N(u)} \mathbf{y}_j) \quad (5)$$

$$\min_{\mathbf{q}^*, \mathbf{p}^*} \sum_{(u,t) \in K} (\omega_{ut} - \hat{\omega}_{ut})^2 + \lambda (\|\mathbf{q}_t\|^2 + \|\mathbf{p}_u\|^2 + \sum_{j \in N(u)} \|\mathbf{y}_j\|^2) \quad (6)$$

其中, $N(u)$ 为用户 u 产生隐式交互行为的题目合集, \mathbf{y}_j 为用户对于题目的隐藏喜好偏置。最后便可通过梯度下降算法进行模型参数的求解,从而得到用户隐向量 \mathbf{u}_{cf} 及题目隐向量 \mathbf{t}_{cf} 。

3.2.2 基于改进的TransR算法的知识图谱的建立与表征学习

HKGCF采用改进的TransR方法获得题目信息的特征向量表征。此类基于距离的翻译模型的基本思想是:如果

两个实体在同一三元组 $\langle h, r, l \rangle$ 中, 则头实体向量加上关系向量后, 应该与尾实体向量接近, 即 $\mathbf{h} + \mathbf{r} \approx \mathbf{l}$ 。根据这一思想, 可得翻译模型的损失函数, 如式(7)所示:

$$f_r(\mathbf{h}, \mathbf{l}) = \|\mathbf{h} + \mathbf{r} - \mathbf{l}\|_{l_1/l_2} \quad (7)$$

其中, \mathbf{h} 与 \mathbf{l} 分别是三元组中头实体 h 与尾实体 l 的向量化表示。因此, 可以通过梯度下降的方法来学习 $\sum_{\langle h, r, l \rangle \in S} f_r(\mathbf{h}, \mathbf{l})$ 最小化时的相关参数, 进而得到实体的向量表示。

与其他翻译模型相比, 本文算法采用的改进的 TransR 算法可以通过关系矩阵来链接不同语义空间, 如图 4 所示。在知识图谱结构中, 一个实体是多种属性的综合体; 不同关系关注实体的不同属性; 实体与关系存在于不同的空间。例如, 在题目知识图谱中, 假设“题目 A”与“题目 B”是两个相似的实体, 但是 \langle 题目 A, 包含, 背包问题 \rangle 与 \langle 题目 B, 包含, 动态规划 \rangle 却有极大不同。而 TransR 通过矩阵 \mathbf{M} 将实体从实体空间转到关系空间, 并在关系空间中进行计算, 从而更精确地区分两个语义相近的实体在某些关系上的不同。

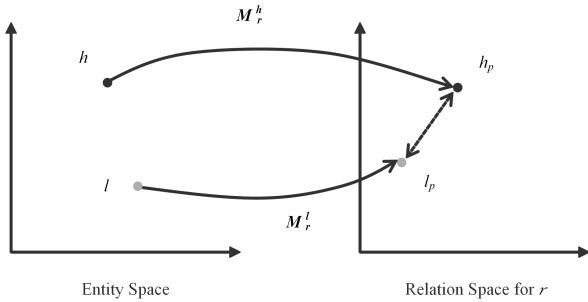


图 4 改进的 TransR 模型

Fig. 4 Improved TransR model

具体算法是将头实体向量 \mathbf{h} 和尾实体向量 \mathbf{l} 映射为 $\mathbf{h}_r = \mathbf{M}_r \mathbf{h}$ 和 $\mathbf{l}_r = \mathbf{M}_r' \mathbf{l}$, 因此 TransR 模型的损失函数如式(8)所示:

$$f_r(\mathbf{h}, \mathbf{l}) = \|\mathbf{h}_r + \mathbf{r} - \mathbf{l}_r\|_{l_1/l_2} \quad (8)$$

即计算 $\mathbf{h}_r + \mathbf{r}$ 到 \mathbf{l}_r 的 1-范数或 2-范数距离。

在 TransR 的基础上, 考虑到真实数据集中头尾实体数量会有较大差异, 如 \langle 题目, 包含, 知识点 \rangle 的三元组中题目实体的数量远大于知识点实体的数量, 最终可能会出现数量多的实体过拟合而数量少的实体欠拟合的情况。为了防止计算结果出现较大误差, 本文模型在原有 TransR 基础上参考 TransSparse 模型^[23], 对翻译模型进行了改进, 即将原有的变换矩阵 \mathbf{M} 更换为能够降低自由度的稀疏矩阵进行空间转换。模型在头尾实体空间转换时需要使用不同的矩阵, 设 θ 为稀疏矩阵的稀疏因子, \mathbf{M}_r^h 和 \mathbf{M}_r^l 分别表示头实体与尾实体的投影矩阵, 则头尾实体转换公式如式(9)、式(10)所示:

$$\mathbf{h}_p = \mathbf{M}_r^h(\theta_r^h) \mathbf{h} \quad (9)$$

$$\mathbf{l}_p = \mathbf{M}_r^l(\theta_r^l) \mathbf{l} \quad (10)$$

两个矩阵的稀疏度由式(11)决定。

$$\theta_r = \{1 - (1 - \theta_{\min}) N_r^i / N_r^{\max}\}_{i=h, l} \quad (11)$$

其中, N_r^i 表示关系 r 连接的头实体与尾实体数量, 最大数量的头尾实体分别对应为 N_r^{\max} 和 N_r^{\min} , θ_{\min} 表示最大值的情况下投影矩阵的稀疏因子。则最后的得分函数可以修改为式(12)。

$$f_r(\mathbf{h}, \mathbf{l}) = \|\mathbf{h}_p + \mathbf{r} - \mathbf{l}_p\|_{l_1/l_2} \quad (12)$$

由于图谱中都是真实的三元组, 因此在进行负采样时需要随机替换头尾实体, 损失函数可推导为式(13)。

$$\mathcal{L}_{kg} = \sum_{\langle h, r, l \rangle \in S} \sum_{\langle h', r, l' \rangle \in S'} [\gamma + f_r(\mathbf{h}, \mathbf{l}) - f_r(\mathbf{h}', \mathbf{l}')] \quad (13)$$

3.2.3 引入注意力机制的向量融合

由知识图谱表征学习单元和协同过滤单元分别计算得出稠密嵌入向量后, 便可在精准排序单元进行向量融合。向量融合过程中可以直接使用简单的加和方式来聚合向量信息, 或者加入平均池化层取向量的平均值, 但这样的做法会使向量中大量的特征信息丢失。本文引入注意力机制, 对用户与题目在知识图谱表征和协同过滤分解过程中得出的属性角度和交互角度的向量分别进行特征交互, 为特征向量加上权重 a_{ij} 后进行向量的加权平均操作。由于不同的交叉特征对结果的影响程度不同, 注意力单元通过学习可以很好地捕捉到影响因子的大小, 从而在不损失特征信息的情况下完成向量融合。

属性角度和交互角度的用户与物品向量通过元素积的方式实现特征交互, 其过程如式(14)所示:

$$f_{PI}(\boldsymbol{\varepsilon}) = \{\mathbf{u}_i \odot \mathbf{t}_j\}_{i, j=cf, kg} \quad (14)$$

设用户对题目的最终兴趣分数为 y_{ut} , 加入注意力得分后的平均加权过程如式(15)所示:

$$\hat{y}_{ut} = f_{\Lambda}(f_{PI}(\boldsymbol{\varepsilon})) = \sum_{i, j=cf, kg} (\mathbf{u}_i \odot \mathbf{t}_j) a_{ij} \quad (15)$$

本文模型中的注意力网络采用的是单全连接层与 softmax 输出层相叠加的结构。需要学习的参数为单层注意力网络的权重矩阵 \mathbf{N} 、偏置向量 \mathbf{b} 以及全连接层到输出层的权重向量 \mathbf{h}_a 。此外, 得出的注意力参数还需要通过 softmax 函数进行归一化。最终注意力单元的参数学习过程如式(16)、式(17)所示:

$$a'_{ij} = \mathbf{h}_a^T \text{Relu}(\mathbf{N}((\mathbf{u}_i \odot \mathbf{t}_j) + \mathbf{b})) \quad (16)$$

$$a_{ij} = \frac{\exp(a'_{ij})}{\sum_{i, j=cf, kg} \exp(a'_{ij})} \quad (17)$$

最终的目标函数如式(18)所示, 通过梯度下降方法即可学习优化公式中参数, 从而得到兴趣值 \hat{y}_{ut} 并排序得出 Top-K 推荐结果。

$$\mathcal{L}_a = \sum_{u, t \in K} (\hat{y}_{ut} - y_{ut})^2 + \lambda_a \|\mathbf{N}\|^2 \quad (18)$$

4 实验验证

本文实验依托北京航空航天大学在线评测系统 AcCoding, 采用常规离线评估方案的同时, 也在推荐算法上线后进行了线上 A/B 测试评估, 综合评估了 HKGCF 模型的准确率和可解释性。

4.1 准确率离线评估

4.1.1 数据集

为测试模型的有效性, 本文采用北京航空航天大学在线评测系统 AcCoding 的用户-题目交互行为数据集和题目特征信息数据集, 数据集记录了 2015 年以来, 20000 余名用户、约 5000 道题目、上百万条的交互记录以及题目的特征详细信息, 足以支撑对本文算法的验证与分析。由于系统在 2015—2016 年间用户注册数量较少, 交互数据异常稀疏, 因此采用

2017年以来的用户-题目交互行为数据作为本文研究数据集。

同时,本文通过数据预处理,将部分不符合规则的题目特征信息和同一代码多次提交的问题交互记录进行清洗,并对数据进行划分。我们将80%的数据集作为训练集用于训练模型,将20%的数据集作为测试集用于评价算法性能。

数据集的详细信息如表5所列。

表5 数据集信息

Table 5 Dataset information

Datasets		Size
User-Topic interaction behavior dataset	Users	26901
	Tasks	4653
	Effective interactions	2722398
Topic feature information dataset	Entities	5323
	Entity types	4
	Relations	18656
	Relation types	4

4.1.2 实验环境与参数设置

本文模型的硬件环境为Intel Core i5-8250U@ 1.60GHz的CPU与8GB内存,操作系统为Ubuntu 18.04,实现语言为Python,使用TensorFlow框架。

本文模型在进行实验前,使用Xavier初始化方法^[24]获得模型初始参数,并设置学习率 L_r 为0.005,批次大小batch为1024,正则化系数 λ 为 10^{-5} , λ_s 为 10^{-3} 的参数设置,且通过实验验证模型中各类参数皆为最优设置。

4.1.3 实验结果评价体系

评估推荐结果准确程度的指标为在不同 K 近邻下的Top- K 推荐场景下的模型精确率(precision),以及点击通过率(Click-Through-Rate,CTR),预估场景下的F1-score和曲线下面积(Area Under Curve,AUC)。

F1-score是综合反映精确率和召回率(recall)的实验指标,为二者的调和平均值,因此能够合理地评估推荐算法。其中,精确率指推荐正确的题目占被模型分类器判定为正样应推荐题目个数的比例,召回率体现了推荐正确的题目占真正总推荐题目个数的比例。该指标可以从混淆矩阵中推导得出,结果如表6所列。

表6 混淆矩阵

Table 6 Confusion matrix

Results	Sample Adopted	Sample Not Adopted
Sample Recommended	TP	FP
Sample Not Recommended	FN	TN

其中,TP(True Positive)表示题目被推荐并且被用户采用的样本,TN(True Negative)表示题目不被推荐而用户自己也没有选用的样本,FP(False Positive)表示题目被推荐但不被用户采用的样本,FN(False Negative)表示题目不被推荐但用户选用的题目样本。在Top- K 的推荐场景中,认定最终的Top- K 结果即是模型判定的正样本数。则推荐系统精确率、召回率和F1-score可分别表示为式(19)式(21)。

$$Precision = \frac{TP}{TP + FP} \quad (19)$$

$$Recall = \frac{TP}{TP + FN} \quad (20)$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (21)$$

AUC指标表示受试者工作特征曲线(the Receiver Operating Characteristic,ROC)下的面积,其横坐标为假阳性率(False Positive Rate,FPR),纵坐标为真阳性率(True Positive Rate,TPR),其计算式如式(22)所示:

$$FPR = \frac{FP}{N} \quad (22)$$

$$TPR = \frac{TP}{P}$$

该指标反映了当任意取一个正样本和负样本时,分类器予以正样本的得分大于负样本的概率,因此可以准确反映出模型的性能。

4.1.4 对比方法

为了验证实验结果的有效性,将本文算法与经典协同过滤推荐算法进行了实验结果的对比,同时也包括近年来较为先进的协同过滤或注意力网络与知识图谱相结合的可解释算法CKE,KGAT和Ripplet等,以评估本文模型的性能。具体的对比方法的介绍如下。

(1)基于物品的协同过滤算法(ItemCF)。基于用户相似度的协同过滤算法,适合具有较强社交特性的系统进行热点挖掘并跟踪热点趋势。

(2)基于用户的协同过滤算法(UserCF)。该算法是互联网推荐场景中应用最广、效率最高的推荐算法,基于物品相似度计算进行Top- K 推荐,能够初步挖掘用户的兴趣偏好。

(3)RKGE(Recurrent Knowledge Graph Embedding)。本文模型提出了基于循环神经网络的知识图谱元路径推荐方法,能够较为准确地提取推理路径信息,在准确率的提升和解决冷启动问题上有很好的表现。

(4)CKE。该模型将知识图谱引入到推荐系统中,借助图谱中实体的内在关联提升推荐质量,相比基础的矩阵分解模型,推荐效果得到了明显的提升。

(5)RippleNet。该模型使用知识图谱结构并关注特征的偏好传播,充分利用了知识图谱内的关系信息,将每一层的实体向量进行加权池化,更为细致地刻画了用户偏好,能够为用户提供推荐解释。

(6)KGAT(Knowledge Graph Attention Network for Recommendation)。该模型提出了一种融合知识图谱特征及注意力网络的推荐算法,弥补了深度推荐算法忽略了用户交互数据间关系的不足。

4.1.5 实验结果分析

Top- K 推荐场景下的实验结果如表7和图5所示。CTR预估场景下的实验结果如表8所列。

表7 算法精确度对比

Table 7 Algorithm accuracy comparison

Model	Precision@1	Precision@5	Precision@10	Precision@20
ItemCF	0.1534	0.1442	0.1393	0.1011
UserCF	0.1425	0.1453	0.1386	0.1032
RKGE	0.1091	0.0951	0.0753	0.0746
CKE	0.1023	0.0977	0.0792	0.0705
RippletNet	0.1150	0.1042	0.0959	0.0894
KGAT	0.1676	0.1574	0.1202	0.1181
HKGCF	0.1883	0.1723	0.1672	0.1525

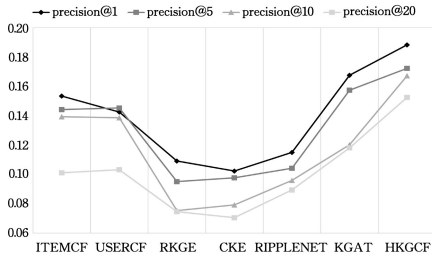


图 5 算法精确度对比

Fig. 5 Algorithm accuracy comparison

表 8 CTR 预估场景下的实验结果对比

Table 8 Results comparison under CTR prediction

Model	F1	AUC
ItemCF	0.1103	0.6246
UserCF	0.0956	0.5988
RKGE	0.0879	0.5834
CKE	0.0857	0.6037
RippleNet	0.0843	0.5712
KGAT	0.1249	0.6468
HKGCF	0.1344	0.6509

从图 5、表 7 中可以看出, HKGCF 模型与 ItemCF, UserCF, RKGE, CKE, RipPLENET, KGAT 等推荐算法相比, 不同 K 近邻下的平均准确率分别提高了 3.6%, 3.8%, 8.2%, 8.3%, 2.9% 和 6.9%。根据实验结果的分析对比可以发现, 较为成熟的可解释推荐算法 (CKE, RipPLENET) 为了保证可解释性而牺牲了准确率, 其推荐效果不如融合了协同过滤思想的算法 (ItemCF, UserCF, CKE, HKGCF); 同时, 加入了注意力机制的推荐算法 (KGAT, HKGCF) 在精确率上都取得了更好的结果, 说明结合了注意力机制的推荐算法更能够精准表达用户和物品的特征信息。

从表 8 中可以看出, HKGCF 模型的推荐表现在 F1-score 与 AUC 指标下都具有明显优势。同时, KGAT 与 HKGCF 表现接近且皆优于其他对比算法, 说明注意力机制与知识图谱的融合能够一定程度上提升推荐算法的性能。实验结果表明, 本文模型通过有效结合交互角度和属性角度两方面的用户与题目的特征向量, 可以更好地捕捉用户兴趣, 得出最优的推荐结果。

4.2 线上 A/B 测试

由于单独使用离线评估并不能完全模拟线上的实际项目环境, 因此本文也通过将线上用户随机分成实验组和对照组, 施以不同的推荐算法对本文 HKGCF 模型进行全面评估。

在 A/B 测试过程场景中, 对具有相似用户画像的活跃

用户以 1:1 的比例分别施以传统协同过滤算法与 HKGCF 算法的推荐场景, 通过观察用户的题目点击率和通过率来对比判断两者的推荐结果是否符合用户需求。在持续两个月的线上测试中, HKGCF 推荐场景中的用户点击率和通过率各以 10.8% 和 9.5% 的优势超过协同过滤推荐场景, 具体数据如表 9 所列。因此, 可以确认 HKGCF 得出的 Top-K 推荐结果更能够满足用户的做题需求。

表 9 线上 A/B 测试的实验结果对比

Table 9 Results comparison from online A/B tests

Model	Task Hits	Task Pass Rate
ItemCF	0.1540	0.1034
HKGCF	0.2621	0.1981

4.3 可解释性评价

推荐结果是否具有可解释性可通过推荐解释的生成情况与用户体验测试两方面进行评判。

HKGCF 模型采用的具有高度可解释性的协同过滤算法和知识图谱集成结构可以根据最终排序单元中的注意力权重系数的大小, 反映出不同单元提供的特征信息的影响力高低, 从而为后续推荐解释路径的形成提供依据。系统通过基于规则的后处理式推荐解释方法生成用户与被推荐物品间的关联路径, 使用户与题目之间产生从推荐结果直接通往推荐理由的推理路径, 最终达到教学推荐精准性和可解释性的统一, 推荐解释的生成过程如图 6 所示。

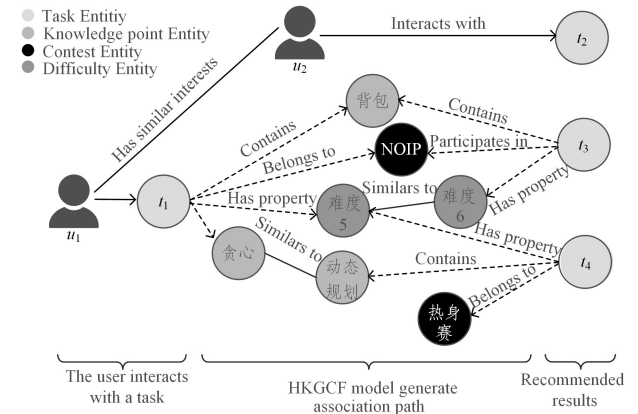


图 6 推荐解释的生成过程

Fig. 6 Process of generating recommendation explanation

为了验证本文模型具有良好的可解释性, 通过随机采样抽取多个测试样本进行 Top-5 推荐, 其中一个采样样本的验证结果如表 10 所列。

表 10 推荐解释实验验证结果

Table 10 Validation results of recommendation explanation experiment

Recommended Tasks	Highest Weight Unit	Recommendation Explanations
Repeat String Pairs	Collaborative filtering unit	A classmate with similar interests to yours also did this programming task
Tree Structure	Knowledge graph feature learning unit	Based on the programming task you did, "Mdd's Chain Table", we recommend this programming task with similar difficulty
Minimum spanning tree algorithm	Knowledge graph feature learning unit	Based on your mastered knowledge point of the "KD tree", we recommend this programming task with similar knowledge point
Multipacks	Knowledge graph feature learning unit	Based on your mastered knowledge point of the "Knapsack problem", we recommend this programming task with similar knowledge point
Crowdsourcing Issues	Collaborative filtering unit	A classmate with similar interests to yours also did this programming task

同时,实验通过设计问卷的形式对使用 CKE,RippleNet 与 HKGCF 推荐算法分别获得推荐结果的 57 名用户进行了用户问卷调查,通过用户对 4 个问题的打分结果进行量化来评估本文算法的用户满意度和可解释性。由于 CKE 与 RippleNet 无法产生推荐解释,因此部分测验结果不纳入量化评分范围。测试的题目包括:

- (1)对推荐结果是否满意?
- (2)推荐的题目是否能够吸引您尝试做题并提交?
- (3)根据推荐结果,对系统推荐此题目的原因是否理解?
- (4)推荐解释是否能帮助你了解被推荐题目的相关信息?

问卷采用五分制,其中 1 至 5 分分别表示“非常不认同”“不认同”“不确定”“认同”及“非常认同”,其测试结果如表 11 所列。

表 11 用户问卷调查结果
Table 11 Results of user survey

Model	Qa	Qb	Qc	Qd
CKE	3.521	2.631	1.830	—
RippleNet	3.493	2.711	1.225	—
HKGCF	3.965	3.052	4.137	3.406

根据以上实验和评估可以发现,HKGCF 模型能够生成用户易于理解的推荐解释,其推荐结果具有良好的可解释性。

4.4 实验结果总结

本文通过离线实验及线上 A/B 测试两种方式对所提算法进行了测试和评价。其中,离线实验在 Top-K 推荐与 CTR 预测两种场景下,将本文设计的 HKGCF 算法与 6 种广泛使用的现有算法进行了比对,结果发现 HKGCF 模型以平均精确率 17.01%、F1-score 值 0.1344 及 AUC 值 0.6509 的结果优于所有对比算法。线上 A/B 测试发现,在用户题目点击率和题目通过率方面,使用 HKGCF 算法的推荐比经典协同过滤算法有明显提升。因此,本文算法在编程题目推荐场景中具有优秀的准确度。

此外,本文使用随机采样验证和用户问卷调查的方法,验证了本文模型具有良好的可解释性。本文设计的 HKGCF 算法,能够使推荐题目通过关联路径推导产生推荐解释,用户满意度调查优于其他对比算法。

综上所述,本文设计改进的 HKGCF 模型由于能够有效挖掘隐藏在交互行为和属性信息后的用户潜在兴趣,在保证推荐结果准确率的同时,具有良好的可解释性,在在线编程评测系统推荐场景中具有创新性和先进性。

结束语 本文设计了一种基于知识图谱与协同过滤混合策略的在线编程评测系统推荐模型。该模型使用协同过滤算法发掘用户在交互过程中的潜在兴趣,借助知识图谱结构挖掘用户的做题偏好,最终利用注意力机制将不同维度的特征向量进行融合,得到具有高准确率和可解释性的个性化推荐结果。实验结果表明,该编程题目推荐算法的准确性和可解释性均优于现有典型模型。

未来工作将主要考虑利用用户交互行为的时序信息捕捉用户的动态兴趣变化和知识掌握程度,以提升现有模型的表现,并探索将 HKGCF 算法应用于其他在线教育推荐场景的可能性。

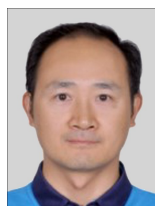
参考文献

- [1] MIRZAYANOV M, PAVLOVA O, MAVRIN P, et al. Codeforces as an Educational Platform for Learning Programming in Digitalization [J]. *Olympiads in Informatics*, 2020, 14(10): 133-142.
- [2] LI W X, GUO W. Peking university online judge and its applications [J]. *Journal of Changchun Post and Telecommunication Institute*, 2005(S2): 170-177.
- [3] PARK Y. Predicting personalized student performance in computing-related majors via collaborative filtering [C] // *Proceedings of the 19th Annual SIG Conference on Information Technology Education*. Florida: ACM Press, 2018: 151-151.
- [4] RECHKOSKI L, AJANOVSKI V V, MIHOVA M. Evaluation of grade prediction using model-based collaborative filtering methods [C] // *2018 IEEE Global Engineering Education Conference (EDUCON)*. Santa Cruz de Tenerife: IEEE Press, 2018: 1096-1103.
- [5] SU X, KHOSHGOFTAAR T M. A survey of collaborative filtering techniques [J]. *Advances in artificial intelligence*, 2009, 2009(1): 1-19.
- [6] WU Q, HUANG M, MU Y. A Collaborative Filtering Algorithm Based on User Similarity and Trust [C] // *2017 14th Web Information Systems and Applications Conference (WISA)*. Liuzhou: IEEE Press, 2017: 263-266.
- [7] YU X, CHEN W. Research on three-layer collaborative filtering recommendation for Online Judge [C] // *2016 Seventh International Green and Sustainable Computing Conference (IGSC)*. Hangzhou: IEEE Press, 2016: 1-4.
- [8] HE M, SUN W, XIAO R. A collaborative filtering recommendation algorithm fusing clustering and user interest preferences [J]. *Computer Science*, 2017, 44(S2): 391-396.
- [9] HE M, XIAO R, LIU W S, et al. Collaborative filtering recommendation algorithm fusing category information and user interest degree [J]. *Computer Science*, 2017, 44(8): 230-235, 269.
- [10] KIM B H, VIZITEI E, GANAPATHI V. GritNet: Student performance prediction with deep learning [J]. *arXiv:1804.07405*, 2018.
- [11] SAITO T, WATANOBE Y. Learning Path Recommendation System for Programming Education Based on Neural Networks [J]. *international journal of distance education technologies*, 2020, 18(1): 36-64.
- [12] XIAO J, YE H, HE X, et al. Attentional factorization machines: Learning the weight of feature interactions via attention networks [J]. *arXiv:1708.04617*, 2017.
- [13] ZHOU G, ZHU X, SONG C, et al. Deep interest network for click-through rate prediction [C] // *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. London: ACM Press, 2018: 1059-1068.
- [14] CHOUNTA I A, CARVALHO P F. Square it up! How to model step duration when predicting student performance [C] // *Proceedings of the 9th International Conference on Learning*

- Analytics & Knowledge. Tempe; ACM Press, 2019: 330-334.
- [15] LIANG H H, GU T L, BING C Z, et al. Joint learning of user-side and project-side knowledge graphs for personalized recommendation [J]. *Computer Science*, 2021, 48(5): 109-116.
- [16] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data [J]. *Advances in neural information processing systems*, 2013, 26(1): 2787-2795.
- [17] WANG Z, ZHANG J, FENG J, et al. Knowledge graph embedding by translating on hyperplanes [C] // *Proceedings of the AAAI Conference on Artificial Intelligence*. Quebec City: AAAI Press, 2014, 28(1): 1112-1119.
- [18] CHEN X J, XIANG Y. STransH: An improved knowledge representation model based on translation model [J]. *Computer Science*, 2019, 46(9): 184-189.
- [19] WANG H, ZHANG F, ZHAO M, et al. Multi-task feature learning for knowledge graph enhanced recommendation [C] // *The World Wide Web Conference*. San Francisco: ACM Press, 2019: 2000-2010.
- [20] PALUMBO E, MONTI D, RIZZO G, et al. entity2rec: Property-specific knowledge graph embeddings for item recommendation [J]. *Expert Systems with Applications*, 2020, 151(15): 1-18.
- [21] ZHANG F, YUAN N J, LIAN D, et al. Collaborative knowledge base embedding for recommender systems [C] // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco: ACM Press, 2016: 353-362.
- [22] KANG Y, LI T, LI H, et al. A recommendation model combining knowledge graph and collaborative filtering [J]. *Computer Engineering*, 2020, 46(12): 73-79, 87.
- [23] JI G, LIU K, HE S, et al. Knowledge graph completion with adaptive sparse transfer matrix [C] // *Thirtieth AAAI Conference on Artificial Intelligence*. Phoenix: AAAI Press, 2016: 985-991.
- [24] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks [C] // *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, 2010: 249-256.



LIU Zejing, born in 1997, postgraduate. His main research interests include recommendation system and knowledge graph.



SONG You, born in 1973, Ph.D, professor. His main research interests include data analysis techniques, information processing, knowledge graph, behavior prediction for specific fields.

(责任编辑: 喻藜)