

融合偏好传播的多任务推荐模型研究

杨本臣¹, 叶洪宇¹, 孟祥福²

(1. 辽宁工程技术大学 软件学院; 2. 辽宁工程技术大学 电子与信息工程学院, 辽宁 葫芦岛 125105)

摘要: 针对知识图谱可以有效地从多源异构数据中还原出实体的三元组关系, 却不利于推荐任务且采用单任务学习又很难挖掘数据潜在关联的问题, 提出一种融合偏好传播的多任务推荐模型(MAPKR)。首先, 利用涟漪传播从知识图谱中提取用户的偏好特征集; 其次, 依据相似近邻结构实现潜在特征共享, 经交叉压缩单元提取项目和高阶特征表示; 最后, 以多任务学习交替训练推荐模块和知识图谱嵌入模块, 将提取的特征向量经归一化内积操作后进行预测、推荐。在3个公开数据集上进行实验并与5个基线模型进行比较。与MKR和RippleNet相比, 在MovieLens-1M数据集上, AUC和ACC分别提高了0.68%、0.31%和0.77%、0.54%; 在Book-Crossing上, AUC和ACC分别提高了3.48%、2.66%和4.51%、7.21%; 在Last.FM上, AUC和ACC分别提高了3.44%、6.25%和2.70%、2.62%。实验结果表明, 提出的模型与MKR、RippleNet等其他基线模型相比推荐性能良好。

关键词: 推荐系统; 深度学习; 知识图谱; 偏好传播; 多任务学习

DOI: 10.11907/rjdk.231526

开放科学(资源服务)标识码(OSID):



中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2024)006-0009-09

Research on Multi-task Recommendation Model Fused with Preference Propagation

YANG Bencheng¹, YE Hongyu¹, MENG Xiangfu²

(1. School of Software, Liaoning Technical University; 2. School of Electronic & Information Engineering, Liaoning Technical University, Huludao 125105, China)

Abstract: To address the problem that the knowledge graph can effectively reduce the triadic relationships of entities from multi-source heterogeneous data, but is not conducive to recommendation tasks and it is difficult to explore the potential association relationships of data using single-task learning, a multi-task recommendation model with fused preference propagation (MAPKR) is proposed. Firstly, the user's preference feature set is extracted from the knowledge graph using ripple propagation; secondly, the potential features are shared based on the similar nearest neighbor structure, and the higher-order feature representations of items and entities are extracted by cross-compression units; finally, the recommendation module and the knowledge graph embedding module are trained alternately with multi-task learning, and the extracted feature vectors are predicted and recommended after normalized inner product operation. Experiments are conducted on three publicly available datasets and compared with five baseline models. Compared with MKR and Ripple Net, the AUC and ACC are improved by 0.68%, 0.31% and 0.77%, 0.54% on MovieLens-1M dataset; 3.48%, 2.66% and 4.51%, 7.21% on Book-Crossing, respectively; on Last.FM, AUC and ACC improved by 3.44%, 6.25% and 2.70%, 2.62%, respectively. The experimental results show that the proposed model has good recommendation performance compared with other baseline models such as MKR and RippleNet.

Key Words: recommendation system; deep learning; knowledge graph; preference propagation; multi-task learning

收稿日期: 2023-05-31

基金项目: 国家自然科学基金面上项目(61772249)

作者简介: 杨本臣(1975-), 男, 辽宁工程技术大学软件学院副教授、硕士生导师, 研究方向为大数据、物联网、图像处理; 叶洪宇(1998-), 男, 辽宁工程技术大学软件学院硕士研究生, 研究方向为推荐系统; 孟祥福(1981-), 男, 辽宁工程技术大学电子与信息工程学院教授、博士生导师, 研究方向为大数据分析、时空数据管理、Web数据库top-k查询、中医药知识图谱、大数据分析可视化技术。

0 引言

信息爆炸时代,如何高效地获取用户偏好信息是人们关注的焦点问题,由此能够快速、准确地筛选出用户偏好信息的推荐系统应运而生。相对于用户明确的需求,它从用户历史行为中获取偏好,主动地为用户进行个性化推荐^[1]。然而,以协同过滤为代表的传统推荐方法虽然有效,但在处理多源异构数据时却很难挖掘到数据之间深层关联,稀疏性和冷启动等问题逐渐暴露。为解决这些问题,研究人员尝试在推荐系统中融入包括社交网络^[2]、属性^[3]、多媒体^[4](例如文本、图像)等多源异构辅助信息(Side Information)。

知识图谱(Knowledge Graph, KG)作为一个用节点和边分别表示实体及实体间关系的有向异构图,包含丰富的项目属性和关系^[5]。以知识图谱为辅助信息的推荐模型能充分挖掘知识图谱中所蕴涵的信息,并结合已有信息增强数据维度并提高推荐准确性,从而缓解数据稀疏和冷启动等问题。例如,Zhang等^[6]提出的CKE模型,采用TransR^[7]方法从知识图谱中学习物品潜在表示,提高了推荐性能。Zhu等^[8]利用实体间的高阶关系捕获知识图谱中的语义信息,提出了一个知识图注意力网络增强序列推荐的框架。Wang等^[9]提出一种知识图谱卷积网络KGCN,利用用户特定的关系选择性地迭代聚合知识图谱上的邻域信息,建模高阶邻近信息,从而捕捉远距离用户间的潜在偏好,其通过挖掘知识图谱中的相关属性以有效获得各项目之间的相关性,有效提高了推荐准确率。

然而,上述工作虽然辅助增强了推荐任务,提高了推荐性能,但由于大多数采用单任务学习方式,导致模型泛化能力差。与单任务学习不同,多任务学习凭借信息共享能极大程度减少数据量和参数规模^[10],提高模型学习效率。因此,为弥补上述方式各自的不足,本文提出了一种融合偏好传播的多任务推荐模型。首先,利用涟漪传播思想获取用户的向量表示;其次,通过交叉压缩单元提取项目和实体的潜在特征向量与涟漪传播共同促进用户和实体的信息交互;最后,以多任务学习的方式进行训练,将提取的特征向量经归一化内积操作后进行预测、推荐。实验结果表明,在推荐系统中融入知识图谱和多任务学习能保证推荐结果更加精确的同时也更具可解释性。

本文主要贡献有:①提出了一种基于涟漪传播思想的多任务推荐算法,将知识图谱引入推荐系统构建成协同知识图谱,利用涟漪传播思想挖掘用户和实体的潜在关联,极大提升了模型推荐效果;②在神经网络中,依据知识图谱项目间相似的近邻结构,利用交叉压缩单元和涟漪传播促进用户和实体的信息交互,增强挖掘用户和项目高阶特征的能力,从而提高推荐性能,增强系统可解释性,最后采用多任务学习方式训练,提高了模型的学习能力;③

在3个真实数据集上进行点击率预测和Top-K推荐实验,与其他基线模型进行比较,证明了该模型的有效性;④在MovieLens-1M数据集上模拟稀疏场景进行实验,证明了模型在稀疏场景下仍然具有良好的推荐性能。

1 相关工作

1.1 融合多任务学习的推荐模型

自推荐系统诞生以来,传统的推荐算法可以分为基于内容的推荐^[11-12]、协同过滤^[13-15]及混合推荐^[16],30多年的发展使得这些传统推荐算法取得了不错的进展。例如,He等^[17]运用神经网络方法极大提升了推荐算法精确度。Gai等^[18]在传统的推荐算法领域进行创新提出了MLR算法,在广告领域对CTR预估算法创新性升级。Naumov等^[19]提出了DLRM模型,利用深度学习模型完成训练。然而,上述工作均采用单任务学习方式,具有一定局限性。而多任务学习通过共享表示信息同时学习多个相关任务,往往比单任务推荐能取得更好的效果。例如,Caruana^[20]基于多任务学习的概念提出了一种多任务学习算法,通过在3个领域进行实验展现了强大的学习能力。申联等^[21]采用深度多任务学习方法同时训练图像美学评价和情感分析任务。Bachmann等^[22]提出了一种多模态多任务掩码自动编码器的预训练策略。

综上所述,与传统的单任务推荐算法相比,在推荐系统中采用多任务学习方式更能提高模型学习效率。因此,本文采用多任务学习方式辅助增强推荐任务以增强模型泛化性,从而提高模型推荐性能。

1.2 融合知识图谱的推荐模型

随着多源异构数据的不断涌现,以传统协同过滤为代表的浅层模型很难挖掘到数据之间的高阶关系^[23],从而导致推荐性能较差,传统的推荐算法面临着亟需突破的难题。对此,本文主要针对融合知识图谱的推荐模型展开研究。

根据知识图谱使用方式的不同,基于知识图谱的学习方法可以分为:基于嵌入的方法^[24]、基于路径的方法^[25]和基于混合的方法^[26]。基于嵌入的方法直接使用知识图谱信息丰富用户项目表示,不需要手动设计路径。例如,Wang等^[27]提出的深度知识感知网络(DKN)采用知识图谱嵌入方法,利用知识图谱学习到的特征训练实体和关系嵌入,将CNN融合到深度推荐模型中进行学习预测。Esposito等^[28]针对问答系统提出了一种基于词汇资源和词嵌入的混合查询扩展方法。

此外,Zhang等^[29]提出CFKG方法,从用户一项目知识图中学习实体和关系的嵌入表示。然而,这类方法在信息提取过程中,由于异构信息网络中信息呈高维性和异质性,通常要通过知识图谱嵌入(Knowledge Graph Embedding, KGE)方法进行预处理^[30]。此外,该方法在使用知识

图谱辅助推荐任务中更适合链路预测等任务而非推荐任务。基于路径的方法通过构造元路径(Meta-path)^[31]或元图(Meta-graph)^[32]挖掘潜在关系。例如,Yu等^[33]提出的PER模型使用基于元路径的潜在特征表示不同关系用户—项目连通性,虽然充分直观地利用网络结构为系统提供可解释性,却太依赖于手动设计元路径或元图,对于复杂模型实现起来较为困难。鉴于上述两种方式的不足,基于混合的方式借鉴了知识图谱语义嵌入和语义路径模式各自的长处,利用传播思想细化了项目或用户的表示。Wang等^[34]提出的RippleNet模型就采用了混合的方式,将基于嵌入和基于路径的方法相结合,根据用户历史行为在知识图谱上模仿波纹传播从而获取用户的潜在偏好集。该模型更关注用户—项目交互矩阵,但却忽略了项目特征对用户兴趣建模的重要性。此外,Wang等^[35]提出了一种以知识图嵌入辅助增强推荐任务的多任务训练模型MKR,通过共享潜在特征实现高阶项目—实体信息交互,从而获得高阶特征表示,提升了系统的性能。该模型更注重知识图谱结构信息的交替学习模式,但忽略了评分矩阵中的关键信息及实体间的邻域关系。

综上可知,基于嵌入的方式在链路预测等方面具有独特的优势却不利于推荐任务,而基于路径的方式能够充分利用知识图谱的结构信息使得结果更具解释性,然而却太依赖于手动设计元路径或元图。因此,本文受上述文献启发汲取了各自的优势,采用基于混合的方法在知识图谱中利用涟漪传播思想获取实体及其邻域的关系,并采用多任务学习的方式进行训练使得模型更具泛化性,最终得到更加准确且更具解释性的结果。

2 模型框架

2.1 符号定义

本文用户集和项目集分别使用 U 和 I 进行表示,将基于隐式反馈的用户项目交互矩阵定义为 Y 。其中, $U = \{u_1, u_2 \dots u_M\}$, $I = \{i_1, i_2 \dots i_N\}$, $Y = \{y_{ui} | u \in U, i \in I\}$ 且 Y 同时满足以下条件:

$$y_{ui} = \begin{cases} 1, & (u, i) \text{ 存在交互} \\ 0, & (u, i) \text{ 无交互} \end{cases} \quad (1)$$

当 $y_{ui} = 1$ 时,表示用户和项目存在如点击、浏览等隐式交互;若不存在则 $y_{ui} = 0$ 。此外,本文将知识图谱定义为 G ,用 E 和 R 表示 G 中的实体集和关系集。 G 由关系—实体三元组(head, relation, tail)构成,简写为 (h, r, t) 。其中, $h \in E, r \in R, t \in E$,分别代表三元组中的头部实体、实体间关系和尾部实体。

MAPKR模型根据交互矩阵 Y 和知识图谱 G ,通过预测函数 $F(u, v | \theta, Y, G)$ 预测用户 u 是否对未交互的项目 i 产生潜在交互兴趣的概率 y_{ui}^Δ , θ 表示 F 的参数,表示为:

$$y_{ui}^\Delta = F(u, v | \theta, Y, G) \quad (2)$$

2.2 MAPKR模型框架

本文提出的MAPKR模型通过将交替学习和联合学习两种特征学习方法同时集成到系统中,在知识图谱上传播偏好时无需手动设计元路径,极大程度地挖掘了实体深层关系,优化了挖掘效率,提高了推荐质量。MAPKR模型框架主要包括推荐模块、知识图嵌入模块(KGE模块)、交叉压缩单元,具体如图1所示。

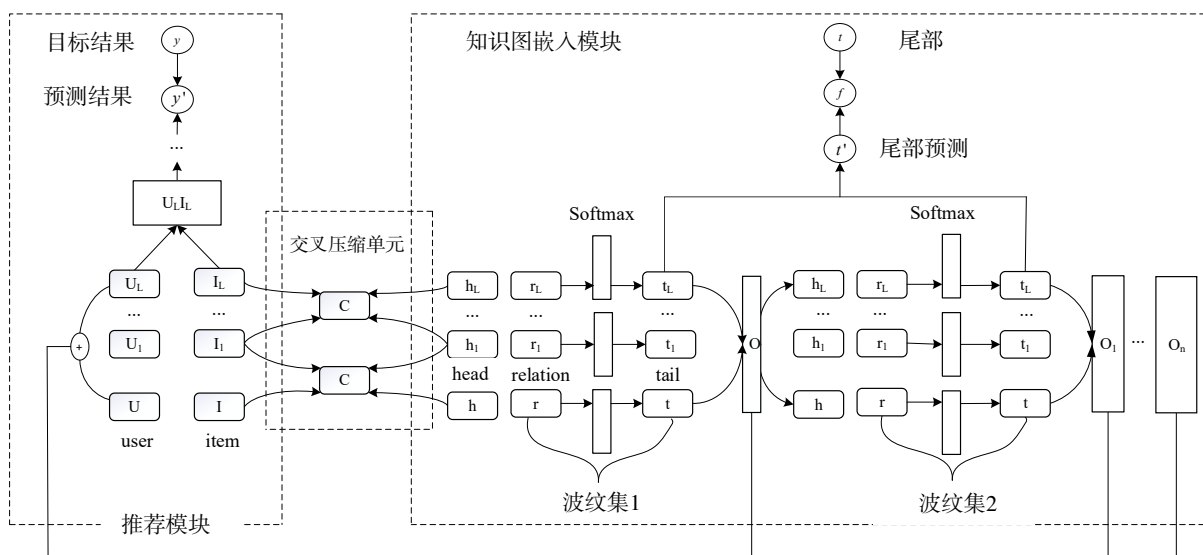


Fig. 1 Framework of MAPKR

图1 MAPKR模型框架

由图1可见,知识图嵌入模块以头部向量 h 和关系向量 r 作为输入, t^Δ 是尾部 t 的输出预测结果;推荐模块以用户 U 的交互历史即经过波纹传播后的用户偏好集和项目 I

作为输入,使用交叉压缩单元进行链接。最后,采用多任务学习方式交替训练从而进行预测、推荐, y_{ui}^Δ 为模型最终预测结果。

2.2.1 交叉压缩单元

交叉压缩单元模块如图2所示。

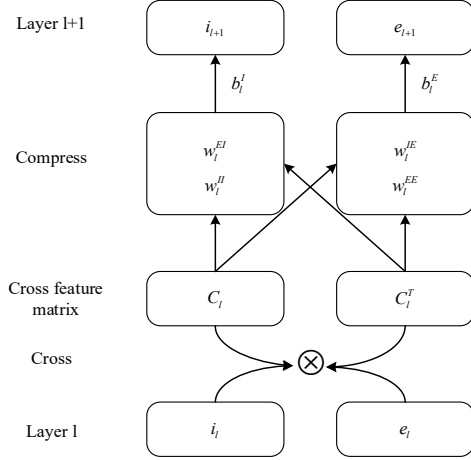


Fig. 2 Cross & compress unit

图2 交叉压缩单元

该模块分为交叉和压缩两部分,以项目 $i \in R^d$ 及其关联实体 $e \in R^d$ 作为输入,首先通过交叉进行链接,构造 L 层特征矩阵 $C_l \in R^{d \times d}$ 。

$$C_l = i_l e_l^T = \begin{bmatrix} i_l^{(1)} e_l^{(1)} & \dots & i_l^{(1)} e_l^{(d)} \\ \vdots & \ddots & \vdots \\ i_l^{(d)} e_l^{(1)} & \dots & i_l^{(d)} e_l^{(d)} \end{bmatrix} \quad (3)$$

然后通过压缩将得到的交叉特征矩阵投影到潜在空间输出下一层项目和实体的特征向量:

$$\begin{bmatrix} i_{l+1} \\ e_{l+1} \end{bmatrix} = \begin{bmatrix} w_l^{II} & w_l^{EI} \\ w_l^{IE} & w_l^{EE} \end{bmatrix} \begin{bmatrix} C_l \\ C_l^T \end{bmatrix} + \begin{bmatrix} b_l^I \\ b_l^E \end{bmatrix} \quad (4)$$

经整理得:

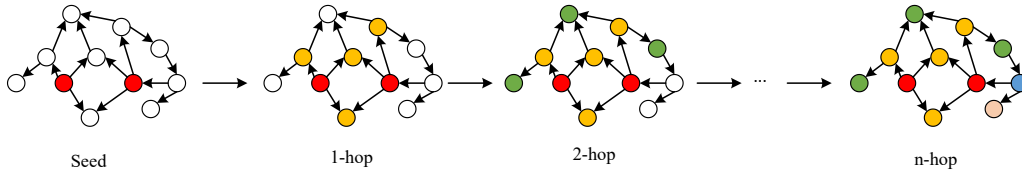


Fig. 3 Propagation process of ripple

图3 涟漪传播流程

将基于用户历史兴趣的潜在偏好(红色标记)作为种子集,在知识图谱上模拟水波逐层进行偏好传播,每一次传播的结果标记为不同的颜色,传播效果随着 n 值的增加逐渐减弱。在涟漪传播过程中,对于给定的知识图谱 G 和交互矩阵 Y ,首先将用户 U 的 n -hop 的相关实体定义为:

$$\varepsilon_u^m = \{t | (h, r, t) \in G, h \in \varepsilon_u^{m-1}\} \quad m = 1, 2, \dots, n \quad (11)$$

其中,当 $m = 0$ 时表示用户 u 的历史兴趣集 I_u ,即 $\varepsilon_u^0 = I_u = \{I | y_{ui} = 1\}$,并将其视为种子然后沿链接扩展。经过 n -hop 后形成多个波纹集 ripple set,将 n -hop 波纹集定义为:

$$S_u^m = \{(h, r, t) | (h, r, t) \in G, h \in \varepsilon_u^{m-1}\} \quad m = 1, 2, \dots, n \quad (12)$$

$$\begin{cases} i_{l+1} = C_l w_l^{II} + C_l^T w_l^{EI} + b_l^I = i_l e_l^T w_l^{II} + e_l i_l^T w_l^{EI} + b_l^I \\ e_{l+1} = C_l w_l^{IE} + C_l^T w_l^{EE} + b_l^E = i_l e_l^T w_l^{IE} + e_l i_l^T w_l^{EE} + b_l^E \end{cases} \quad (5)$$

其中, $w_l^{II} \in R^d$, $b_l^I \in R^d$ 分别代表权重和偏置向量。最终将交叉压缩单元简化为:

$$[i_{l+1}, e_{l+1}] = C(i_l, e_l) \quad (6)$$

2.2.2 KGE 模块

对于给定的三元组 (h, r, t) ,该模块将头部 h 和关系 r 作为输入,首先使用交叉压缩单元和非线性层提取原始特征,然后将 h 和 r 的对应向量进行组合,再通过 k 层 MLP 预测尾部 t :

$$h_l = E_{i-S(h)}[C^L(i, h)][e] \quad (7)$$

$$r_l = \text{Mlp}(\text{Mlp}(\dots \text{Mlp}(r))) = \text{Mlp}^l(r) \quad (8)$$

$$t^\wedge = \text{Mlp}^k\left(\begin{bmatrix} h^l \\ r^l \end{bmatrix}\right) \quad (9)$$

其中, $S(h)$ 是实体 h 的关联集, C^L 表示 L 层的交叉压缩计算, $[e]$ 是经过交叉压缩单元输出的实体特征向量, $t^\wedge$ 是尾部 t 的预测向量, Mlp^l 代表 L 层的多层感知机运算。 $\text{Mlp}(x) = \sigma(wx + b)$ 是以 σ 为非线性激活函数、 w 为权重、 b 为偏置的全连接神经网络。最后,利用相似度函数 f_{kg} 对三元组进行评估:

$$s(h, r, t) = f_{kg}(t, t^\wedge) \quad (10)$$

其中, $f_{kg}(t, t^\wedge) = \sigma(t^T t^\wedge)$, σ 为 sigmoid 激活函数将提取的向量进行归一化内积。

2.2.3 推荐模块

在推荐模块,用涟漪传播代替原有的用户向量表示方法,将得到的最终用户表示和项目 I 共同作为该模块的输入。涟漪传播过程如图3所示。

且对于给定的项目嵌入 I 和用户 u 的 1-hop 波纹集 S_u^1 ,为了比较 I 与 S_u^1 三元组 (h_i, r_i, t_i) 中 h_i, r_i 的相关性,为其分配相关概率:

$$P_m = \text{soft max}(v^T R_m h_m) = \frac{\exp(v^T R_m h_m)}{\sum_{(h, r, t) \in S_u^1} \exp(v^T R_m h_m)} \quad (13)$$

其中, $R_m \in R^{d \times d}$, $h_m \in R^d$ 分别表示关系和头部嵌入,然后将 S_u^1 中尾部及其对应的相关概率进行加权可以得到用户 u 的一阶表示:

$$o_u^1 = \sum_{(h, r, t) \in S_u^1} P_m t_m \quad (14)$$

以此类推,可以得到最终的用户表示:

$$u = \sum_{m=1}^n o_u^m \quad (15)$$

式(11)和式(12)这种将用户的兴趣偏好沿着 S_u^n 中的链接从其历史集 I_u 转移到 n-hop 相关实体集 $\mathcal{E}_u^n$ 的过程称作偏好传播。并且,对于用户 u 和项目 i ,分别使用 L 层 MLP 交叉压缩单元提取 u 和 i 的潜在特征:

$$U_L = \text{Mlp}^L(u) \quad (16)$$

$$I_L = E_{e \sim S(h)}[C^L(i, e)][i] \quad (17)$$

迭代地进行波纹传播,最后通过激活函数预测点击概率:

$$\begin{aligned} y_{ui}^\wedge &= \sigma(U_L^T, I_L^T) \\ &= \sigma \left(\text{Mlp}^L \left(\sum_{n=1}^H \sum_{(h_n, r_n, t_n) \in S_u^n} \frac{\exp(v^T R_m h_m)}{\sum_{(h, r, t) \in S_u^n} \exp(v^T R_m h_m)} \right) \right) \end{aligned} \quad (18)$$

其中, $\sigma(x) = \frac{1}{1 + \exp(-x)}$ 为 Sigmoid 函数。

2.3 算法流程

该模型完整的损失函数包括推荐模块的损失值、KGE 模块的损失值以及为了防止过拟合引入的正则化项 3 部分。具体构造如下:

$$\begin{aligned} \text{Loss} &= \text{loss}_{RS} + \text{loss}_{KGE} + \text{loss}_{REG} \\ &= \sum_{u \in U, i \in I} \varphi(y_{ui}^\wedge, y_{ui}) \\ &\quad + \lambda_1 \left(\sum_{(h', r, t') \notin G} s(h', r, t') - \sum_{(h, r, t) \in G} s(h, r, t) \right) \\ &\quad + \lambda_2 \|w\|_2^2 \end{aligned} \quad (19)$$

其中 u, v 代表用户和项目集, φ 代表交叉熵函数, λ_1, λ_2 为平衡参数。MAPKR 模型算法如下:

输入:交互矩阵 Y , 知识图谱 G

输出:预测函数 $F(u, v | \theta, Y, G)$

1. 初始化所有参数;
2. 根据式(11)–(12)为每个用户 u 计算其相关实体 $\{\mathcal{E}_u^m\}_{m=1}^n$

以及波纹集 $\text{ripple set}\{\mathcal{S}_u^m\}_{m=1}^n$;

3. for 迭代次数 do

//推荐任务

4. for t do

5. 获取交互矩阵 Y 中的正负小批量样本;

6. 从小批量样本中获取所有项目 I 的历史交互集 $e \sim S(I)$;

7. 利用 Adam 算法,通过式(3)–(6), (13)–(19)反向传播更新参数 F ;

8. end for

//知识图谱嵌入任务

9. 获取知识图谱 G 中的正负小批量样本;

10. 从小批量样本中获取所有头部 h 的历史交互集 $i \sim S(h)$;

11. 利用 Adam 算法,通过式(7)–(10), (19)实现梯度下降更新参数 F ;

12. end for

13. return $F(u, v | \theta, Y, G)$

在算法实现上,首先对所有参数进行初始化,构建并获取用户 u 的 n-hop 相关实体 $\{\mathcal{E}_u^m\}_{m=1}^n$ 及其波纹集 ripple

$\text{set}\{\mathcal{S}_u^m\}_{m=1}^n$, 然后进行迭代训练。在每次训练迭代中,为提高计算效率,采用负采样策略。通过正向传播输入交互矩阵 Y 和知识图谱 G , 然后分别从 Y 和 G 中随机抽取小批量正负交互样本以及真/假三元组,基于采样的小批量样本采用梯度下降法通过反向传播更新所有参数,最终得到预测函数 $F(u, v | \theta, Y, G)$ 。其中,推荐任务需要重复训练 t 次 ($t > 1$),参数 t 取当预测结果在指定范围内达到局部最优时所对应的次数。

3 实验与结果分析

为验证模型有效性,本文在电影、书籍、音乐 3 个数据集上进行实验评估,并与其他基线模型进行比较且通过对比实验证明参数选取的有效性,最后给出实验结果与实验分析。本文实验环境设置如表 1 所示。

Table 1 Experimental environment setup

表 1 实验环境设置

软硬件	环境配置
处理器	Intel® Core™ i5-6300HQ CPU@ 2.30GHz
显卡	Nvidia GeForce GTX 960M(4 GB)
操作系统	Windows10 64 位
内存	8 GB
编程语言	Python 3.7
编程环境	Pycharm 2021
深度学习框架	Tensorflow-gpu 1.14.0

3.1 数据集

MovieLens-1M: MovieLens-1M 是电影推荐中广泛使用的基线数据集,由 MovieLens 网站上的大约 100 万个明确评分(范围从 1 到 5)组成。

Book-Crossing: Book-Crossing 数据集包含 Book-Crossing 社区中用户对书籍的 100 多万条显式评分记录(范围从 0 到 10)。

Last.FM: Last.FM 数据集包含 Last.FM 在线音乐系统中的 92 834 个明确的音乐评分和一组 2 000 名用户。

本文对数据集进行了预处理。在 MovieLens 1M 数据集中,将用户对电影评分大于等于 4 的认定为正向交互记作 1,表示为用户的偏好,反之记作 0。且由于 Book-Crossing、Last.FM 两个数据集比较稀疏,因此只要用户参与评分项目,则将交互记作 1,反之记作 0。此外,在构建知识图谱时,使用 Microsoft Satori 为每个数据集进行构建。首先,从整个知识图谱中选择一个置信度大于 0.9 的三元组子集。对于 MovieLens-1M 和 Book-Crossing,从子知识图谱中额外选择了一个三元组的子集,其关系名称分别包含 film 或 book,以进一步减少知识图谱的大小。对于上述三个数据集而言,给定子知识图谱,采用关联词与三元组 (head, film. film. name, tail) (head, book. book. title, tail) 或 (head, type. object. name, tail) 尾部相匹配的方法收集所有有效电影、书籍或音乐家的 ID,并将其与三元组的首、尾相

匹配,最终筛选出最适合的三元组。3个数据集的基本信息如表2所示。

Table 2 Dataset statistics

表2 数据集统计信息

Dataset	users	items	interactions	entities	relations	KG triples
MovieLens-1M	6 036	2 445	753 772	6 279	7	20 195
Book-Crossing	17 860	14 967	139 746	24 039	10	19 793
Last.FM	1 872	3 846	42 346	9 366	60	15 518

3.2 参数设置与评估指标

本文将每个数据集以6:2:2分为训练集、评估集和测试集3部分。超参数设置如表3所示,其中 d 表示嵌入维度, L 表示低层数, H 表示高层数, n 表示跳数, t 表示推荐模块训练次数, μ_1 表示推荐任务的学习率, μ_2 表示KGE任务的学习率, λ_1 表示KGE的权重, λ_2 表示正则化权重, s 表示波纹集大小。超参数的选取通过优化验证集上的AUC加以确定。将训练后的模型应用于测试集,输出结果为预测点击概率。在点击率(CTR)预测中,使用AUC(Area Under Curve)和准确率(Accuracy, ACC)评估预测性能;在Top-K推荐中,本文使用精确率(Precision@K)和召回率(Recall@K)对推荐集进行评估。

Table 3 Hyperparameter settings

表3 参数设置

Dataset	d	L	H	n	t	μ_1	μ_2	λ_1	λ_2	s
MovieLens-1M	16	1	2	2	2	$2e^{-2}$	$1e^{-2}$	$1e^{-7}$	$1e^{-6}$	32
Book-Crossing	8	1	1	3	3	$2e^{-4}$	$2e^{-5}$	$1e^{-5}$	$1e^{-6}$	32
Last.FM	8	2	2	3	2	$3e^{-4}$	$4e^{-4}$	$1e^{-6}$	$1e^{-6}$	32

3.3 对比模型

KGCN^[9]:通过在知识图谱上迭代地聚合邻域信息并学习邻域的结构信息和语义信息,从而发现用户的潜在兴趣。RippleNet^[34]:以联合训练的方式模拟水波在知识图谱上沿着链接不断传播用户兴趣,从而自动学习用户潜在偏好。MKR^[35]:Wang等^[35]提出一种将知识图谱作为辅助信息,以端到端的方式交替训练的深度推荐模型。KGNN-LS^[36]:在KGCN的基础上增加了标签平滑度(LS),以提高模型鲁棒性,防止过拟LibFM^[37]:一种用于CTR场景基于特征的因子分解模型,将用户和项目ID以及通过TransR算法学习的相应实体嵌入连接作为模型输入。MAPKR:本文提出融合偏好传播的多任务推荐模型。

3.4 实验结果及分析

3.4.1 CTR 预测

各模型在3个数据集上的CTR预测结果如表3所示。可以看出,本文提出的模型相比其他基线模型在CTR预测上AUC和ACC均有所提高。在MovieLens-1M数据集中与MKR和RippleNet相比,AUC和ACC分别提高了0.68%、0.31%和0.77%、0.54%。与KGCN、KGNN-LS、LibFM相比,在AUC指标上分别提高了1.75%、0.80%、2.77%,在ACC指标上提高了1.05%、0.53%、3.33%。在Book-Crossing和

Last.FM两个数据集中,由于用户—项目交互数据相对稀疏,预测结果相对于MovieLens-1M并没有特别明显提升,但是相对于基线模型也有很好的效果。这是因为在偏好传播过程中,传播效果随着波纹集的增加逐渐变弱且容易引入噪声。上述结果证明了融合后的模型是有效的,通过偏好传播与交叉压缩单元的交互作用,模型能够更好地挖掘知识图谱的潜在特征。

Table 3 The results of AUC and ACC in CTR prediction

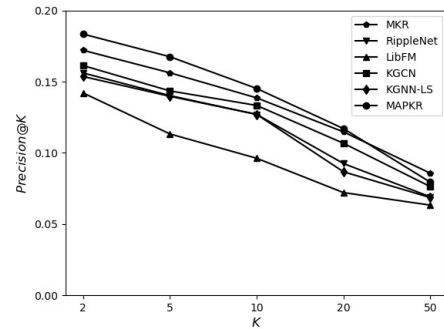
表3 CTR预测中AUC和ACC的结果

Model	MovieLens-1M		Book-Crossing		Last.FM	
	AUC	ACC	AUC	ACC	AUC	ACC
MKR	0.913 9	0.842 6	0.717 2	0.667 2	0.765 7	0.702 9
RippleNet	0.913 0	0.840 3	0.685 4	0.640 2	0.737 6	0.703 7
KGCN	0.903 2	0.835 2	0.690 1	0.654 4	0.782 0	0.713 1
KGNN-LS	0.912 7	0.840 4	0.689 9	0.670 5	0.795 3	0.726 0
LibFM	0.892 4	0.812 4	0.682 1	0.673 1	0.777 3	0.709 0
MAPKR	0.920 7	0.845 7	0.672 0	0.712 3	0.800 1	0.729 9

3.4.2 Top-K 推荐

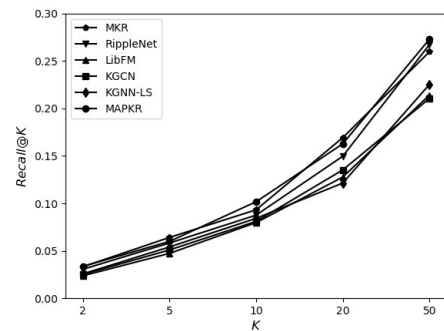
针对K取不同值时,各模型在3个数据集上Top-K推荐的Precision和Recall结果如图4—图6所示。

图4(a)、图5(a)、图6(a)3个图分别描述了各模型在3个数据集上Precision随K值的变化情况,可以看出,随着K



(a) Precision@k result

(a) Precision@k 结果

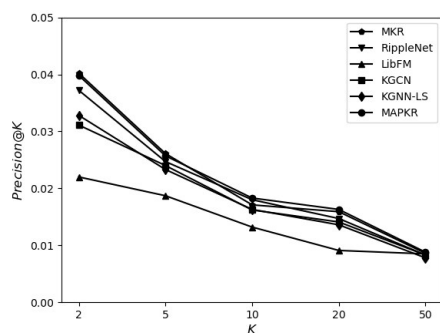


(b) Recall@k result

(b) Recall@k 结果

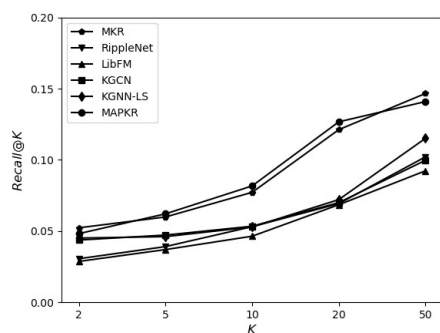
Fig. 4 The results of Precision@k and Recall@k on MovieLens-1M in Top-K recommendation

图4 Top-K推荐中Precision@k和Recall@k在MovieLens-1M结果



(a) Precision@k result

(a) Precision@k 结果



(b) Recall@k result

(b) Recall@k 结果

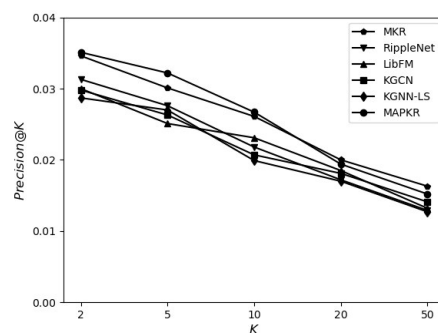
图 5 The results of Precision@k and Recall@k on Book-Crossing in Top-K recommendation

图 5 Top-K 推荐中 Precision@k 和 Recall@k 在 Book-Crossing 结果值的增加, 各模型精确率逐渐下降。在 MovieLens-1M 数据集上 MAPKR 模型提升效果最为明显, 当 $K = 10$ 时, 较 MKR、RippleNet 两个模型在精确率上分别提升了 0.66%、0.50%, 较 KGCN、KGNN-LS、LibFM 分别提升了 1.19%、1.82%、4.90%。

图 4(b)、图 5(b)、图 6(b) 分别是各模型在 3 个数据集上 Recall 随 K 值的变化情况。图中可以看出, 随着 K 值的增加, 各模型的 Recall 值均呈上升趋势。在 MovieLens-1M 数据集上, 当 $K = 10$ 时, 较 MKR、RippleNet 两个模型召回率分别提升了 0.85%、4.25%。MAPKR 模型在 3 个数据集上的精确率和召回率均高于 MKR、RippleNet 两个模型, 这表明利用偏好传播获取到的用户向量替代原有的用户表示起到了良好的效果, 同时也证明了同时融合交叉压缩单元和偏好传播对特征提取的有效性。

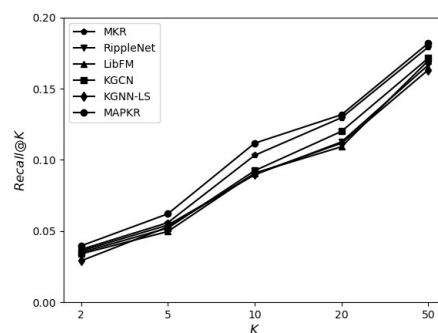
3.4.3 稀疏场景下模型性能验证

在模型中融合知识图谱作为辅助信息的一个主要目标是缓解推荐系统的数据稀疏和冷启动问题。因此, 本文为验证 KGE 模块在稀疏场景下的性能, 通过在 MovieLens-1M 数据集上固定验证集和测试集的大小, 只调节训练集的比率进行实验。MovieLens-1M 数据集上稀疏场景下模型性能的变化情况如图 7 所示。



(a) Precision@k result

(a) Precision@k 结果



(b) Recall@k result

(b) Recall@k 结果

图 6 The results of Precision@k and Recall@k on Last.FM in Top-K recommendation

图 6 Top-K 推荐中 Precision@k 和 Recall@k 在 Last.FM 的结果

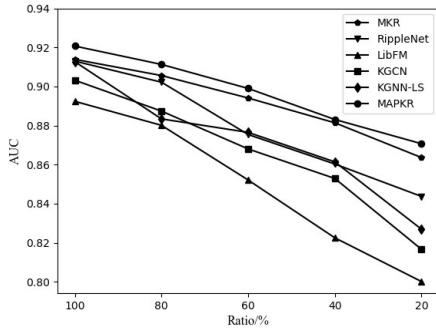
图 7 描述了在 MovieLens-1M 数据集上不同比率训练集下, 模型在 CTR 预测中 AUC 曲线和 AUC 下降比率曲线变化情况。可以看出, 当 $ratio = 20\%$ 时, AUC 曲线下下降幅度最为明显。其中, MKR、RippleNet、LibFM、KGCN 和 KGCN-LS 分别下降了 5.03%、6.93%、9.23%、8.66%、8.58%; MAPKR 模型仅下降了 5.00%, 这表明了 MAPKR 模型在用户一项目交互稀疏场景下的推荐性能仍优于其他基线模型。

3.4.4 超参数敏感性

为验证模型对训练次数 t 、跳数 n 以及波数集大小 s 的敏感性, 在 MovieLens-1M 数据集上进行对比试验。结果如图 8—图 10 所示。

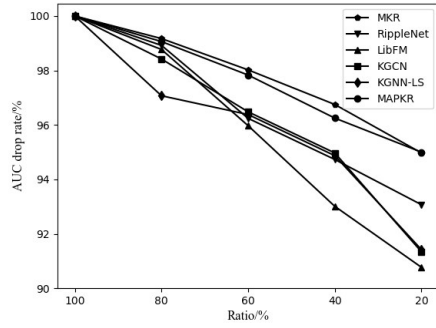
在训练 KGE 模块之前, 需要重复训练模型的推荐任务, 这一点在本文算法流程部分已经提出说明。图 8 描述了训练次数 t 的变化对模型性能的影响, 可以看出: 当 $t = 2$ 时, 模型性能达到最优; 当训练次数过大时, 模型性能逐渐下降。其原因在于, 当训练次数过多, 会误导 MAPKR 模型的目标函数, 也会导致训练时间太长从而影响训练效果; 当训练次数过小, 模型也无法充分利用知识图谱中的有效信息从而导致训练效果较差。

知识图谱中任意两个项目近邻结构的相似性可以帮



(a) AUC result

(a) AUC结果



(b) AUC drop rate

(b) AUC下降比率

Fig. 7 AUC variation of the model in sparse scenes on the MovieLens-1M dataset

图7 MovieLens-1M数据集上稀疏场景下模型的AUC变化情况

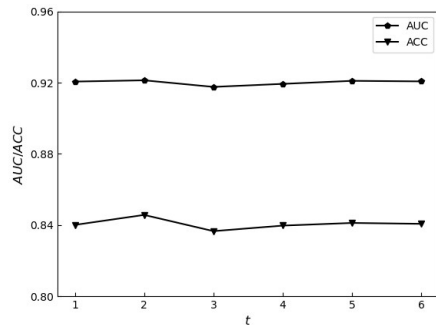


Fig. 8 Sensitivity results of the model to parameter t on the MovieLens-1M dataset

图8 在MovieLens-1M数据集上模型对参数 t 的敏感性结果

助测量其在推荐系统中的相关性,两个项目的近邻结构又随着跳数的增加变得更加相似,因此本文针对最大跳数对模型性能的影响进行实验。图9描述了MAPKR模型对最大跳数 n 的敏感性变化情况,可以看出:当 $n=2$ 时,模型性能最好;当 $n>2$ 时,模型性能开始下降。其原因在于,随着 n 不断增大,模型中的噪声也逐渐增多,从而更多的项目邻居实体对中心实体的特征表示产生影响,生成较长的关系路径。同时,太长的关系路径也不利于推断项目间

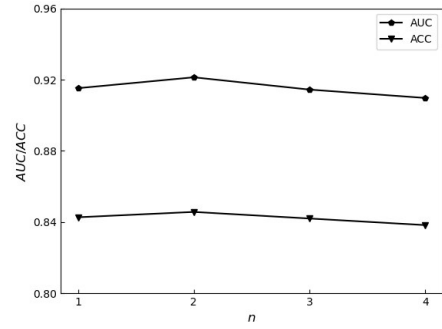


Fig. 9 Sensitivity results of the model to parameter n on the MovieLens-1M dataset

图9 在MovieLens-1M数据集上模型对参数 n 的敏感性结果

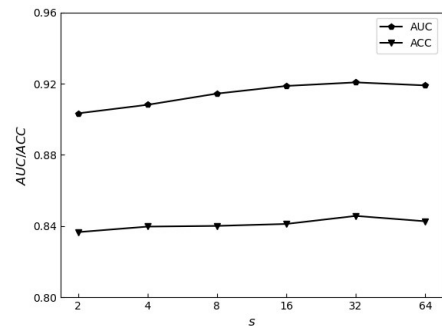


Fig. 10 Sensitivity results of the model to parameter s on the MovieLens-1M dataset

图10 在MovieLens-1M数据集上模型对参数 s 的敏感性结果

的相似性,从而导致模型性能下降。而 n 过小时,模型也无法充分利用项目的邻居实体信息,不利于构建项目的特征向量。

图10描述了涟漪传播过程中波纹集大小 s 对模型性能的影响,可以看出:当与其他参数相同时,随着波纹集 s 的增大,模型性能逐渐提高;当 s 取32时,模型性能达到最优;当波纹集过大时,模型性能逐渐下降。其原因在于,波纹集越大,知识图谱中编码的知识就越多,也越容易引入噪声;波纹集过小又会导致模型在知识图谱中可利用的有效信息变少,从而导致训练效果较差。

4 结语

大多数融合知识图谱的推荐模型采用传统的特征训练方法,不能够充分挖掘用户和实体间的潜在关系。传统的单任务学习在推荐任务中,逐渐暴露出容易过拟合和泛化能力差等缺点。针对上述问题,本文提出了一种融合偏好传播的多任务推荐模型,将交替训练和联合训练两种特征学习方式同时融入到模型中,在知识图谱上利用偏好传播挖掘实体及其邻域信息以丰富实体表示并采用多任务学习方式训练,提升模型泛化性。最后,通过在3个公开数据集上进行实验并与其他基线模型进行对比,证明

了本文提出的模型在推荐性能方面有所提升且优于其他基线模型。

本文提出的模型方法虽然在3个数据集上取得了一定优势,在数据稀疏场景下也有较好的性能,然而目前该模型采用的数据是固定时间的,且是静态的。在现实中,用户的喜好并不总是一直不变,可能随着时间或者社交关系的影响而发生偏移。因此,参考动态注意网络,如何增加模型的动态性将着重成为下一步研究方向。

参考文献:

- [1] HUANG L W, JIANG B T, LYU S Y, et al. A review of recommender systems based on deep learning[J]. Journal of Computers, 2018, 41(7): 1619-1647.
黄立威,江碧涛,吕守业,等. 基于深度学习的推荐系统研究综述[J]. 计算机学报, 2018, 41(7): 1619-1647.
- [2] JAMALI M, ESTER M. A matrix factorization technique with trust propagation for recommendation in social networks[C]//Proceedings of the fourth ACM Conference on Recommender Systems, 2010: 135-142.
- [3] WANG H, ZHANG F, HOU M, et al. Shine: signed heterogeneous information network embedding for sentiment link prediction[C]//Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018: 592-600.
- [4] WANG H, WANG N V, YEUNG D Y. Collaborative deep learning for recommender systems[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015: 1235-1244.
- [5] ZHANG F Z, JING N YUAN, 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, 2016: 353-362.
- [6] 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, 2016: 353-362.
- [7] LIN Y K, LIU Z Y, SUN M S, et al. Learning entity and relation embeddings for knowledge graph completion[C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence, 2015: 2181-2187.
- [8] ZHU X, ZHAO P, XU J, et al. Knowledge graph attention network enhanced sequential recommendation[C]//Proceedings of 4th International Joint Conference on Web and Big Data, 2020: 181-195.
- [9] WANG H, ZHAO M, XIE X, et al. Knowledge graph convolutional networks for recommender systems[C]//Proceedings of the World Wide Web Conference, 2019: 3307-3313.
- [10] ZHANG Y, LIU J W, ZUO X. Multitask learning[J]. Journal of Computer Science, 2020, 43(7): 1340-1378.
张钰,刘建伟,左信. 多任务学习[J]. 计算机学报, 2020, 43(7): 1340-1378.
- [11] RICCI F, ROKACH L, SHAPIRA B, et al. Recommender systems handbook[M]. New York: Springer US, 2011.
- [12] PAZZANI M J. Content-based recommendation systems[M]. Berlin: Springer, 2007.
- [13] RESNICK P, IACOVOU N, SUCHAK M, et al. Grouplens: an open architecture for collaborative filtering of netnews[C]//Proceedings of ACM Conference on Computer Supported Cooperative Work, 1994: 175-186.
- [14] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms[C]//Proceedings of the Tenth International World Wide Web Conference, 2001: 285-295.
- [15] KOREN Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model[C]//Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008: 426-434.
- [16] PANDYA S, SHAH J, JOSHI N, et al. A novel hybrid based recommendation system based on clustering and association mining[C]//2016 10th International Conference on Sensing Technology (ICST), 2016: 1-6.
- [17] HE X, LIAO L, ZHANG H, et al. Neural collaborative filtering[C]//Proceedings of International Conference on World Wide Web, 2017: 173-182.
- [18] GAI K, ZHU X, LI H, et al. Learning piece-wise linear models from large scale data for ad click prediction[DB/OL]. <https://arxiv.org/abs/1704.05194>, 2017.
- [19] NAUMOV M, MUDIGERE D, SHI H-J M, et al. Deep learning recommendation model for personalization and recommendation systems[DB/OL]. <https://arxiv.org/abs/1906.00091>, 2019.
- [20] CARUANA R. Multitask learning[J]. Machine Learning, 1997, 28(1): 41-75.
- [21] SHEN Z, CUI C R, DONG G X, et al. Joint prediction of image aesthetics and emotion based on deep multi-task learning[J]. Journal of Software, 2023, 34(5): 2494-2506.
申联,崔超然,董桂鑫,等. 基于深度多任务学习的图像美感与情感联合预测研究[J]. 软件学报, 2023, 34(5): 2494-2506.
- [22] BACHMANN R, MIZRAHI D, ATANOV A, et al. Multimae: multi-modal multi-task masked autoencoders[C]//Proceedings of 17th European Conference on Computer Vision, 2022: 348-367.
- [23] ZHANG L, XIE Y, LUAN X D, et al. Multi-source heterogeneous data fusion[C]//2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), 2018: 47-51.
- [24] CAO Y, WANG X, HE X, et al. Unifying knowledge graph learning and recommendation: towards a better understanding of user preferences[C]//Proceedings of The World Wide Web Conference, 2019: 151-161.
- [25] HU B, SHI C, ZHAO W X, et al. Leveraging meta-path based context for top-n recommendation with a neural co-attention model[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018: 1531-1540.
- [26] WANG X, HE X, CAO Y, et al. Kgat: knowledge graph attention network for recommendation[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019: 950-958.
- [27] WANG H, ZHANG F, XIE X, et al. DKN: deep knowledge-aware network for news recommendation[C]//Proceedings of the 2018 World Wide Web Conference, 2018: 1835-1844.
- [28] ESPOSITO M, DAMIANO E, MINUTOLO A, et al. Hybrid query expansion using lexical resources and word embeddings for sentence retrieval in question answering[J]. Information Sciences, 2020, 514: 88-105.
- [29] ZHANG Y, AI Q, CHEN X, et al. Learning over knowledge-base embeddings for recommendation[DB/OL]. <https://arxiv.org/abs/1803.06540>, 2018.
- [30] WANG Q, MAO Z, WANG B, et al. Knowledge graph embedding: a survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743.
- [31] SUN Y, HAN J. Meta-path-based search and mining in heterogeneous information networks[J]. Tsinghua Science and Technology, 2013, 18(4): 329-338.
- [32] ZHAO H, YAO Q, LI J, et al. Meta-graph based recommendation fusion over heterogeneous information networks[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017: 635-644.
- [33] YU X, REN X, SUN Y Z, et al. Personalized entity recommendation: a heterogeneous information network approach[C]//Proceedings of the 7th ACM International Conference on Web Search and Data Mining, 2014: 283-292.
- [34] WANG H W, ZHANG F Z, WANG J L, et al. Ripple net: propagating user preferences on the knowledge graph for recommender systems[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018: 417-426.
- [35] WANG H W, ZHANG F Z, ZHAO M, et al. Multi-task feature learning for knowledge graph enhanced recommendation[C]//Proceedings of The world Wide Web Conference, 2019: 2000-2010.
- [36] WANG H, ZHANG F, ZHANG M, et al. Knowledge-aware graph neural networks with label smoothness regularization for recommender systems[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019: 968-977.
- [37] RENDLE S. Factorization machines with libfm[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2012, 3(3): 1-22.

(责任编辑:孙娟)