

DOI: 10.11992/tis.202203039

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20230627.1703.002>

融合图卷积注意力机制的协同过滤推荐方法

朱金侠, 孟祥福, 邢长征, 张霄雁

(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: 图卷积神经网络 (graph convolutional neural network, GCN) 因其强大的建模能力引起了广泛关注, 在商品推荐中, 现有的图卷积协同过滤技术忽略了邻居节点在传播聚合过程中的重要性, 使得用户和商品的嵌入向量表达不够合理。为了解决这一问题, 本文提出一种融合图卷积注意力机制的协同过滤推荐模型。首先通过图嵌入技术将用户-项目的交互信息映射到低维稠密的向量空间; 其次通过堆叠多层的图卷积网络学习用户与项目间的高阶交互信息; 同时融合注意力机制为邻居节点自适应地分配权重, 不仅可以捕获更具代表性的邻居影响, 还使得在聚合邻居节点的特征信息时, 仅依赖于节点之间的特征表达, 使其独立于图结构, 提高了模型的泛化能力; 最后设计了分层聚合函数, 将图卷积层学习到的多个嵌入向量加权聚合, 使用内积函数得到用户-项目之间的关联分数。在 3 个真实数据上进行的泛化实验, 实验结果验证了该方法的有效性。

关键词: 图嵌入技术; 图卷积神经网络; 注意力机制; 协同过滤; 用户偏好; 协同过滤; 高阶交互; 邻域聚合

中图分类号: TP311 **文献标志码:** A **文章编号:** 1673-4785(2023)06-1295-10

中文引用格式: 朱金侠, 孟祥福, 邢长征, 等. 融合图卷积注意力机制的协同过滤推荐方法 [J]. 智能系统学报, 2023, 18(6): 1295-1304.

英文引用格式: ZHU Jinxia, MENG Xiangfu, XING Changzheng, et al. Collaborative filtering recommendation approach fused with graph convolutional attention mechanism[J]. CAAI transactions on intelligent systems, 2023, 18(6): 1295-1304.

Collaborative filtering recommendation approach fused with graph convolutional attention mechanism

ZHU Jinxia, MENG Xiangfu, XING Changzheng, ZHANG Xiaoyan

(School of Electronics and Information Engineering, Liaoning Technical University, Huludao 125105, China)

Abstract: The graph convolutional neural network (GCN) has attracted extensive attention due to its powerful modeling capabilities. In item recommendation, existing graph convolution collaborative filtering techniques ignore the importance of neighbor nodes in the propagation aggregation process, making the embedding vector representation of user and item unreasonable. Therefore, this paper proposes a collaborative filtering recommendation model fused with graph convolutional attention to address this problem. First, user-item interaction information was mapped to a low-dimensional, dense vector space using graph embedding techniques. Further, the high-order interaction information between the user and the item was learned using stacking multiple layers of GCN. The model also fused attention mechanisms to adaptively assign weights to neighbor nodes, thereby capturing the influence of highly representative neighbors. Simultaneously, the model could rely only on feature expressions between nodes when aggregating feature information from neighboring nodes, increasing the independence of the graph structure and improving the generalization capability of the model. Finally, a hierarchical aggregation function that aggregated multiple embedding vectors, which was learned from the graph convolution layer by weighting, was designed, and the inner product function was used to obtain the association score between the user and the item. Results of the extensive experiments conducted on three real datasets have demonstrated the effectiveness of the proposed approach.

Keywords: graph embedding technology; graph convolutional network; attention mechanism; collaborative filtering; user preference; collaborative filtering; high-order interaction; neighbor aggregation

收稿日期: 2022-03-22. 网络出版日期: 2023-06-28.

基金项目: 国家重点研发计划项目 (2018YFB1402901); 国家自然科学基金项目 (61772249); 辽宁省教育厅一般项目 (LJ2019QL017).

通信作者: 孟祥福. E-mail: marxi@126.com.

©《智能系统学报》编辑部版权所有

大数据时代下, 为了缓解信息过载问题, 推荐系统被广泛应用于个性化信息过滤。协同过滤 (collaborative filtering, CF) 是应用最广泛的推荐算法, 核心思想是从兴趣相似的用户中预测用户偏

好。早期的协同过滤推荐模型^[1-2]直接使用用户和项目ID作为嵌入向量,嵌入函数缺乏对用户与项目间交互信息的明确编码,导致所学习的嵌入表达具有局限性,不足以产生令人满意的CF嵌入。随后,SequentialMF模型^[3]将更多的邻居信息整合到嵌入过程,提高了嵌入质量,但模型受线性乘积的影响,无法有效地捕捉交互数据中的非线性结构。广义矩阵分解模型^[4]在浅层结构引入的神经网络赋予模型具有学习非线性交互的能力,使得它可以捕获用户与项目间的二阶交互关系,但却无法捕获包含更多信息的高阶交互,导致模型性能受限。

图卷积神经网络是一种针对图结构数据的深度学习模型^[5],具有更抽象的学习能力,其强大的表征能力,可以捕获包含许多丰富信息且更为抽象的高阶关系。通过多次迭代图卷积层还可以学习嵌入表征的全局结构,挖掘海量数据中深层次的属性特征,建模用户与项目之间的高阶交互的非线性关系。PinSage^[6]、NGCF^[7]和LightGCN^[8]等均为典型的图卷积协同过滤技术。但是,它们聚合邻居节点的方式依赖于图结构,局限了模型的泛化能力,并且为所有邻居统一分配固定的权重,不能区分邻居信息的重要性。另外,数据稀疏性对模型的推荐效果也会产生一定的消极影响。

为了解决上述提到的问题,本文提出一种融合图卷积注意力机制的协同过滤推荐模型,利用图嵌入技术将初始的用户和项目信息嵌入到低维稠密的向量空间,缓解了数据稀疏性对模型性能的消极影响,通过构建用户-项目交互图使得嵌入函数可以显式地学习用户与项目间的交互信息,以产生令人满意的CF嵌入;其次利用用户-项目交互图的拓扑结构来建模用户与项目间的高阶交互信息,通过堆叠多层图卷积网络以递归的方式学习此类高阶关系,以此揭示用户与项目之间的耦合关系;此外融合注意力机制聚合邻居节点信息,根据邻居节点的贡献值为其动态分配权重,捕获更具代表性的邻居影响,提高推荐结果的可解释性,而且这种聚合方式仅依赖于节点之间的特征表达,独立于图结构,可以提高模型的泛化能力;并设计了分层聚合函数,将图卷积层学习到的多个嵌入向量加权聚合,学习每一层嵌入向量对最终嵌入向量的影响;最后使用内积函数得到用户-项目之间的关联分数。

1 相关工作

基于协同过滤的推荐算法。矩阵分解^[9]是经

典的协同过滤方法,它将用户/项目表示在潜在向量的低维空间,通过用户/项目的潜在向量的内积来预测用户对项目的评分。随着深度学习的发展,基于深度学习的推荐模型可以更好地增强模型的表达能力。它将用户/项目信息经神经网络的表征学习,得到更具代表的用户/项目表达。文献[10]和文献[11]是经典的基于深度学习的协同过滤推荐模型。

基于图嵌入的推荐算法。图嵌入的基本思想是使用降维技术将节点的高维信息映射到低维稠密的向量空间,以此完成聚类^[12]和分类^[13]等任务。最初的嵌入方式使用随机游走^[14]来映射向量的低维嵌入,但是随机游走每一步生成所有项目的排名分数,导致了效率低下问题。随后研究者开始利用网络嵌入模型^[15-16]来缓解这一限制。文献[15]提出了异构偏好嵌入模型,将用户偏好和查询意图编码到低维向量空间中,通过相似度计算来生成推荐列表。Hop-Rec^[16]将矩阵分解和基于图的学习模型融合在一起,通过将观察到的用户-项目的直接交互因子化来获取用户偏好,从用户-项目交互图中随机游走来提取间接偏好。

基于图卷积神经网络的推荐算法。图卷积网络已经在多种领域取得了成功,最近的研究趋势是将卷积推广到不满足平移不变性的图结构数据中,使用卷积算子学习图结构数据。GraphSage^[17]是一个通用的图卷积推荐框架,利用一个广义聚合函数从节点的局部邻域采样并聚集特征来生成嵌入向量,使得图卷积网络在推荐领域取得了进一步成功。文献[18]通过将用户-项目的交互数据转换为3个异构图来学习异构图之间的共享特征表示,利用图卷积技术来聚集邻居的特征,保持了图数据的局部性质。

基于注意力机制的推荐算法。图注意力网络(graph attention network, GAT)^[19]不仅可以过滤输入数据中的噪声数据,还可以加强某些代表性数据的权重,因此在推荐系统中应用注意力机制逐渐成为推荐新技术。文献[20]提出一种基于自注意力机制的属性异构信息网络嵌入的商品推荐模型,模型通过属性异构信息网络嵌入学习用户与项目的低维嵌入表示,利用自注意力机制考虑不同的属性信息对推荐结果有不同程度的影响,克服了传统属性异构信息网络嵌入推荐模型的局限性。文献[21]是一种基于双层注意力机制的推荐模型,通过引入局部注意力和相互注意力学习用户与评论之间所存在的潜在相关性,提高了推荐结果的可解释性。

2 问题定义与解决方案

2.1 问题定义

定义1 用户集与项目集: 用 $U=\{u_1, u_2, \dots, u_M\}$ 表示用户集, $I=\{i_1, i_2, \dots, i_N\}$ 表示项目集, 其中, M 为用户数量, N 为项目数量。

定义2 用户-项目交互图: 令 $G_{ui}=\{U \cup I, E_{ui}\}$ 表示用户-项目交互图, 其中 U 和 I 为顶点集合, $|U|$ 和 $|I|$ 分别表示用户与项目的数量, E_{ui} 为用户与项目之间边的集合。

2.2 解决方案

本文的总体解决方案的流程图如图1。

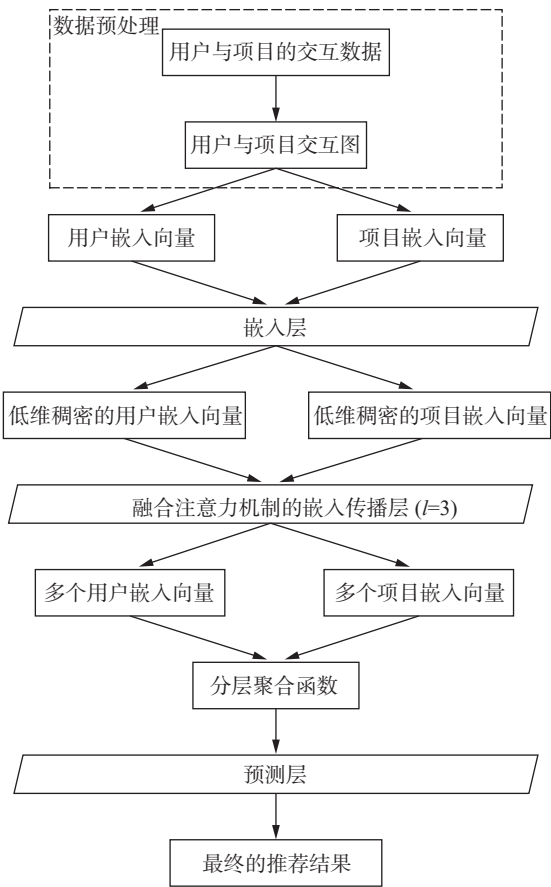


图1 模型流程
Fig.1 Model flow

模型流程共分3步。

1) 构建模型的嵌入层。首先将用户与项目的原始交互数据, 以 one-hot 编码的方式将其转换为数据; 其次根据用户与项目间的交互记录构造用户-项目交互图; 最后利用图嵌入技术将用户与项目信息嵌入到低维稠密的向量空间。

2) 构建融合注意力机制的嵌入传播层。将嵌入层得到的低维稠密向量送入嵌入传播层, 通过堆叠多层图卷积网络在嵌入空间构建信息交互, 强化用户与项目间的传播嵌入; 融合注意力

机制根据邻居节点的贡献值为其动态的分配权重。

3) 构建预测层。经嵌入传播层学习将得到多个用户和项目的嵌入向量, 将其送入分层聚合函数得到最终的嵌入向量表达; 最后采用内积交互函数, 得到最终的推荐结果。

本文设计的融合图卷积注意力机制的协同过滤模型, 框架如图2, 主要由3个模块组成: 1) 嵌入层、初始化用户和项目的嵌入向量; 2) 融合注意力机制的嵌入传播层, 通过图卷积技术细化用户与项目间的高阶交互信息, 融合注意力机制学习邻居节点在传播聚合过程中的重要性; 3) 预测层, 采用内积交互函数, 得到用户与项目之间的关联分数。

2.2.1 嵌入层

模型在嵌入层通过在用户-项目交互图上传播嵌入来强化嵌入向量的潜在特征。根据嵌入模型^[22-23], 一个用户 u (或项目 i) 的嵌入向量可表示为 $\mathbf{c}_u \in \mathbf{R}^d$ ($\mathbf{c}_i \in \mathbf{R}^d$), 其中 d 表示嵌入维度, 那么由用户和项目嵌入向量组成的参数矩阵作为初始嵌入向量查找表, 如下式:

$$\mathbf{E}_m = [\mathbf{e}_{u_1} \mathbf{e}_{u_2} \dots \mathbf{e}_{u_M} \mathbf{e}_{i_1} \mathbf{e}_{i_2} \dots \mathbf{e}_{i_N}] \quad (1)$$

式中: M 表示用户数量, N 为项目数量。嵌入查找表作为用户嵌入和项目嵌入的初始状态, 以端到端的方式进行传递和优化。

2.2.2 融合注意力机制的嵌入传播层

本文设计了融合注意力机制的嵌入传播层, 通过学习用户-项目交互图的拓扑结构来构建用户与项目间的高阶交互信息, 图3给出了用户-项目交互图。

图3(a)为用户-项目交互图, 图3(b)为目标用户 u_1 的高阶交互关系图(即交互图的拓扑结构)。 u_1 的高阶交互是指从路径长度 $L>1$ 的任何节点到达目标节点 u_1 的路径, 此类高阶交互承载着 u_1 更为丰富的交互信号。例如, 路径 $u_1 \leftarrow i_3 \leftarrow u_3$ 和 $u_1 \leftarrow i_3 \leftarrow i_6 \leftarrow u_4$ 表示 u_1 与 u_3 和 u_4 之间具有行为相似性, 但 u_1 与 u_4 之间有两次交互, 由此可以推断 u_1 与 u_4 具有更高的亲密度, u_4 更能影响 u_1 的决策, 也就是说 u_4 应该被分配更高的权重。然而, 以往的模型^[7-8] 为邻居节点分配统一的静态权重, 无法区分用户亲密度对用户决策的影响。例如, $i_4(u_1 \leftarrow i_2 \leftarrow u_2 \leftarrow i_4$ 和 $u_1 \leftarrow i_3 \leftarrow u_3 \leftarrow i_4)$ 与 $i_7(u_1 \leftarrow i_3 \leftarrow u_4 \leftarrow i_7$ 和 $u_1 \leftarrow i_6 \leftarrow u_4 \leftarrow i_7)$ 均有两条路径可以到达 u_1 , 传统模型无法区分 i_4 与 i_7 哪一个项目更能引起 u_1 的兴趣。本文模型融合注意力机制, 可以根据用户与邻居节点的亲密度, 自适应地为其分配权重。因此, 邻居节点 u_4 相较于 u_2 拥有更高的权重, 由此预测 i_7 更能引起 u_1 的兴趣。

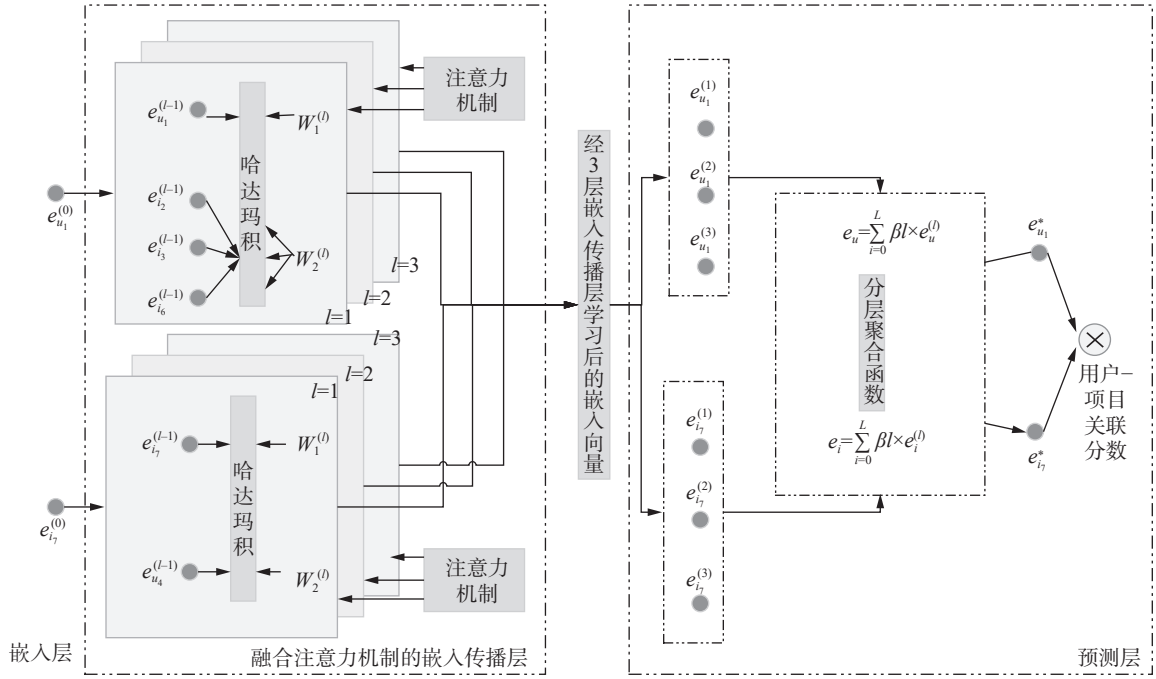


图 2 融合图卷积注意力机制的协同过滤推荐模型框架

Fig. 2 Framework of collaborative filtering recommendation model fused with graph convolutional attention

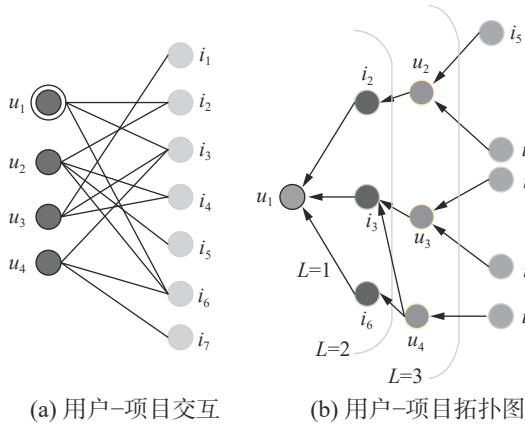


图 3 用户-项目交互图

Fig. 3 User-Item interaction diagram

此外,随着路径的增加每一层嵌入传播层在反映用户偏好时有不同的贡献,例如,路径 \$L=1\$ 为用户直接交互的项目,此类信息更能反映用户偏好,因此在聚合多个嵌入向量对最终嵌入向量的影响时,应该被赋予更高的比重。本文设计的分层聚合函数可以捕获不同层的嵌入表达,以实现路径越短对最终向量的影响越大。例如,由图 3 (b) 可知 \$i_4\$ 的两条路径在 \$L=3\$ 层分化,而 \$i_7\$ 的两条路径在 \$L=2\$ 层分化,则可以预测 \$i_7\$ 更能引起 \$u_1\$ 的兴趣。

模型通过堆叠多层图卷积层来构建嵌入传播层,为了更好地构建高阶传播,首先建模单阶传播。在现实世界中,用户交互的项目直接反映了用户偏好。由图 3 可知对于用户-项目交互图中

存在连接的用户-项目对 \$(u, i)\$, 可将从项目 \$i\$ 到用户 \$u\$ 的消息传递定义为

$$m_{u \leftarrow i} = f(e_i, e_u, p_{ui}) \quad (2)$$

式中: \$u \leftarrow i\$ 表示从项目到用户的方向传递, \$m\$ 表示传递消息的嵌入向量, \$f(\bullet)\$ 是一个消息编码函数, \$p_{ui}\$ 为相关系数控制 \$(u, i)\$ 边上每次传播时的衰减因子。

根据经典的图卷积推荐技术原理^[7-8], 相关性系数 \$p\$ 可定义为拉普拉斯范数:

$$p_{ui} = \frac{1}{\sqrt{|N_u| |N_i|}} \quad (3)$$

其中 \$N_u\$ 和 \$N_i\$ 分别表示用户 \$u\$ 交互的项目集以及项目 \$i\$ 交互的用户集。

本文的嵌入传播层为了更好的建模用户与项目间的交互信息, 利用哈达玛积 \$(e_i \odot e_u)\$ 将用户与项目的交互编码到消息传递机制中, 考虑了 \$e_i\$ 与 \$e_u\$ 之间的交互信息, 使得模型在传递消息时依赖于 \$e_i\$ 和 \$e_u\$ 之间的亲密度值, 也就是说相似的用户/项目之间可以传递更多的消息。此外, 融合注意力机制自适应地为邻居节点分配权重, 以此捕获邻居节点的重要性。由此消息编码函数 \$f(\bullet)\$ 可定义为

$$m_{u \leftarrow i} = p_{ui} \left(\sigma \left(\sum_{i \in N_u} \alpha_{ui} W_1 e_i + W_2 (e_i \odot e_u) \right) \right) \quad (4)$$

式中: \$W_1, W_2 \in \mathbf{R}^{d' \times d}\$ 为可训练的权重矩阵, 用于提取重要的嵌入信息; 矩阵维度为 \$d' \times d\$; \$d'\$ 为转换大小; \$\alpha_{ui}\$ 为注意力权重; \$\odot\$ 表示逐元素相乘。

依据 GAT^[19] 原理,在构建单阶传播时,邻居节点 i 对目标节点 u 的影响,可以由注意力相关分数表示:

$$e_{ui} = \text{LeakyRelu}(e_u || e_i) \quad (5)$$

式中: $||$ 表示向量的拼接, LeakyRelu ^[24] 为激活函数。为了得到对应的注意力权重,模型使用 softmax 函数计算目标用户 u 的所有邻居信息,对各个邻居节点的贡献值进行归一化,获得每个邻居节点 i 的权重,计算方式为

$$\alpha_{ui} = \text{softmax}(e_{ui}) = \frac{\exp(e_{ui})}{\sum_{k \in N_u} \exp(e_{uk})} \quad (6)$$

模型通过聚合目标用户 u 的邻居节点特征来更新用户 u 的嵌入表达 e_u , 采用 LeakyRelu 激活函数显式地利用 $m_{u \leftarrow i}$ 连接信息来关联用户和项目嵌 (LeakyRelu 函数允许对消息传递过程中的正负信号进行编码), 则聚合函数可以抽象为

$$e_u^{(1)} = \text{LeakyRelu}(m_{u \leftarrow i}) \quad (7)$$

其中 e_u 表示用户 u 在单层传播之后得到的用户嵌入向量。

本文模型通过叠加多个单阶传播层来挖掘高阶的交互信息,以此构建模型的高阶传播。由图 3(b) 可知,堆叠两层可捕获用户间的行为相似性(如 $u_1 \leftarrow i_3 / i_6 \leftarrow u_4$),堆叠 3 层可挖掘潜在的推荐优先级(如 i_7 的两条路径 $u_1 \leftarrow i_3 \leftarrow u_4 \leftarrow i_7$ 和 $u_1 \leftarrow i_6 \leftarrow u_4 \leftarrow i_7$)。模型利用堆叠的 l 层融合注意力机制的嵌入传播层,获得用户(或项目) l -阶邻居节点的特征信息,在第 l 层以递归的方式传播此类特征信息,计算方式为

$$e_u^{(l)} = \text{LeakyRelu} \left(\sum_{i \in N_u} m_{u \leftarrow i}^{(l)} \right) \quad (8)$$

其中,根据单阶嵌入传播层的构建,高阶邻居节点影响 $\sum_{i \in N_u} m_{u \leftarrow i}^{(l)}$ 的具体表达形式可定义为

$$\sum_{i \in N_u} m_{u \leftarrow i}^{(l)} = p_{ui} \left(\left(\sigma \left(\sum_{i \in N_u} \alpha_{ui} W_1 e_i \right) \right)^{(l-1)} + W_2 (e_i^{(l-1)} \odot e_u^{(l-1)}) \right) \quad (9)$$

图 4 给出了图 3(b) 中三阶路径 $u_1 \leftarrow i_3 / i_6 \leftarrow u_4 \leftarrow i_7$ 在嵌入传播时被捕获的过程。分析图 3(b) 可知, u_1 与 u_4 具有更高的亲密度,也就是说在采用注意力机制聚合邻居节点的贡献值时, u_4 被分配了更高的权重,由此预测 i_7 更能引起 u_1 的兴趣。因此,在所有到达 u_1 的三阶路径中,路径 $u_1 \leftarrow i_3 / i_6 \leftarrow u_4 \leftarrow i_7$ 被捕获。同样,当堆叠 l 层嵌入传播层时,用户与项目之间的高阶交互信息仍可以在嵌入传播的过程中被捕获。

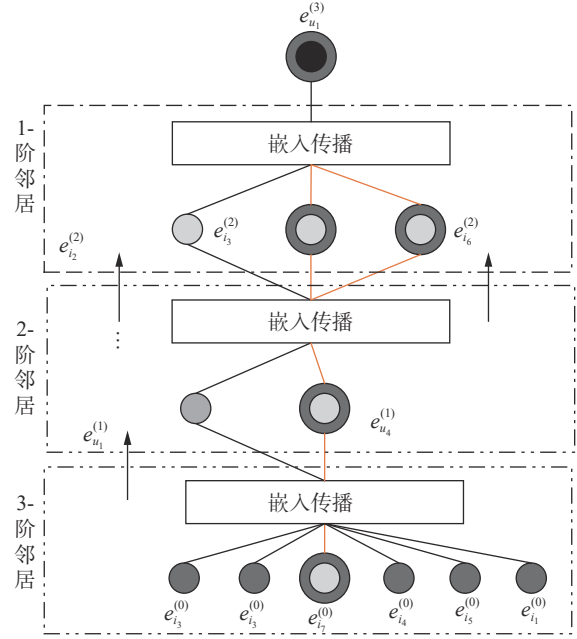


图 4 三阶嵌入传播

Fig. 4 Illustration of third-order embedding propagation

2.2.3 模型预测层

在经过 l 层传播后,得到了用户 u 的多个嵌入表达,即 $\{e_u^{(1)}, e_u^{(2)}, \dots, e_u^{(l)}\}$ 。由图 3 分析可知不同嵌入层获得的嵌入向量对用户偏好有不同的影响,因此本文设计了分层聚合函数进一步结合在每一层捕获的嵌入向量,构成用户最终的向量表达。对经过 l 层传播后得到的多个项目嵌入表达,同理得到最终的项目嵌入向量表达,分别如下所示:

$$e_u = \sum_{i=0}^L \beta_i \times e_u^{(i)} \quad (10)$$

$$e_i = \sum_{i=0}^L \beta_i \times e_i^{(i)} \quad (11)$$

式中: $\beta_i \geq 0$ 表示第 l 层嵌入对最终嵌入表达的重要性,在本文中设置 $\beta_l = 1/(1+L)$ 。

最后,在模型预测层,采用内积交互函数得到用户对目标项目的偏好,计算方式如下:

$$\hat{y}_{ui} = e_u^T \times e_i \quad (12)$$

2.2.4 模型优化

本文模型使用贝叶斯损失函数^[25]对模型进行优化,原因是贝叶斯损失函数考虑了观察到的交互和未观察到的交互之间的相对顺序。具体地说,它假设观察到的交互对用户偏好的影响更大,优化目标如下:

$$L = \sum_{u,i,j \in O} -\ln \sigma(\hat{y}_{ui} - \hat{y}_{uj}) + \lambda ||\Theta||^2 \quad (13)$$

其中, $O = \{(u, i, j) | (u, i) \in R^+, (u, j) \in R^-\}$ 被定义为用户与项目交互数据集,每一个数据都是一个三元

组。 \mathbf{R}^+ 表示观察到的用户-项目交互, \mathbf{R}^- 表示未观察到的用户-项目交互, $\sigma(\cdot)$ 是 sigmoid 函数。 $\lambda\|\Theta\|^2$ 为 L2 正则化项, λ 为正则化系数, 它控制 L2 正则化强度防止过拟合, $\Theta\{E, \{W_1^{(l)}, W_2^{(l)}\}_{l=1}^L\}$ 表示所有可训练的模型参数。

Dropout 策略可以在模型训练时有效防止模型的过拟合, 本文采用消息丢弃的 Dropout 策略。在消息丢弃方法中, 按照概率 p 随机舍弃式(9)计算出的消息, 使其不聚合到用户/项目的嵌入向量中。

3 实验及仿真结果

3.1 数据集

实验数据集采用 Gowalla、Yelp2018 以及 Amazon-Book 3 个公开的数据集, 对于每一个数据集, 随机选择每个用户的 80% 的交互历史构成训练集, 剩下的 20% 作为测试集, 随机选取训练集中的 10% 作为验证集, 用于参数的调试。表 1 给出了 3 个数据集的统计信息。

表 1 实验数据信息
Table 1 Statistics of the datasets

数据集	用户	项目	交互记录	稀疏度
Gowalla	29858	40981	1027370	0.00084
Yelp2018	31831	40841	1666869	0.00128
Amazon-Book	52643	91599	2984108	0.00062

3.2 对比算法

本文提出的图卷积注意力机制协同过滤 (GACF) 模型分别与 5 种推荐模型比较:

1) MF^[9] 模型: 该模型利用奇异值方法对矩阵进行分解, 通过贝叶斯个性化排名损失函数对模型进行优化, 并将用户-项目交互作为交互函数的目标值。

2) NeuMF^[11] 模型: 该模型是一个经典的基于深度学习的协同过滤推荐算法, 它结合了传统矩阵分解和多层感知机, 通过元素层面上的用户和项目嵌入和串联多个隐藏层来捕获用户-项目之间的非线性特征交互, 能够捕获用户和项目间的关联关系。

3) Hop-Rec^[10] 模型: 该模型基于图神经网络, 通过在图上的随机游走并结合图顶点的度, 以一定概率采样用户和项目的高阶关系, 可以显式地捕获用户和项目之间的高阶偏好, 丰富了训练数据, 提高了推荐结果的可解释性。

4) GC-MC^[26] 模型: 该模型基于消息传递的图自动编码, 通过编码器生成用户和项目表示, 并

利用用户和项目的一阶交互信息来捕捉它们之间的隐性特征, 提高了推荐模型的泛化能力。

5) LightGCN^[8]: 该模型基于文献 [7] 所提出的模型, 文献 [7] 模型中特征变换和非线性激活两个复杂的设计, 降低了模型的训练难度。

6) NGCF^[7] 模型: 该模型是基于图卷积神经网络的最新推荐模型, 它利用了用户与项目交互图的拓扑结构将交互信息嵌入到学习过程中, 并构建了 3 层图卷积神经网络对交互信息进行嵌入传播学习, 通过将传播层学习的嵌入向量进行聚合, 得到用户与项目之间的关联分数。

3.3 实验设置

本文模型和算法使用 Python 语言基于 TensorFlow 框架实现, 机器配置为 CPU i7-8700K 3.7 GHz, 操作系统为 Windows 10。在实验中, 基于以往模型^[7]的经验, 正则化系数设置为 1×10^{-4} , 每次处理的数据量大小为 1024, 嵌入大小为 64, 使用 64×64 的图卷积层, 迭代次数为 100 次时模型收敛。采用 Adam^[27] 作为优化器, 参数初始化采用 Xavier 方式。经验证集测试后, 学习率 l_r 设置为 0.001, Dropout 率 p 设置为 0.1。

3.4 评估指标

对于测试集中的每个用户, 将用户未交互的所有项目视为负样本, 将用户已经交互的项目作为正样本。本文中每种模型均输出用户对所有项目的偏好分数, 为了评估模型在 top- k 推荐效果方面的优越性, 实验采用召回率 ($R_{\text{Recall}@K}$) 和归一化折现积累收益 ($N_{\text{NDCG}@K}$) 作为推荐效果方面的评价指标。实验中, k 分别取值为 20、40、60、80、100, 模型总体推荐效果取其平均值。

召回率计算的是所有“被正确推荐的项目”占所有“应该被推荐的正确的项目”的比例, 计算公式为

$$R_{\text{Recall}@K} = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u |T(u)|} \quad (14)$$

式中: $R(u)$ 是为用户推荐的项目集合, $T(u)$ 为测试集上用户感兴趣的项目的集合。

$N_{\text{NDCG}@K}$ 是一种基于排名的测试指标, 排名靠前的项目得分更高。

$$N_{\text{NDCG}@K} = \frac{1}{n} \sum_{i=1}^K \frac{2^{R_{\text{rel}}(i)} - 1}{\log_2(i+1)} \quad (15)$$

式中: $R_{\text{rel}}(i)$ 表示用户对第 i 个项目的评分, $\log_2(i+1)$ 是一个位置递减权重。

3.5 模型推荐效果比较

3.5.1 总体比较

将本文模型与对比模型分别在3个不同的数据集进行实验,实验结果如表2所示。从模型推

荐效果的总体比较来看,本文提出的模型在3种数据集上的2种评价指标的表现均大幅度优于其他对比方法,说明了模型设计的合理性,同时也证明了模型的高效性和良好的泛化能力。

表2 总体比较

Table 2 Overall performance comparison

模型	Gowalla		Yelp2018		Amazon-Book	
	$R_{Recall@K}$	$N_{NDCG@K}$	$R_{Recall@K}$	$N_{NDCG@K}$	$R_{Recall@K}$	$N_{NDCG@K}$
MF	0.1291	0.1878	0.0317	0.0617	0.0250	0.0518
NeuMF	0.1326	0.1985	0.0331	0.0840	0.0253	0.0535
GC-MC	0.1395	0.1960	0.0365	0.0812	0.0288	0.0551
Hop-Rec	0.1399	0.2128	0.0388	0.0857	0.0309	0.0606
LightGCN	0.1830	0.1554	0.0649	0.0530	0.0411	0.0315
NGCF	0.1547	0.2237	0.0438	0.0926	0.0344	0.0630
GACF	0.2285	0.2597	0.1038	0.1455	0.0634	0.0936
GACF较NGCF的提升率/%	47.51	16.09	136.99	57.13	84.30	48.57

3.5.2 嵌入传播层数的影响

实验通过改变模型的深度,探究最佳的嵌入传播层数,在实验中,分别设置了{1,2,3,4}等不同的层数。观察表3可知,GACF模型相较于NGCF模型均取得了更好的推荐效果。此外,随着图卷积层的层数的增加,模型的推荐效果也有显著的提升。但是当层数增加到4层时,在Gowalla数据

集和Amazon-Book数据集上,模型出现了过拟合现象。原因是当嵌入传播层加深时,导致节点间的特征表达过度平滑化,使之缺少节点表征,进而缺少区分性,降低了模型的推荐效果,其次当堆叠过深的嵌入传播层,导致过多的噪声数据参与模型训练,使模型出现过拟合现象,也从侧面验证了堆叠三层的嵌入传播层足以捕获有效的CF信号。

表3 层数对模型推荐效果的影响

Table 3 Effect of number of layers on the model recommendation

层数	方法	Gowalla		Yelp2018		Amazon-Book	
		$R_{Recall@K}$	$N_{NDCG@K}$	$R_{Recall@K}$	$N_{NDCG@K}$	$R_{Recall@K}$	$N_{NDCG@K}$
1层	NGCF	0.1511	0.2218	0.0417	0.0889	0.0315	0.0618
	GACF	0.2197	0.2557	0.0975	0.1383	0.0579	0.0916
	提升率/%	+45.4	+15.28	+133.81	+55.57	+83.81	+48.22
2层	NGCF	0.1535	0.2238	0.0429	0.0926	0.0319	0.0622
	GACF	0.2246	0.2587	0.1005	0.1448	0.0588	0.0924
	提升率/%	+46.32	+15.59	+134.27	+56.37	+84.33	+48.55
3层	NGCF	0.1547	0.2237	0.0438	0.0926	0.0344	0.0630
	GACF	0.2285	0.2597	0.1038	0.1455	0.0634	0.0936
	提升率/%	+47.51	+16.09	+136.99	+57.13	+84.3	+48.57
4层	NGCF	0.1560	0.2240	0.0427	0.0907	0.0342	0.0636
	GACF	0.2294	0.2600	0.1010	0.1424	0.0630	0.0945
	提升率/%	+47.05	+16.07	+136.53	+57.00	+84.21	+48.58

3.5.3 模型top-k推荐有效性评估

为了探究融合注意力机制的有效性以及分层聚合函数的合理性,实验对模型的top-k推荐进行了 $R_{Recall@K}$ 和 $N_{NDCG@K}$ 两方面的评估。图5的实验结果表明,本文模型在3个数据集上所表

现出的推荐效果均远高于其他模型,表明本文方法所推荐的前top-k列表更具个性化。原因是模型在嵌入传播层融合注意力机制,衡量了邻居节点的重要性,自适应地为其动态分配权重,帮助目标用户选择更具代表性的项目。其次,

模型通过分层聚合函数,将图卷积层学习到的多个嵌入向量加权聚合,学习其对最终嵌入表

达的重要性,实现了路径越短对最终向量的影响越大。

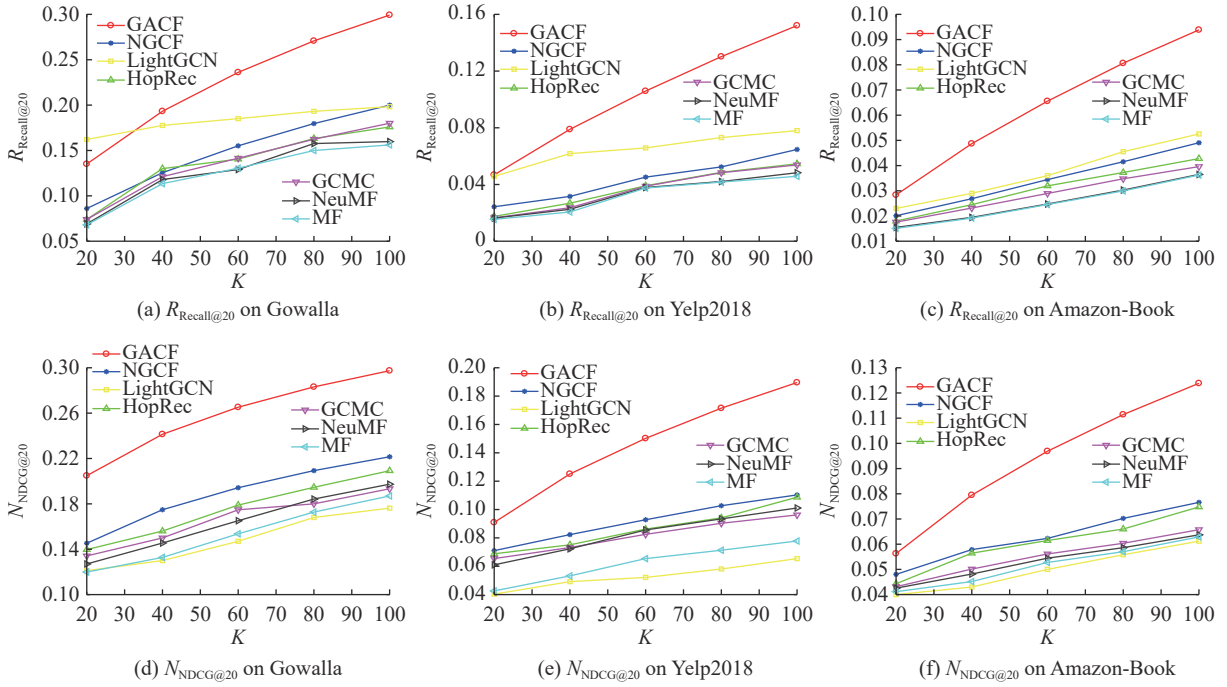


图 5 两个不同数据集上 top-k 推荐效果

Fig. 5 top-k recommendation performance on two different data sets

3.6 超参数学习

3.6.1 学习率的影响

模型的学习率直接影响着神经网络模型的收敛状态,进而影响着模型学习的性能表现,因此选择一个合适的学习率对于一个神经网络模型的影响至关重要。本文在实验中,将其设置为一个超参数,通过动态学习率策略调整训练出一个最优的学习率大小。

从图 6 的实验结果观察可知当学习率 $l_r=0.001$ 时,模型的总体效果达到最佳,当学习率 $l_r=0.01$ 时,模型的总体效果最差。在 Gowalla 数据集上,当 $l_r=0.01$ 时 $R_{Recall@20}$ 和 $N_{NDCG@20}$ 的结果分别为 0.1750 和 0.2057;当 $l_r=0.001$ 时 $R_{Recall@20}$ 和 $N_{NDCG@20}$ 的结果分别为 0.2269 和 0.2584,评价指标分别提高了 29.66% 和 25.62%。在 Yelp2018 数据集上,当 $l_r=0.01$ 时 $R_{Recall@20}$ 和 $N_{NDCG@20}$ 的结果分别为 0.0727 和 0.1140;当 $l_r=0.001$ 时 $R_{Recall@20}$ 和 $N_{NDCG@20}$ 的结果分别为 0.1028 和 0.1455,评价指标分别提高了 41.40% 和 27.63%。在 Amazon-Book 数据集上,当 $l_r=0.01$ 时 $R_{Recall@20}$ 和 $N_{NDCG@20}$ 的结果分别为 0.0382 和 0.0623;当 $l_r=0.001$ 时 $R_{Recall@20}$ 和 $N_{NDCG@20}$ 的结果分别为 0.0611 和 0.0924,评价指标分别提高了 59.95% 和 48.31%。由此可见,合适的学习率对于模型效果的表现有直接影响。

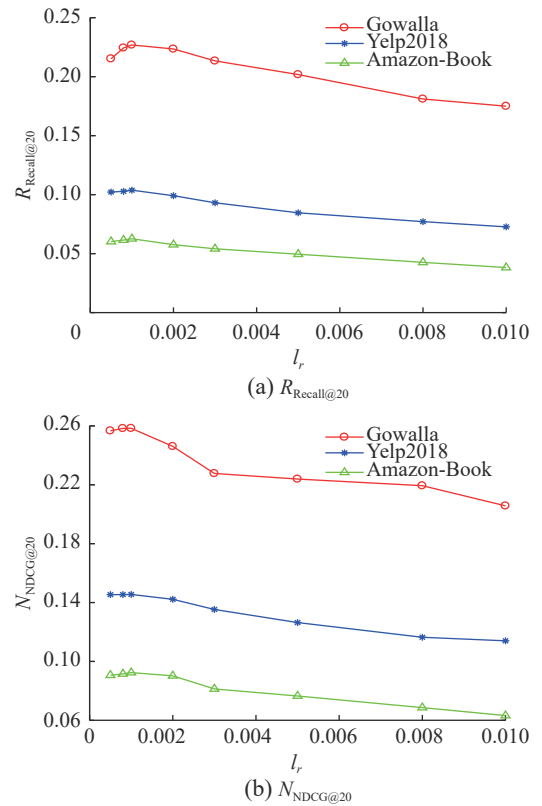


图 6 学习率测试结果

Fig. 6 Learning rate test results

3.6.2 Dropout 率学习

Dropout 策略可以在模型训练时有效防止模型的过拟合,本文采用消息丢弃 Dropout 策略。

图7给出节点丢弃率 p 作用于3个数据集上所产生的不同影响。在Gowalla、Yelp2018和Amazon-

Book数据集上,当 $p=0.1$ 时获得了最佳的召回率,当 $p=0.3$ 时获得了最差的召回率。

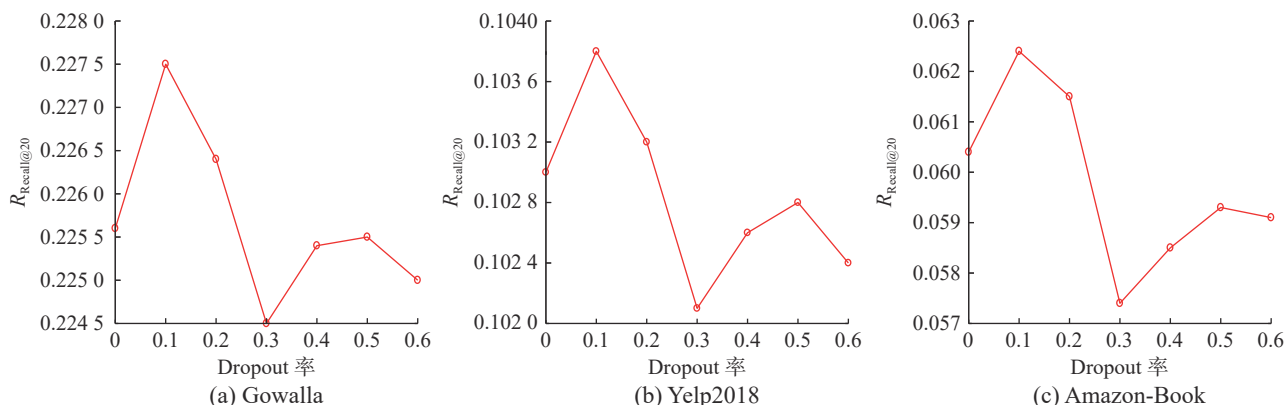


图7 节点丢弃率的影响

Fig. 7 Effect of message dropout ratios

4 结束语

本文提出了一种图卷积注意力机制协同过滤推荐模型,模型首先在嵌入层考虑用户与项目的交互信息,其中在嵌入传播层融合注意力机制,综合考虑了邻居节点对目标节点影响的重要性,通过分层聚合函数将嵌入层学习到的多个嵌入向量加权聚合,学习其对最终嵌入表达的重要性。最后,利用内积运算对用户与项目之间的关联分数进行预测。实验结果表明,与现有的主流协同过滤推荐模型相比,本文模型取得了更好的推荐效果。在未来的工作中,将考虑影响推荐性能的更多属性,如用户的社会关系、用户兴趣的动态变化等,尝试通过图卷积神经网络协同过滤技术缓解推荐系统中的冷启动问题,进一步提升推荐的准确性。

参考文献:

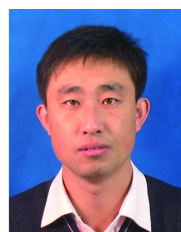
- [1] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms[C]//Proceedings of the 10th International Conference on World Wide Web. New York: ACM, 2001: 285–295.
- [2] AMBULGEKAR H P, PATHAK M K, KOKARE M B. A survey on collaborative filtering: tasks, approaches and applications[C]//Proceedings of International Ethical Hacking Conference 2018. Singapore: Springer, 2019: 289–300.
- [3] SUN Guangfu, WU Le, LIU Qi, et al. Recommendations based on collaborative filtering by exploiting sequential behaviors[J]. *Journal of software*, 2014, 24(11): 2721–2733.
- [4] HE Xiangnan, LIAO Lizi, ZHANG Hanwang, et al. Neural collaborative filtering[C]//Proceedings of the 26th International Conference on World Wide Web. New York: ACM, 2017: 173–182.
- [5] SHUMAN D I, NARANG S K, FROSSARD P, et al. The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains[C]//IEEE Signal Processing Magazine. New York: IEEE, 2013: 83–98.
- [6] YING R, HE Ruining, CHEN Kaifeng, et al. Graph convolutional neural networks for web-scale recommender systems[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 974–983.
- [7] WANG Xiang, HE Xiangnan, WANG Meng, 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, 2019: 165–174.
- [8] HE Xiangnan, DENG Kuan, WANG Xiang, et al. LightGCN: simplifying and powering graph convolution network for recommendation[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020: 639–648.
- [9] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems[J]. *Computer journal*, 2009, 42(8): 30–37.
- [10] 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. New York: ACM, 2008: 426–434.
- [11] HE Xiangnan, DU Xiaoyu, WANG Xiang, et al. Outer product-based neural collaborative filtering[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2018: 2227–2233.

- [12] PEROZZI B, AL-RFOU R, SKIENA S. DeepWalk: on-line learning of social representations[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2014: 701–710.
- [13] GROVER A, LESKOVEC J. Node2vec: scalable feature learning for networks[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 855–864.
- [14] GORI M, PUCCI A. ItemRank: a random-walk based scoring algorithm for recommender engines[C]//Proceedings of the 20th International Joint Conference on Artificial Intelligence. New York: ACM, 2007: 2766–2771.
- [15] CHEN C M, TSAI M F, LIN Y C, et al. Query-based music recommendations via preference embedding[C]//Proceedings of the 10th ACM Conference on Recommender Systems. New York: ACM, 2016: 79–82.
- [16] YANG J H, CHEN C M, WANG Chuanju, et al. HOP-rec: high-order proximity for implicit recommendation[C]//Proceedings of the 12th ACM Conference on Recommender Systems. New York: ACM, 2018: 140–144.
- [17] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 1025–1035.
- [18] XU Yanan, ZHU Yanmin, SHEN Yanyan, et al. Learning shared vertex representation in heterogeneous graphs with convolutional networks for recommendation[C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2019: 4620–4626.
- [19] PETAR V, GUILLEM C, ARANTXA C, et al. Graph attention networks[C]//Proceedings of International Conference on Learning Representations. Vancouver: ICLR, 2018, 23–35.
- [20] 王宏琳, 杨丹, 聂铁铮, 等. 自注意力机制的属性异构信息网络嵌入的商品推荐 [J]. 计算机研究与发展, 2022, 59(7): 1509–1521.
WANG Honglin, YANG Dan, NIE Tiezheng, et al. Attributed heterogeneous information network embedding with self-attention mechanism for product recommendation[J]. Journal of computer research and development, 2022, 59(7): 1509–1521.
- [21] LIU Donghua, LI Jing, DU Bo, et al. DAML: dual attention mutual learning between ratings and reviews for item recommendation[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2019: 344–352.
- [22] 高海燕, 毛林, 窦凯奇, 等. 基于图嵌入模型的协同过滤推荐算法 [J]. 数据采集与处理, 2020, 35(3): 483–493.
GAO Haiyan, MAO Lin, DOU Kaiqi, et al. Graph embedding model based collaborative filtering algorithm[J]. Journal of data acquisition and processing, 2020, 35(3): 483–493.
- [23] 邢长征, 赵宏宝, 张全贵, 等. 融合评论文本层级注意力和外积的推荐方法 [J]. 计算机科学与探索, 2020, 14(6): 947–957.
XING Changzheng, ZHAO Hongbao, ZHANG Quanguai, et al. Review text hierarchical attention and outer product for recommendation method[J]. Journal of frontiers of computer science and technology, 2020, 14(6): 947–957.
- [24] MAAS A L. Rectifier nonlinearities improve neural network acoustic models[C]//Proceedings of International Conference on Machine Learning. New York: ACM, 2013, 28–34.
- [25] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback[C]//Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence. New York: ACM, 2009: 452–461.
- [26] VAN DEN BERG R, KIPF T, WELLMING M. Graph convolutional matrix completion[C]//Proceedings of Computing Research Repository. New York: ACM, 2017, 1706–02263.
- [27] KINGMA D, BA J. Adam: a method for stochastic optimization[C]//Proceedings of the International Conference on Learning Representations. San Diego: ICLR, 2015, 1412–1427.

作者简介:



朱金侠, 硕士研究生, 主要研究方向为推荐系统。



孟祥福, 教授, 博士, 主要研究方向为 Web 数据库 top-k 查询、空间数据管理、推荐系统和大数据可视化。主持国家自然科学基金 2 项, 主持辽宁省各类基金项目 3 项。发表学术论文 60 余篇, 出版学术专著 1 部。



邢长征, 教授, 主要研究方向为分布式数据库与数据管理、流聚类、推荐系统。主持国家重点研发计划项目子课题任务 1 项, 省部级科研项目 3 项, 矿山安全应用类课题 10 余项。获辽宁省优秀教学成果一等奖 1 项。发表学术论文 40 余篇, 编写出版教材 3 部。