

融合协同知识图谱高阶邻居特征的推荐模型^①



于嘉玮, 薛 涛

(西安工程大学 计算机科学学院, 西安 710048)
通信作者: 薛 涛, E-mail: 954761809@qq.com

摘 要: 在推荐时引入知识图谱中的实体及关系信息是有效缓解冷启动问题的方法. HAN 模型首次将基于注意力机制的图神经网络用于异构图, 但是并没有充分利用节点的高阶邻居信息. 为了解决该问题, 提出了一种融合协同知识图谱高阶邻居特征的推荐模型 CKG-HAN. 该模型用元路径来连接项目节点, 将协同知识图谱分成多个子图, 模型的节点注意力层用于聚合子图中每个节点的高阶邻居特征, 关系注意力层给不同元路径下的节点特征分配不同的权重, 最终得到充分融合语义信息的节点嵌入表示. 在 MovieLens-1M 数据集上进行了 Top-K 推荐, 结果表明本文提出的模型能够有效提高推荐结果的准确性.

关键词: 协同知识图谱; 图注意力网络; 推荐系统; 注意力机制; 元路径; 特征融合

引用格式: 于嘉玮, 薛涛. 融合协同知识图谱高阶邻居特征的推荐模型. 计算机系统应用, 2022, 31(6): 252-258. <http://www.c-s-a.org.cn/1003-3254/8489.html>

Recommendation Model Fused with High-order Neighbor Features of Collaborative Knowledge Graph

YU Jia-Wei, XUE Tao

(School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: Introducing the entity and relationship information in the knowledge graph during recommendation is an effective way to alleviate the problem of cold start. The HAN model introduces the attention mechanism-based graph neural networks into heterogeneous graphs for the first time. However, it does not make full use of the high-order neighbor information of nodes. To solve this problem, the study proposes a recommendation model CKG-HAN that integrates the high-order neighbor features of the collaborative knowledge graph. The model employs meta-paths to connect project nodes and divides the collaborative knowledge graph into multiple subgraphs. The high-order neighbor features of each node in the subgraph are aggregated in the node attention layer of the model, and different weights are assigned to node features on different meta-paths by the relation attention layer. Finally, a node embedding representation is obtained which fully integrates semantic information. The Top-K recommendation is performed on the MovieLens-1M data set, and the results show that the model proposed in this study can effectively improve the accuracy of the recommendation results.

Key words: collaborative knowledge graph; graph attention network; recommendation system; attention mechanism; meta-path; feature fusion

1 引言

随着互联网技术和相关产业的迅速发展, 接入互

联网中的设备和应用呈现指数上升的趋势. 推荐系统 (recommendation system) 作为一种有效过滤信息的手

^① 基金项目: 陕西省 2020 年技术创新引导专项基金 (2020CGXNG-012)

收稿时间: 2021-08-15; 修改时间: 2021-09-13; 采用时间: 2021-09-22; csa 在线出版时间: 2022-05-26

段,从被提出到现在,诞生了大量的理论、技术和应用.协同过滤(collaborative filtering)作为推荐系统的经典算法,基于用户-项目交互矩阵建模用户偏好,为相似的用户推荐相似的物品.因为其易于使用并且能有效捕捉用户偏好被广泛应用在实际项目中,然而,基于协同过滤的推荐存在数据稀疏和冷启动问题^[1],并且在用户量越来越大的情况下,其计算的复杂度也急剧增加.

近年来,将知识图谱(knowledge graph)中的实体及关系信息引入推荐系统中是有效缓解冷启动问题的方法^[2].目前,融合知识图谱语义信息的方法大致可分为两类:基于嵌入的方法和基于路径的方法.其中,基于嵌入的方法主要是以 TransE^[3], TransR^[4]为代表的距离翻译模型,其核心思想是将实体和关系映射到连续的向量空间中,同时可以保留知识图谱原有的结构.基于路径的方法以构造元路径^[5]为核心,有很好的推荐效果以及可解释性,缺点是当关系信息很多时需要手动构建元路径,工作量较大.以上方法从语义关联的角度出发,得到的节点嵌入表示更适合知识图谱的一些内部任务使用,如:知识图谱补全、三元组分类等,而不是用于推荐任务.为了使得到的语义信息更适合于推荐任务,一种可行的方法是结合用户-项目二部图与知识图谱,构建协同知识图谱(CKG)^[6].从协同知识图谱中挖掘语义信息,并将获取到的信息传递给推荐模型.

从知识图谱中提取信息可以从图结构的角度出发.近年来,图卷积神经网络被认为能够很好地建模图的结构属性和节点特征信息^[7],空域图卷积直接在图结构上使用卷积操作获取潜在的关系信息.文献[8]提出的GAT模型将图卷积操作定义为利用注意力机制(attention)对邻居节点进行有区别地聚合.但是GAT模型没有充分利用边上的信息,且只能用于同构图(homogeneous graph),在后续的研究中仍有较大的改进空间.文献[9]改进了GAT只能用于同构图的问题,首次将基于注意力机制的图神经网络用于异构图,但是没有很好地聚合高阶邻居信息.文献[10]基于共同邻居数目对知识图谱中的每个实体邻域进行排序采样,再沿着关系路径逐层融合邻居信息.文献[11]提出一种基于知识图谱驱动的端到端的图神经网络推荐模型,利用一个评分函数计算影响因子改变不同邻居实体的聚合权重.

本文在HAN模型的基础上,提出了一种融合协同知识图谱高阶邻居特征的推荐模型CKG-HAN(collaborative knowledge graph hierarchical attention network).

该模型首先将二部图和外部知识图谱融合成一个协同知识图谱,在协同知识图谱中根据手动构建的元路径将其分成多个同构子图,在每个子图中使用注意力机制聚合节点邻居信息,针对原HAN模型不能获取高阶信息的问题,加入了一个信息聚合层,该层能够聚合多跳邻居信息.最后融合不同子图的特征得到充分融合知识图谱语义信息的节点嵌入表示.在MovieLens-1M数据集上与原模型以及一些主流模型进行了实验对比,CKG-HAN在Top-K推荐任务中,准确率提升了1.1%,召回率提升了1.2%.

2 CKG-HAN模型

2.1 协同知识图谱

协同知识图谱 G 由用户-项目二部图 G_I 和外部知识图谱 G_2 融合而成.二部图表示为 $G_I = \{(u, interact, i) | u \in U, i \in I\}$, U 代表用户集合, I 代表项目集合, $interact$ 代表用户 u 与项目 i 有交互.知识图谱表示为 $G_2 = \{(h, r, t) | h, t \in E, r \in R\}$, E 代表实体集合, R 代表实体间关系的集合.

协同知识图谱的一个示例如图1所示.

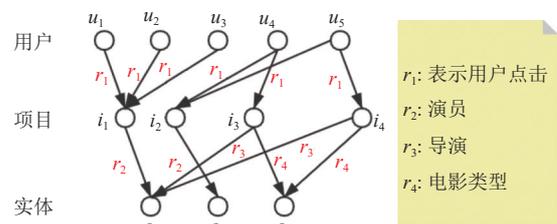


图1 协同知识图谱示例

从图1中可以看出,用户和项目之间的关系是单一的,项目和实体之间包含多种不同的关系.例如,项目节点代表电影,实体节点代表演员,图1中的演员 e_1 可以同时主演两部电影 i_1 和 i_3 ,此时 i_1 和 i_3 就可以通过元路径电影-演员-电影建立连接,所有通过电影-演员-电影这条元路径建立连接的项目构成一个子图.通过其他类型的元路径建立连接的项目构成另外的子图,我们需要综合同一项目在不同的元路径下的特征得到其最终特征.

2.2 模型概述

为了使推荐结果有更好的准确性和可解释性,我们不仅要用户对项目的交互进行建模,还要考虑项目之间的关系信息.知识图谱在提供额外属性信息的同时也蕴含大量的关系信息,CKG-HAN模型首先要将知识图谱与用户点击记录充分融合成协同知识图谱,

作为输入推荐模型的数据,之后利用改进的分层图注意力机制充分挖掘语义信息和高阶邻居信息,得到最终的特征表示后,计算特征间的相似度得到交互概率. CKG-HAN 模型的整体架构如图 2 所示.

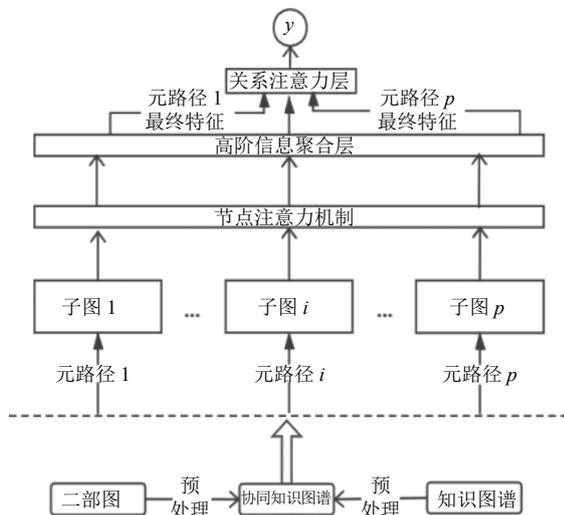


图 2 CKG-HAN 模型架构

从图中可以看出,模型的处理过程分为两个阶段:

在构建协同知识图谱阶段,需要预处理知识图谱,使知识图谱中包含全部的项目以及与之相关的实体作为补充.之后扩展二部图,加入一个代表关系的维度,和知识图谱中的三元组按行拼接在一起形成协同知识图谱.

在推荐阶段,推荐模块的处理分为节点注意力层和关系注意力层.对于节点注意力层,计算中心节点及其邻居节点之间的注意力系数,得到节点级注意力机制的特征表达.对于关系注意力层,聚合来自不同子图的特征.最后通过一个全连接层得到交互概率.

2.3 节点注意力层

节点注意力层要解决的问题是如何学习节点邻居的重要性并进行融合.由于知识图谱中的节点具有异构性,不同类型的节点具有不同的特征空间,所以需要 一个变换矩阵 M_{Φ_i} ,将不同类型节点的特征投影到同一特征空间中.变换过程如下所示:

$$h'_i = M_{\Phi_i} \cdot h_i \quad (1)$$

其中, h_i 表示原始特征; h'_i 表示投影特征.通过变换操作,节点注意力机制可以处理不同类型的节点.

在考虑邻居节点对目标节点的重要性时,注意力机制^[12]不仅能够加速模型训练的过程,而且能够很好地捕捉局部信息,很适合学习邻居节点对于目标节点的重要程度.对于节点对 (i, j) 和其对应的元路径 Φ 可以

构造如下表示:

$$e_{ij}^{\Phi} = LeakyReLU\{p_{\Phi}^T \cdot [h'_i \| h'_j]\} \quad (2)$$

其中, p_{Φ}^T 表示一个可训练的向量; $\|$ 表示拼接操作; e_{ij}^{Φ} 表示节点 j 对目标节点 i 的重要程度,是一个具体的数值;激活函数采用 $LeakyReLU$ ^[13].对上述注意力的绝对值进行归一化操作后的结果用 α_{ij}^{Φ} 表示.根据得到的相关性权重系数 α_{ij}^{Φ} ,对同一元路径下的节点邻居信息进行有区别地聚合.

此外,由于异构图具有无标度性,节点之间的连接分布不均,少数节点有很多与之相连的节点,多数节点则连接较少.为解决上述问题,可以重复若干次求注意力系数的操作,即多头注意力机制.使用多头注意力机制的好处是可以稳定模型学习的过程,获取更多关于邻居节点的信息,具体计算公式如下:

$$n_i^{\Phi} = \parallel_{k=1}^K \sigma \left(\sum_{j \in N_i^{\Phi}} \alpha_{ij}^{\Phi} \cdot h'_j \right) \quad (3)$$

其中, \parallel 表示拼接操作,是一种集成多个注意力头输出结果的方式.通过重复 K 次式 (3) 中的操作并拼接每次的结果,得到多头注意力机制的嵌入表示.

但是,上述过程只是聚合了节点周围一跳邻居的信息,无法捕获更深层的、潜在的关系.所以我们设置了一个聚合层用于聚合高阶信息,通过叠加多个聚合层,使消息传递能够从一跳扩展到多跳.聚合层采用一个聚合函数 $f_{relation}(h'_i, n_i^{\Phi}; \Phi)$ 实现,在阅读一些文献并经过实验验证后,筛选出函数 f 的 3 种实现形式以适用于不同的情况:

(1) GCN 聚合函数^[14]直接将节点投影特征与聚合邻居信息后的特征表达相加,并经过一个非线性变换得到结果:

$$f_{GCN} = LeakyReLU(W(h'_i + n_i^{\Phi})) \quad (4)$$

(2) GraphSage 聚合函数^[15]将两种特征进行拼接,并经过一个非线性变换:

$$f_{GraphSage} = LeakyReLU(W(h'_i \| n_i^{\Phi})) \quad (5)$$

(3) Bi-Interaction 聚合函数^[6]考虑两种特征之间的相互作用:

$$f_{Bi-Interaction} = LeakyReLU(W_1(h'_i + n_i^{\Phi}) + LeakyReLU(W_2(h'_i \circ n_i^{\Phi})) \quad (6)$$

其中, W_1 和 W_2 是线性变换矩阵; \circ 表示两个矩阵对应位置相乘的操作. 从后续的实验中可以得出结论, Bi-Interaction 聚合函数效果最好, 因为其丰富了特征交互. 最后我们经过多个聚合层, 每次聚合相当于向外扩展了一跳, 因为从第二层开始, 每个聚合层的输入是上一层计算后的输出, 具体的层数作为超参数.

元路径集合 $\{\Phi_1, \dots, \Phi_p\}$ 经过节点注意力层, 可以得到同一节点在不同元路径下的特征表达, 用 $\{n_{\Phi_1}, \dots, n_{\Phi_p}\}$ 表示, 将其输入关系注意力层进行进一步的处理. 节点注意力层的处理流程如图3所示.

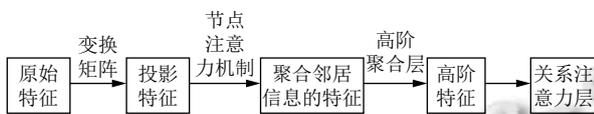


图3 节点注意力层处理流程

2.4 关系注意力层

在经过节点注意力层的处理后, 我们得到了同一节点在不同元路径下的特征表达, 每条元路径代表一种关系信息, 每个节点都可能包含在多种不同的元路径中. 关系注意力层主要解决的问题是如何聚合不同元路径下的关系信息.

与节点注意力层中的操作类似, 定义一个可学习的向量 q 与一个全连接层, 计算得到注意力分数, 计算方法如下:

$$w_{\Phi_p} = \frac{1}{|V|} \sum_{i \in V} q^T \cdot \tanh(W \cdot n_{\Phi_i} + b) \quad (7)$$

对注意力分数进行标准化处理, 得到不同元路径下邻居节点对目标节点的相关性权重值, 最后利用学习到的权重作为系数, 我们可以融合这些不同元路径下的节点信息, 以获得最终的节点嵌入表示.

模型选用的损失函数为 BPR 损失函数^[16], 使用 BPR 损失函数的目的是让正样本和负样本之间得分之差尽可能大, 为观测到的用户与项目间的交互分配更高的分数, 损失函数定义如下:

$$L = \sum_{(u,i,j) \in \Omega} -\ln \sigma(y(u,i) - y(u,j)) \quad (8)$$

其中, $\Omega = \{(u,i,j) | (u,i) \in I^+, (u,j) \in I^-\}$ 表示训练集, I^+ 表示正样本, I^- 表示负样本, $y(u,i)$ 表示用户 u 与项目 i 存在交互的概率, σ 表示非线性激活函数. 关系注意力层的处理流程如图4所示.

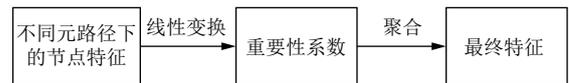


图4 关系注意力层处理流程

3 实验分析

3.1 实验环境与参数设置

本文使用的实验环境为 Intel i7-9750H 处理器, NVIDIA GTX1660Ti 显卡, 16 GB 运行内存. 软件环境为 Python 3.6, PyTorch 1.6.0.

本文模型在超参数的设置上经多次试验得出一组较好的取值, 迭代次数 200, 学习率 0.005, 多头注意力机制的头数 H 为 4, 聚合层的层数为 3, 使用 Adam 优化器进行学习率的自适应调节.

3.2 实验数据

本文使用的数据集是电影推荐场景 MovieLens-1M 数据集, 该数据集包含 6 000 个用户与近 4 000 部电影的 100 万条交互记录. 根据数据集中出现的电影, 搜集电影相关的属性信息来构建知识图谱, 数据集中已有关于电影的属性信息, 所以本文在 IMDB 上使用爬虫获取了与电影相关的导演和演员信息, 数据集的情况统计如表1所示.

表1 数据集概况

MovieLens-1M	知识图谱
用户: 6040	实体: 12215
电影: 3883	关系: 3
交互: 100209	三元组: 23362

其中, 知识图谱中的 3 种关系可以对应 3 种不同的元路径, 分别为电影-演员-电影 MAM, 电影-导演-电影 MDM, 电影-类型-电影 MTM. 对于每一位用户, 选取其有交互记录的电影的 80% 作为训练集, 20% 作为测试集, 并且在训练时加入一个随机的无交互记录的负例, 在测试时去除每个用户有交互记录的正样本, 对其余项目给出预测. 本文采用的评价指标包括准确率 Precision@K 和召回率 Recall@K, K 表示推荐列表的长度, 表示为用户推荐预测评分最高的 K 部电影, K 的取值为 [1, 5, 10, 20].

为了验证模型的有效性, 本文选取了几类主流模型进行对比:

(1) FM^[17]: FM 是一个结合支持向量机 (SVM) 和因式分解的模型, 使用因式分解的参数对变量之间的

关系进行建模,适用于数据稀疏的场景。

(2) NFM^[18]: 由于 FM 模型对于特征的组合仅限于二阶,缺少对特征之间深层次关系的抽取.因此,NFM 提出来就是在 FM 的基础上引入神经网络,实现对特征的深层次抽取。

(3) CKE^[19]: 采用包括异构网络嵌入和深度学习嵌入的两种嵌入方法自动从知识图谱中提取语义表示,通过执行知识图谱嵌入和协作联合过滤,CKE 可以同时从知识图谱中提取特征并且从中获取隐式的用户和项目之间的关系。

(4) HAN^[9]: 改进了 GAT 只能用于同构图的问题,首次将基于注意力机制的图神经网络用于异构图。

3.3 实验结果与分析

在 MovieLens-1M 数据集上,将 CKG-HAN 模型与上述几种推荐算法进行了对比实验.现将对比结果绘制成表格,统计各模型在 $K=20$ 的情况下的准确率与召回率,结果如表 2 所示。

表 2 推荐模型性能对比

模型	Precision@20	Recall@20
FM	0.259	0.209
NFM	0.260	0.210
CKE	0.264	0.220
HAN	0.274	0.232
CKG-HAN	0.286	0.252

在推荐长度为 20 时,本文提出的 CKG-HAN 模型与以上 4 个基准模型相比,在 2 个评价指标上均取得了最佳效果.与原 HAN 模型相比,在准确率上提高了 1.1%,在召回率上提高了 1.2%.CKG-HAN 模型的优势在于:(1) 引入协同知识图谱,将交互记录与知识图谱两个相互独立的内容联系在一起;(2) 新增高阶信息聚合层,将邻居节点的高阶信息有区别地聚合到中心节点的表示中,挖掘了基于元路径的语义信息。

图 5 与图 6 是模型之间的详细对比图.从图中可以看出,4 个基准模型中,FM 与 NFM 由于更多地关注特征交互带来的影响,没有引入额外的信息,效果并不是很好,由此可见引入知识图谱这种结构化数据的必要性.CKE 效果较好,影响其性能的原因可能是引入外部信息之后噪声增多,这也证明了需要对特征进行有区别地聚合的重要性.HAN 效果最佳,因为其加入了注意力机制,能够有区别地进行邻居聚合,并且 HAN 的走势平缓,说明多头注意力机制也能更好地稳定模型训练过程.实验结果证明,CKG-HAN 模型在基于协

同知识图谱的情况下,在公共数据集上的 Top-K 推荐任务中取得了更好的效果。

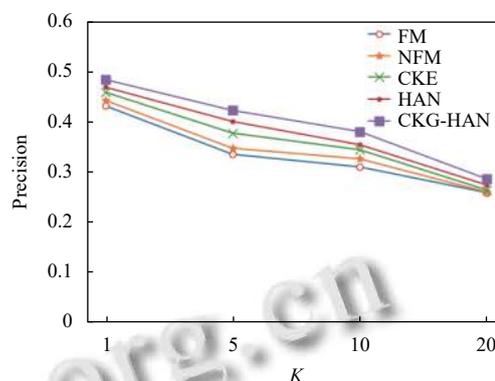


图 5 模型 Precision 性能对比

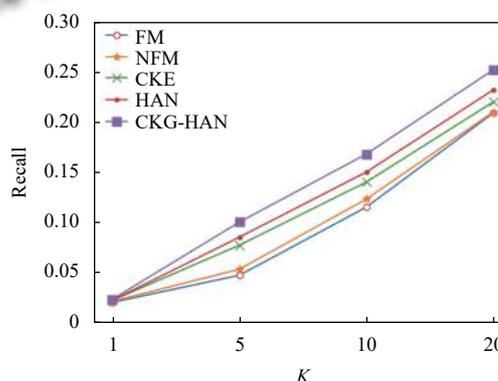


图 6 模型 Recall 性能对比

3.4 超参数敏感性分析

对于模型中的超参数设置,本文也进行了实验验证其取值的影响.在节点注意力层使用了多头注意力机制,图 7 显示了头数 H 在不同取值下对准确率的影响,可以看出 H 的取值为 4 时效果最好,当 H 为 6 时准确率反而降低,这与 GAT 模型中作者建议取值为 2 或 3 的结论相似。

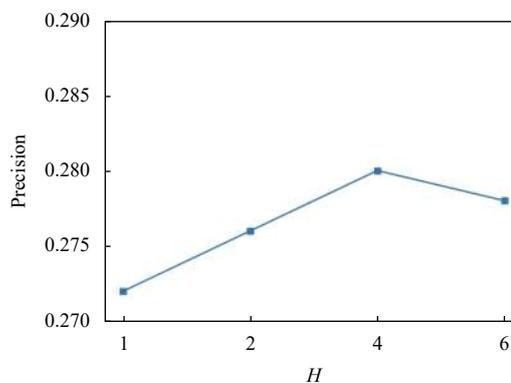


图 7 多头注意力机制的头数性能对比

3.5 聚合函数效果对比

在聚合高阶邻居时,本文使用了3种不同的聚合函数,图8中对比了这3种聚合函数以及不使用聚合函数的情况。

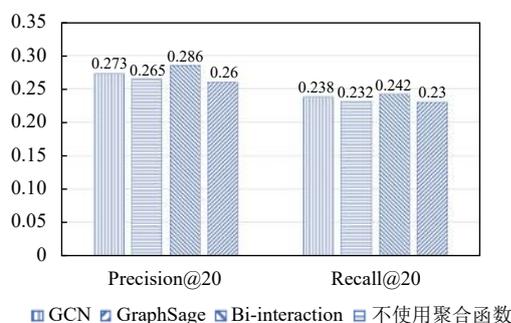


图8 聚合函数效果对比

从图中可以看出,不使用聚合函数的效果最差, f_{GCN} 聚合器在两个指标上都优于 $f_{GraphSage}$,一个可能的原因是GraphSage没有将中心节点特征与其邻居节点特征进行交互。 $f_{Bi-Interaction}$ 则增强了这一点,在聚合多跳特征的同时进行多种形式的特征交互,取得了最好的效果,因此它证明了在执行信息聚合和传播时特征交互的重要性。

4 结语

本文在HAN模型的基础上,主要针对其无法获取高阶信息的问题,新增了一个聚合层聚合多跳邻居信息,从而提出了CKG-HAN模型,该模型能有区别地聚合节点周围高阶邻居的信息,得到蕴含丰富语义的节点特征表示。然后结合MovieLens-1M数据集和外部知识图谱构建协同知识图谱作为输入数据,在Top-K推荐任务中CKG-HAN取得了比原模型和几类基准模型更好的效果,准确率与召回率分别提升了1.1%和1.2%,证明我们成功将模型用于推荐领域并且取得了效果上的提升。

在实验过程中我们发现,随着知识图谱关系类型的增加,需要手动构建大量元路径,生成很多子图。在下一步的工作中,将考虑不使用元路径,直接在同一张图中进行计算;当前用到的知识图谱仅考虑了一些简单的属性信息,引入更多信息会使计算更加复杂,如何引入更丰富的语义信息也是一个值得改进的方向。

参考文献

- Sun Z, Guo Q, Yang J, *et al.* Research commentary on recommendations with side information: A survey and research directions. *Electronic Commerce Research and Applications*, 2019, 37: 100879. [doi: 10.1016/j.elerap.2019.100879]
- 秦川, 祝恒书, 庄福振, 等. 基于知识图谱的推荐系统研究综述. *中国科学: 信息科学*, 2020, 50(7): 937-956.
- Bordes A, Usunier N, Garcia-Durán A, *et al.* Translating embeddings for modeling multi-relational data. *Proceedings of the 26th International Conference on Neural Information Processing Systems*. Lake Tahoe: ACM, 2013. 2787-2795.
- Lin YK, Liu ZY, Sun MS, *et al.* Learning entity and relation embeddings for knowledge graph completion. *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. Austin: AAAI, 2015. 2181-2187.
- Yu X, Ren X, Sun YZ, *et al.* Recommendation in heterogeneous information networks with implicit user feedback. *Proceedings of the 7th ACM Conference on Recommender Systems*. Hong Kong: ACM, 2013. 347-350.
- Wang X, He XN, Cao YX, *et al.* KGAT: Knowledge graph attention network for recommendation. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Anchorage: ACM, 2019. 950-958.
- 徐冰冰, 岑科廷, 黄俊杰, 等. 图卷积神经网络综述. *计算机学报*, 2020, 43(5): 755-780. [doi: 10.11897/SP.J.1016.2020.00755]
- Veličković P, Cucurull G, Casanova A, *et al.* Graph attention networks. *arXiv: 1710.10903*, 2017.
- Wang X, Ji HY, Shi C, *et al.* Heterogeneous graph attention network. *The World Wide Web Conference*. San Francisco: ACM, 2019. 2022-2032.
- 李世宝, 张益维, 刘建航, 等. 基于知识图谱共同邻居排序采样的推荐模型. *电子与信息学报*, 2021, 43(12): 3522-3529.
- 刘欢, 李晓戈, 胡立坤, 等. 基于知识图谱驱动的图神经网络推荐模型. *计算机应用*, 2021, 41(7): 1865-1870.
- Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: ACM, 2017. 6000-6010.
- Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. *Proceedings of the 30th International Conference on Machine Learning*. Atlanta: JMLR, 2013.

- 14 Yang ZL, Cohen WW, Salakhutdinov R. Revisiting semi-supervised learning with graph embeddings. Proceedings of the 33rd International Conference on Machine Learning. New York: JMLR, 2016. 40–48.
- 15 Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017. 1025–1035.
- 16 Rendle S, Freudenthaler C, Gantner Z, *et al.* BPR: Bayesian personalized ranking from implicit feedback. Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. Montreal: ACM, 2009. 452–461.
- 17 Rendle S. Factorization machines. 2010 IEEE International Conference on Data Mining. Sydney: IEEE, 2010. 995–1000.
- 18 He XN, Chua TS. Neural factorization machines for sparse predictive analytics. Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Shinjuku: ACM, 2017. 355–364.
- 19 Zhang FZ, Yuan NJ, Lian DF, *et al.* Collaborative knowledge base embedding for recommender systems. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016. 353–362.