·专　　题·

融合跨平台用户偏好与异质信息网络的推荐算法研究

张　雪[1]　毕达天[1]　陈功坤[1]　杜小民[2*]
（1. 吉林大学商学与管理学院，吉林 长春 130012；2. 海南大学国际商学院，海南 海口 570228）

摘　要：［目的/意义］本文基于跨平台用户的异构大数据，提出一种融合跨平台用户偏好与异质信息网络的推荐算法(CPHAR)，对于缓解个性化推荐的稀疏性和冷启动问题具有重要意义。［方法/过程］首先，根据跨平台用户信息构建核心兴趣朋友圈，使用卷积神经网络和自注意力机制捕捉用户在源平台和目标平台中的信息偏好特征；其次，根据核心兴趣网络以及推荐项目之间的关系构建异质信息网络，使用异质图注意力网络模型进行特征聚合；最后，将以上特征嵌入改进后的矩阵分解模型，计算推荐得分。［结果/结论］模型在自主构建的 4 个跨平台数据集中均表现出优越的性能，本文不仅弥补了推荐领域中跨平台多属性和细粒度数据集的空缺，而且通过引入跨平台特征进一步完善了推荐系统相关的理论与方法体系。

关键词：推荐算法；跨平台；异质信息网络；用户偏好；深度学习
DOI:10.3969/j.issn.1008-0821.2024.09.003
〔中图分类号〕G252.0　〔文献标识码〕A　〔文章编号〕1008-0821（2024）09-0031-11

Research on Recommendation Algorithm Integrating Cross-Platform User Preferences and Heterogeneous Information Networks

Zhang Xue[1]　Bi Datian[1]　Chen Gongkun[1]　Du Xiaomin[2*]
（1. School of Business and Management, Jilin University, Changchun 130012, China;
2. International Business School, Hainan University, Haikou 570228, China）

Abstract: [Purpose/Significance] This paper proposes a recommendation algorithm that integrates cross-platform user preferences and heterogeneous information networks, based on the heterogeneous big data of cross-platform users. It plays a significant role in alleviating the sparsity and cold start problems of personalized recommendation. [Method/Process] Initially, the paper constructed a user core interest social circle based on cross-platform heterogeneous information, captured user information preference features in both the source and target platforms through convolutional neural networks and self-attention mechanisms. Subsequently, it built a heterogeneous information network based on the core interest network and the relationships among recommended items, and it employed a heterogeneous graph attention network model for feature aggregation. Finally, the study integrated the above feature embeddings into an improved matrix factorization model to compute recommendation scores. [Results/Conclusion] The model demonstrates superior performance across four independently constructed cross-platform datasets. This study not only fills the gap in cross-platform, multi-attribute, and fine-grained datasets in the field of recommendation but also enhances the theoretical and methodological system related to recommendation by introducing cross-platform features.

Key words: recommendation algorithm; cross-platform; heterogeneous information networks; user preferences; deep learning

收稿日期：2024-03-07
基金项目：国家社会科学基金项目“基于用户跨社交媒体的信息行为偏好特征挖掘与推荐研究”（项目编号：21BTQ059）。
作者简介：**张雪**（2000-），女，博士研究生，研究方向：深度学习与推荐算法。**毕达天**（1983-），男，教授，博士生导师，研究方向：大数据与管理信息系统。**陈功坤**（2002-），男，博士研究生，研究方向：大数据挖掘与自然语言处理。
通信作者：**杜小民**（1985-），女，副教授，博士，研究方向：大数据与信息管理。

随着社交网络用户规模的急剧扩张和数据资源的爆炸性增长，推荐系统被广泛地应用在各大社交网络平台，成为解决信息过载问题的有效途径。同时，用户不再局限于利用单个社交平台的信息，而是在不同社交平台间进行切换和转移以满足不同的服务需求[1]，形成相应的跨平台行为。用户跨平台数据的迁移共享为个性化推荐服务带来了崭新的机遇与挑战，跨平台推荐系统以同一用户作为连接源平台与目标平台的桥梁，使用用户在源平台中的信息丰富目标平台的数据，辅助模型在目标平台的精准推荐[2-3]。但是，跨平台多源信息间存在交叉关联、重复错节的关系，对用户模糊性和多样化的信息偏好进行准确识别和融合的难度较大[4-5]。面向跨平台异质环境的用户偏好融合与信息推荐研究仍然有大量的理论和关键技术亟待解决。

跨域推荐融合多个辅助领域的信息，通过知识迁移解决目标领域的数据稀疏问题，可以提供更加合理和个性化的推荐服务[3]。在跨域推荐的相关研究中，学者通常平行地在每个领域场景训练模型，或者通过联合协同过滤矩阵、共享参数或共享数据等方法训练一个多领域共享模型来实现信息的跨域流动[5]。前者忽略了用户、项目和内容层面的跨域关联，后者对于不同场景下大规模特征的共性和差异性解读与探索存在明显不足[6]。多数研究基于用户与推荐项目之间的历史交互数据来建模用户兴趣，对跨平台多源异构的辅助信息的利用尚不充分[7]，针对异质性、大规模和分布不均的跨平台用户数据缺少通用的特征提取和迁移融合方法[8]。

跨平台数据对于推荐系统具有重要意义，然而现有关于融合跨平台异构数据的信息推荐框架仍不够完善。鉴于此，本文将跨平台的多领域异质信息引入推荐系统，提出融合跨平台用户偏好与异质信息网络的推荐算法(CPHAR)，旨在全面挖掘跨平台数据要素价值，缓解由数据分布不均产生的稀疏性和用户冷启动问题。本文顺应情报学领域的研究发展趋势，强调多源异构信息的集成整合与融合统一[9]。研究成果将为应对推荐系统实际应用中面临的跨平台数据的复杂特点和解决跨平台信息推荐的瓶颈问题提供新的思路，为实现深度挖掘跨平台数据内的巨大价值提供新的解决途径，进一步提升推荐的效率和准确度。本文的主要贡献如下：

1）本文考虑到不同平台知识独立性和服务差异性的存在，在跨平台用户异质信息融合的基础上开展推荐研究，通过构建用户跨平台的核心兴趣朋友圈，结合卷积神经网络和注意力机制建模用户跨平台的信息偏好，实现了对目标平台冷启动用户进行特征增强的目的，为跨平台多源异构数据的融合和迁移提供了新的解决方案。

2）本文通过提出合理的关系剪枝和补全策略，使用异质图注意力网络(Heterogeneous Graph Attention Network，HAN)提升对异质节点特征的聚合能力。跨平台用户核心兴趣朋友圈有效地降低了网络的噪声与差异，从语义层面和用户行为的角度建立项目的隐式关联，为模型提供了更为全面和深入的推荐依据。

3）优化了矩阵分解模型。经典的矩阵分解模型仅使用用户和项目之间的交互信息来学习对应的潜在因子，对于冷启动用户和未知项目的特征提取能力较弱。本文利用神经网络模型将跨平台用户偏好和异质信息网络中的高阶特征纳入模型之中进行联合矩阵分解，增强模型的预测能力。

1 相关研究

经典的推荐模型包括基于内容的过滤、协同过滤和混合推荐[10]，通常依赖于用户与推荐项目丰富的历史交互进行推荐。大数据环境下稀疏的高维数据以及不断涌入系统的新用户和新项目使传统模型的局限性逐渐突出[11]。学者们通过引入文本、图像、标签、知识图谱等辅助信息，来解决推荐系统存在的上述问题[12-13]。李丹阳等[4]通过神经网络融合多源信息构建项目特征体系，结合加权矩阵分解的潜在因子向量预测用户对项目的偏好。丁浩等[14]使用漂移矩阵捕获用户兴趣随时间的动态变化，提出一种基于时序漂移的潜在因子分解模型。钱聪等[15]考虑到用户兴趣的遗忘，在丁浩等[14]的基础上结合用户多重偏好特征时间权重对模型进行改进。Yang M 等[16]提出，MMDIN 使用多模态模块提取图像特征，利用多头注意力机制从不同维度提取特征，增强了模型的交叉组合和预测能力。为提升推荐算法的时间效率和可扩展性，Das J 等[17]在 Voronoi 图的基础上提出了一种基于分区的推荐方法，在每个分区中单独执行协同过滤算法，将基准协同过滤算法的运行时间缩短了至少 65%，而且保证了较好的推荐质量。

跨域推荐将用户兴趣和项目特征在不同领域之

间进行融合，通过用户偏好的跨域转移解决单域推荐的数据稀疏和冷启动问题[18-19]。Zhang Q 等[20-21]认为，直接将源领域的评分模式转移到无重叠的目标领域可能会导致负迁移，采用领域自适应函数确保转移知识的一致性，并使用内核诱导的知识转移方式来对具有部分用户重叠的目标领域进行推荐。Zhao C 等[19]提出一种基于方面级转移网络的跨领域推荐框架，从评论文档中提取用户和项目抽象的方面级特征，利用重叠用户的方面特征来识别全局跨域方面相关性，以更细的粒度揭示跨领域用户的方面级联系。Xu Z 等[5]提出一种基于层次超图网络的相关偏好转移框架，包括动态项目转移和自适应用户聚合两个核心模块，模型将多域用户项目交互表示为一个统一的超图，利用超边来建立跨领域关系和获取相关知识。

异质信息网络在网络拓扑层面对系统中包含的异质辅助信息进行整合和利用，为推荐算法的进一步优化创造了新的可能性[7,22]。异质信息网络中不同类型的节点和链接代表了不同类型的对象和关系，集成了更为丰富的语义信息，可以通过挖掘高阶关系特征进行充分的语义关联和知识融合[11,23]。Shi C 等[24]将异质信息表示学习的特征向量嵌入矩阵分解模型，相较于传统矩阵分解模型，推荐性能得到有效提升。Li L 等[25]在异质网络中通过提取用户和项目相邻节点来补充元路径的缺失信息，根据卷积层和注意力机制得到的节点和元路径的嵌入进行推荐。熊回香等[26]对异质网络中的关系进行加权，通过对加权异质网络的表示学习进行学术信息的推荐研究。近年来，异质信息网络开始逐渐应用于跨域推荐。易明等[3]在源领域和目标领域分别建立异质信息网络，通过元路径、DeepWalk 算法获取网络中的特征信息，采用扩展的联合矩阵分解模型进行推荐预测。HCDIR 在源领域采用门控递归单元建模用户兴趣，在目标域构建异质信息网络，通过注意力机制和多层感知机学习跨域的特征映射[27]。

综上，推荐系统的研究取得了一定的进展，但仍存在一些不足。首先，跨域推荐对辅助域的信息挖掘不够充分，对于用户跨域多源异构数据的融合和交互缺乏深入研究，在用户偏好迁移的有效性和准确性方面还有较大的改进空间；其次，基于异质信息网络的推荐主要以浅层模型为基础，无法有效捕获大规模、复杂异质网络的语义信息；此外，异质信息网络中的高阶信息聚合方案大多是基于节点的神经网络模型，未能考虑到不同元路径的重要性及其对推荐结果的影响；最后，异质信息网络中普遍存在的噪声和差异问题也没有得到较好的解决，聚合与推荐无关的信息会干扰模型性能[28]。为弥补以上不足，本文一方面通过对用户跨平台产生的属性信息、兴趣知识、社交网络等异质信息进行融合和迁移利用，以全面识别用户的核心兴趣和建模用户偏好；另一方面，使用包含双重注意力的 HAN 聚合复杂的多类型特征和高阶交互信息，识别不同元路径下对推荐有用的异质信息，以共同提升模型的整体性能。

2 模型构建

本文提出的融合跨平台用户偏好与异质信息网络的推荐模型主要包括 3 部分内容：①基于跨平台异质信息融合的用户偏好特征建模：使用用户在不同平台中的属性、内容和社交关系数据构建用户跨平台的核心兴趣朋友圈，利用卷积神经网络模型捕捉用户在源平台和目标平台发布内容中所体现的信息偏好特征，通过注意力机制进行加权融合，得到跨平台迁移后的用户偏好特征；②基于 HAN 的高阶特征聚合：根据用户核心兴趣朋友圈以及用户和推荐项目相关的实体关系构建异质信息网络，使用 TransE 算法学习节点的初始嵌入向量，分别提取异质信息网络中用户和项目相关的元路径，使用 HAN 模型得到多跳路径下的高阶聚合特征；③基于改进矩阵分解模型的推荐预测：将跨平台用户偏好和实体的高阶特征纳入矩阵分解模型中，计算用户与项目之间的推荐概率得分，模型最终为每个用户生成对应的推荐列表。本文所提模型的框架结构如图 1 所示。

2.1 跨平台用户偏好特征建模

跨平台用户偏好特征建模部分通过对用户跨平台的异质信息进行处理，提取具有相同兴趣的跨平台核心兴趣朋友圈，以及获取完整的跨平台用户信息偏好特征。

Nie Y 等[29]提出，用户关注有相似兴趣的朋友，如果两个用户属于同一个体，那么他们在不同平台中将具有部分相似的核心兴趣，并且用户的核心兴趣在不同平台中将会同步改变。用户核心兴趣朋友圈的这种群组思想在社交媒体中的社群发现、用户身份识别、用户推荐和异常用户行为检测等方面得

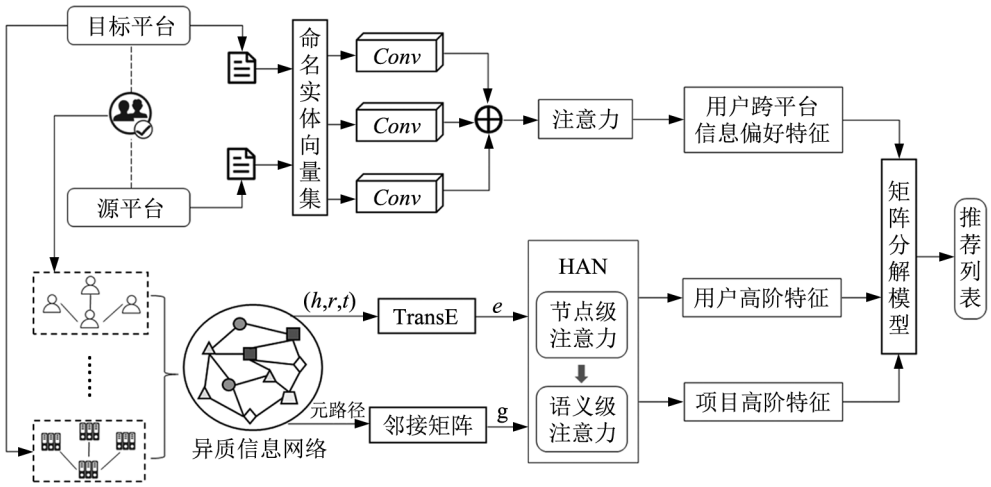

图 1 模型框架图

Fig. 1 Model Diagram

到广泛应用[30-31]。结合已有研究，本研究将同一用户所关注的具有相似跨平台信息和兴趣的朋友认定为该用户的跨平台核心兴趣朋友圈，综合考虑用户跨平台的属性信息、发布内容和社交网络关系构建用户跨平台的核心兴趣朋友圈。构建跨平台核心兴趣朋友圈的流程如下。

对于用户 i 及其关注的用户 j，分别计算 i 和 j 之间的内容相似度 $Csim_{ij}$ 和属性相似度 $Psim_{ij}$。$Csim_{ij}$ 的计算需要合并用户在源平台 s 和目标平台 t 中的发布内容 C，然后根据余弦相似度计算 i 和 j 合并内容的相似度。由于不同平台的用户属性类型差异性的存在，无法对不同平台中同一用户的属性信息进行简单合并，$Psim_{ij}$ 需要分别计算 i 和 j 在源平台 s 和目标平台 t 中属性的相似度，然后对不同平台中的属性相似度赋予一定的权重，W_P^s 和 W_P^t 分别表示源平台 s 和目标平台 t 中属性相似度的权重，$W_P^s+W_P^t=1$。具体的计算过程如式（1）~（3）所示：

$$Csim_{ij}=Sim((C_i^s+C_i^t),(C_j^s+C_j^t)) \tag{1}$$

$$Psim_{ij}=W_P^s\cdot Sim(P_i^s,P_j^s)+W_P^t\cdot Sim(P_i^t,P_j^t) \tag{2}$$

$$Sim(e_i,e_j)=cos(e_i,e_j)=\frac{\sum_{k=1}^{N}e_{ik}e_{jk}}{\sqrt{\sum_{k=1}^{N}e_{ik}^3\cdot\sum_{k=1}^{N}e_{jk}^2}} \tag{3}$$

用户 i 及其关注的用户 j 之间最终的核心兴趣相似度 $Usim_{ij}$ 为 i 和 j 之间的内容相似度 $Csim_{ij}$ 和属性相似度 $Psim_{ij}$ 的加权求和，W_c 和 W_P 分别为对应的权重，$W_c+W_P=1$。用户 i 的跨平台核心兴趣朋友圈 Cof_i 为 i 关注的用户列表 $friends(i)$ 中与 i 的核心兴趣相似度 $Usim_{ij}$ 排序较高的前 top_N 名朋友。具体的计算过程如式（4）和（5）所示：

$$Usim_{ij}=W_c\cdot Csim_{ij}+W_P\cdot Psim_{ij} \tag{4}$$

$$Cof_i=\{f\mid f\in arg\ sort_{desc}(Usim_{ij})[1:top_N],j\in friends(i)\} \tag{5}$$

为进一步探究跨平台的用户特征，需要对用户发布内容中的信息偏好进行有效挖掘。以重叠用户作为链接源平台和目标平台的桥梁，对用户在不同平台发布内容中的命名实体进行处理，使用卷积神经网络和注意力机制提取跨平台用户信息偏好。跨平台用户信息偏好的具体计算流程如下：

一维卷积神经网络常用于文本和序列数据分析，可以自动检测和提取局部特征。为解决静态命名实体向量带来的歧义问题，使用一维卷积神经网络捕获命名实体周围的上下文信息，可以获得更加准确的内容语义表达[32]。假设用户 i 在源平台 s 和目标平台 t 发布内容中的命名实体集分别为 E_i^s 和 E_i^t，E_i^s 和 E_i^t 中的命名实体均按发布内容的时间进行排序。使用多层的一维卷积网络对命名实体集进行处理，计算过程如式（6）和（7）所示：

$$z'_{i(k)}=ReLU(Conv1D(z_{i(k-1)})) \tag{6}$$

$$z_{i(k)}=Dropout(MaxPool(z'_{i(k)})) \tag{7}$$

其中，$z_{i(k)}$ 表示第 k 次卷积操作的输出向量，$z_{i(0)}$ 表示卷积模型的输入向量，$z_{i(0)}=E_i^s$ or E_i^t。第 k 层模型对前一层模型输出结果 $z_{i(k-1)}$ 进行一维卷积操作 $Conv1D$，使用激活函数 $ReLU$ 对结果进行非线性化，然后进行最大池化 $MaxPool$ 和正则化 $Dropout$ 操作，$MaxPool$ 和 $Dropout$ 可以增强模型对实体位置变化的鲁棒性。模型最后一层的输出结果 $z_i=(z_i^s$ or $z_i^t)$ 即用户 i 在不同平台中的偏好特征，使用注

意力机制对用户在不同平台的偏好进行加权融合，得到所有用户的跨平台信息偏好特征表示为 Z。

2.2 基于异质图注意力网络的高阶特征聚合

HAN 将注意力机制从同质图扩展到异质图，是一种包含节点级注意力和语义级注意力的层次注意力异质图神经网络[33]。HAN 使用节点级注意力学习基于元路径的邻居权重，然后联合学习不同元路径的权重，通过语义级注意力融合特定语义下的节点特征嵌入[34]。本节在数据集原有节点关系的基础上，通过整合用户核心兴趣朋友圈中的用户关系、进一步挖掘项目之间的潜在关联关系，构建异质信息网络，使用 HAN 聚合用户和项目特征。

2.2.1 异质信息网络构建与向量化

使用余弦相似度和点互信息(Pointwise Mutual Information，PMI)分别从语义层面和用户行为的角度挖掘项目之间的关联性。余弦相似度计算项目之间的语义相似性，有助于发现语义上相似的推荐项目，计算过程如式（3）所示。*PMI* 通过计算两个项目被同一用户交互与这两个项目各自独立出现的概率比值的对数，来衡量这两个项目之间的相关性。*PMI* 能够揭示出用户偏好背后的隐含关联，项目 I_1 和 I_2 的 *PMI* 的计算如式（8）所示。通过设置合适的阈值，建立具有相似语义和较高相关性的问题之间的链接。

$$PMI(I_1,I_2)=log_2\frac{p(I_1,I_2)}{p(I_1)p(I_2)} \tag{8}$$

然后，使用 TransE 对异质图中的节点和关系特征进行初始化。TransE 模型是一种基于距离的翻译嵌入模型，可以将多关系数据的实体和关系嵌入到低维向量空间[12]。该模型的基本前提是，如果三元组事实$<h,r,t>$成立，则尾部实体 t 的嵌入表示应接近头部实体 h 的嵌入表示和连接它们的关系 r 的总和，即 $h+r\approx t$[35]。TransE 只关注实体之间的直接关系，而不能考虑异质图中存在的复杂的多步骤关系路径，结合使用 HAN 模型一定程度上可以弥补 TransE 模型的不足。

2.2.2 异质图注意力网络模型

根据异质信息网络中用户和项目相关的元路径，分别生成不同元路径下的用户和项目的邻接矩阵，使用 HAN 的节点级注意力和语义级注意力对邻接矩阵进行处理。节点级注意力学习节点与基于元路径的相邻节点之间的特征，得到特定语义下的节点特征嵌入[33]。将节点特征映射为同一维度的特征 η，假设节点 j 是节点 i 在某一元路径 Φ 下的邻居节点集 N_i^{Φ} 中的一个节点，则 i 和 j 之间节点级别的注意力代表在某一元路径 Φ 下节点 j 对 i 的重要性，使用 *softmax* 函数进行归一化，得到节点级注意力权重 α_{ij}^{Φ}。计算过程如式（9）所示：

$$\alpha_{ij}^{\Phi}=softmax(att_{node}(\eta_i,\eta_j;\Phi)) \tag{9}$$

其中，att_{node}表示进行节点级注意力的图卷积神经网络，不同的元路径对应不同的 att_{node}。使用 α_{ij}^{Φ}对节点特征加权求和，使用激活函数 σ 进行非线性转化，得到节点 i 基于 Φ 的特征嵌入 h_i^{Φ}。为了使训练过程更加稳定，将节点级注意力扩展到多头注意力，重复进行节点级注意力训练 K 次，将学习到的嵌入拼接得到最终的 h_i^{Φ}。h_i^{Φ} 的计算过程如式（10）所示：

$$h_i^{\Phi}=\Big\|_{k=1}^{K}\sigma\Big(\sum\nolimits_{j\in N_i^{\Phi}}\alpha_{ij}^{\Phi}\cdot\eta_j\Big) \tag{10}$$

对于一个给定的元路径集$\{\Phi_1,\Phi_2,\cdots,\Phi_n\}$，在将节点特征输入节点级注意力后，可以得到 n 组特定元路径语义下的节点嵌入$\{H_{\Phi_1},H_{\Phi_2},\cdots,H_{\Phi_n}\}$，使用语义级注意力学习不同元路径的重要性，融合节点在不同元路径下的语义特征，可以学习更全面的节点嵌入[33]。元路径$\{\Phi_1,\Phi_2,\cdots,\Phi_n\}$的语义级注意力权重$(\beta_{\Phi_1},\beta_{\Phi_2},\cdots,\beta_{\Phi_n})$的表示如式（11）所示：

$$(\beta_{\Phi_1},\beta_{\Phi_2},\cdots,\beta_{\Phi_n})=att_{sem}(H_{\Phi_1},H_{\Phi_2},\cdots,H_{\Phi_n}) \tag{11}$$

其中，att_{sem}表示执行语义级注意力的深度神经网络，用于捕获异质图中的各个元路径的语义信息。

为计算不同元路径的重要性，首先将节点在某条元路径下的嵌入进行非线性转化，乘以使用一个可学习的语义级注意力 q，然后对同一元路径下所有节点的运算结果求平均值，得到特定元路径 $\Phi_p\in\{\Phi_1,\Phi_2,\cdots,\Phi_n\}$的重要性，使用 *softmax* 函数进行归一化处理，得到元路径 Φ_p 的权重表示为 β_{Φ_p}。具体计算过程如式（12）所示：

$$\beta_{\Phi_p}=softmax\left(\frac{1}{V}\sum\nolimits_{i\in V}q^T\cdot tanh(W\cdot h_i^{\Phi_p}+b)\right) \tag{12}$$

其中，V 表示 Φ_p 中的节点数量，W 为线性方程的权重矩阵，b 为偏置向量。β_{Φ_p}代表着元路径 Φ_p 的贡献率，β_{Φ_p}越高，意味着 Φ_p 的重要性越大。将学习到的权重作为系数，对特定元路径下的嵌入

进行融合，得到节点最终的特征表示。计算过程如式（13）所示：

$$H=\sum_{p=1}^{P}\beta_{\Phi_p}\cdot H_{\Phi_p} \tag{13}$$

2.3 推荐预测

通过引入额外的特征和非线性变换可以增强传统矩阵分解方法的表达能力和准确性[36-37]。本文对经典矩阵分解模型进行改进，通过增加跨平台用户偏好和 HAN 输出的高阶特征信息来拓展传统的矩阵分解模型。用户 u 与项目 i 之间交互概率的预测如式（14）所示：

$$\hat{r}_{u,i}=x_u^T\cdot y_i+(Z_u^T r_i^z+H_u^T\cdot r_i^h+r_u^{hT}\cdot H_i) \tag{14}$$

其中，x_u 和 y_i 代表用户 u 与项目 i 的在隐空间的特征向量，Z_u 为用户 u 的跨平台偏好特征，H_u 和 H_i 分别是用户 u 与项目 i 经 HAN 处理后的高阶特征向量。r_i^z、r_i^h 和 r_u^h 分别为与 Z_u、H_u 和 H_i 相对应的隐向量，与 x_u 和 y_i 一样也需要在模型训练的过程中进行学习。

使用二元交叉熵损失函数来训练模型，损失函数的定义如式（15）所示：

$$Loss=-\frac{1}{N}\sum_{r_{u,i}}[r_{u,i}log(\hat{r}_{u,i})+(1-r_{u,i})log(1-\hat{r}_{u,i})] \tag{15}$$

其中，N 为样本总数，$r_{u,i}$ 为 u 与 i 实际标签，二元交叉熵损失函数用于测量模型预测概率与实际标签之间的差异。模型以损失函数最小化为目标，使用 Adam 优化器更新模型参数。模型完成训练后，对用户与项目之间的预测得分进行排序，为用户推荐排名较高的项目列表。

3 实验分析

3.1 数据集

由于目前尚未有公开的与推荐算法相关的跨平台数据集，本研究选取知乎和微博平台分别作为目标平台和源平台，以推荐知乎用户所关注的问题为实验目标，自主构建所需数据集。本文通过网络爬虫技术在知乎中随机爬取生活、娱乐、学习和时政 4 个领域的问题及关注该问题的知乎用户数据。知乎为用户提供了公开其他社交媒体账号的功能，通过解析知乎用户的 JSON 数据可以得到部分用户的微博 ID，以匹配的同一用户作为实验的用户集来源。进一步地，爬取匹配用户在知乎以及微博中的属性和发布内容，由于微博的系统限制，无法获取全部的微博用户关注信息，本文仅爬取知乎用户的关注列表以提取匹配用户之间的社交结构信息。

在获取数据集之后，为降低冗余数据对模型效果的潜在负面影响，在 4 个领域的数据集中分别删除关注量少于 20 的问题和关注问题数量不足 10 的用户。数据集最终的基本统计信息如表 1 所示，本文构造的跨平台信息推荐的数据集规模较大，且信息种类多样，不仅弥补了推荐领域中跨平台多属性和细粒度数据集的空缺，也对实验模型的潜在稳健性提出了较高要求。各数据集中的用户—问题交互关系的稀疏程度均在 99% 以上，稀疏的交互数据对模型性能提出了更高要求。重叠用户的微博内容数据量显著高于知乎内容量，为使用源平台的密集数据解决目标平台推荐的冷启动问题提供契机。本文构造的大规模跨平台数据集不仅体现了研究的广度和深度，也为评估模型在不同数据稠密度下的适应性和稳健性提供了实验基础。

3.2 实验设置

本文根据问题、用户、问题作者、问题标签和问题分词 5 种类型的节点及其之间的关系构建异质信息网络，使用 TransE 模型训练各个节点的初始向量。提取异质网络中以用户和问题分别作为开头和结尾的元路径，不同的元路径代表不同的语义或相互关系，各元路径的语义含义及其对应的关系数量如表 2 所示。HAN 可以捕获异质图中复杂的关系结构、聚合多层次信息以及动态调整关系权重。表 2 中的数据展示出用户间、问题间的多维度关系具有异质性和不均匀性等特点，符合 HAN 能够发挥最大效果的应用场景，模型可以最大化地利用具有丰富多样性和复杂性的数据。

使用 Stanford CoreNLP 对用户的内容文本进行命名实体识别，保留与用户行为密切相关的组织、人员和地点类型的命名实体[1]，将命名实体映射到腾讯 AI 大型中文词向量数据集中进行向量化表示。本模型基于 Pytorch 框架实现。在参数设置方面，经过多轮实验，最终确认参数为：HAN 和一维卷积网络的输出节点特征维度均为 64 维，HAN 的多头注意力数量为 4，隐层单元大小为 4，卷积核大小为 3；使用 Xavier 初始化模型参数，学习率 0.01，批量为 128，迭代训练 30 次。在数据集处理方面，将用户集合划分为 90% 的训练集与 10% 的测试集。随机生成负样本，保证训练集的正负样本比例 1∶1，以达到提高训练稳定性和防止模型过拟

表 1 数据集基本信息

Tab. 1 Basic Information of the Datasets

数据类型	生活领域	娱乐领域	学习领域	时政领域
问　　题	13 924	9 388	9 525	8 071
用　　户	8 587	6 301	6 078	4 877
用户—问题关注数据	414 296	272 801	263 089	200 117
问题作者	9 105	6 611	6 481	5 579
问题标签	832	705	765	647
用户社交关系	191 870	104 819	104 923	69 320
用户的知乎属性	8 587	6 301	6 078	4 877
用户的微博属性	8 587	6 301	6 078	4 877
用户的知乎内容	276 995	229 024	233 592	212 814
用户的微博内容	925 109	663 138	624 977	525 720

表 2 元路径的语义及其对应的关系数量

Tab. 2 Semantics of Meta-paths and Corresponding Number of Relations

元路径	语义	生活领域	娱乐领域	学习领域	时政领域
UFU	同一核心兴趣朋友圈中的两个用户	76 113	51 716	51 338	38 517
UUU	两个用户具有相同的核心兴趣朋友	627 853	401 304	408 399	284 152
QWQ	两个问题具有相同分词	4 550 958	2 676 880	3 509 527	3 065 847
QAQ	两个问题被同一作者提出	39 106	25 326	19 857	17 747
QCQ	两个问题属于同一标签	13 258 188	4 958 116	4 159 255	4 594 287
QSQ	两个问题的语义相似	129 630	126 874	105 607	92 947
QPQ	两个问题具有共现相关性	107 600	84 482	90 067	78 585

合的目的。

3.3 对比模型和评估指标

为验证本文所提模型的有效性，将本模型与以下模型进行对比。

1）MF：经典的矩阵分解模型，将用户—项目交互矩阵分解为低维度的潜在特征向量的乘积。该模型依赖用户—项目交互信息进行因子分解，通过学习用户和项目在潜在空间上的表示，进而预测用户对未知项目的偏好程度。

2）RippleNet：一种基于知识图谱的推荐算法[38]。旨在通过模拟用户兴趣在知识图谱中的“涟漪”传播来提高推荐质量，核心思想是通过图谱传播用户兴趣点，以捕获用户多样化的潜在兴趣，使推荐算法有效地利用图中的结构化信息。

3）PGPR：一种基于强化知识图谱推理的推荐算法[39]。将推荐问题转化为知识图谱上的一个确定性马尔可夫决策过程，提出了一种策略性路径推理的方法，将知识图谱路径推理的思想应用于推荐系统，采用强化学习的方法使智能体学习如何导航到用户潜在感兴趣的项目。

模型将为每个用户生成一个推荐列表，本文采用平均倒数排名（Mean Reciprocal Rank，MRR）和前 K 位命中率 Hits@K 作为评估模型性能的指标。

1）MRR：用于衡量推荐结果排序质量的指标，它通过计算用户实际互动项在推荐列表中排名倒数的平均值来评估推荐系统的效果。具体计算过程如式（16）所示：

$$MRR=\frac{1}{|U|}\sum_{u=1}^{|U|}\frac{1}{rank_u} \tag{16}$$

其中，$|U|$是用户的总数，$rank_u$ 是用户 u 的互动项在推荐列表中的排名。

2）Hits@K：测量前 K 个推荐结果的命中率指

标，表示推荐列表的前 K 项中有正确推荐的概率。具体计算过程如式（17）所示：

$$Hits@K=\frac{1}{|U|}\sum_{u=1}^{|U|}I(rank_u\leqslant K) \tag{17}$$

其中，I 是指示函数，如果 $rank_u\leqslant K$，则 I 为 1，否则为 0。

3.4 实验结果

为更好地体现模型效果，选择在两个平台均有发布内容的用户进行实验，表 3 列出了 4 种模型在不同数据集下得到的 MRR、Hits@1、Hits@3 和 Hits@10 指标。总体来看，MF 模型取得的推荐效果较差，没有在特定指标上表现出突出的优势，MF 主要依赖于用户—项目交互数据，无法充分获取用户偏好和领域知识，限制了其在处理复杂推荐场景时的性能。RippleNet 和 PGPR 都能够利用异质信息网络为推荐提供额外的语义信息，因此在推荐效果上优于 MF。RippleNet 在 MRR 和 Hits@1 指标上表现较好，用户兴趣点在网络中的传播增强了 RippleNet 的推荐的精确度和相关性，但是由于其特征融合和信息利用的效率较低，模型在 Hits@3 和 Hits@10 的表现不佳。PGPR 在 Hits@3 和 Hits@10 的表现较好，PGPR 通过强化学习路径搜寻的方式，在为用户提供多样化推荐方面有一定的优势，但是在精准匹配用户核心需求方面的能力有限。通过高效地融合用户跨平台信息偏好，同时结合 HAN 增强用户和项目特征的表示能力，本文提出的 CPHAR 模型推荐效果均优于以上对比模型，能够有效地解决用户冷启动和项目数据稀疏性的问题，提升推荐结果的准确性、多样性和覆盖度。

表 3 实验结果对比

Tab. 3 Comparison of Experimental Results

数据集	指 标	MF	RippleNet	PGPR	CPHAR
生活领域	MRR	0.197	0.222	0.232	0.339
	Hits@1	0.086	0.138	0.099	0.197
	Hits@3	0.178	0.197	0.250	0.395
	Hits@10	0.487	0.382	0.533	0.671
娱乐领域	MRR	0.192	0.261	0.197	0.306
	Hits@1	0.084	0.177	0.076	0.143
	Hits@3	0.168	0.235	0.185	0.336
	Hits@10	0.420	0.454	0.487	0.656
学习领域	MRR	0.159	0.209	0.217	0.300
	Hits@1	0.046	0.130	0.074	0.130
	Hits@3	0.167	0.148	0.241	0.324
	Hits@10	0.352	0.444	0.565	0.667
时政领域	MRR	0.171	0.221	0.244	0.287
	Hits@1	0.050	0.140	0.100	0.150
	Hits@3	0.170	0.190	0.250	0.300
	Hits@10	0.500	0.370	0.610	0.600

3.5 跨平台用户偏好建模效果分析

为探究模型中跨平台用户偏好建模的效果，使用本模型对仅在源平台和仅在目标平台有内容信息的用户进行推荐，在保证用户数量一致的情况下与具有跨平台内容信息用户的推荐结果进行比较，实验结果如图 2、图 3 所示。总体来看，相较于仅在单平台中具有内容信息的用户，模型对于具有跨平台内容的用户推荐效果更好，说明本模型能够有效地融合和利用跨平台内容中的关键信息，实现更优的推荐效果。同时，模型对于仅在源平台有数据的用户也实现了较好的推荐效果，这一意外的实验发现不仅说明引入用户在其他平台的内容信息对目标平台用户数据进行补充具有一定的合理性，验证了 Nie Y 等[29]提出的用户在不同平台中具有相似兴趣

偏好的论点，也进一步证明了本模型对于目标平台中完全冷启动的用户同样具有较好的推荐性能，模型具有一定的普适性。

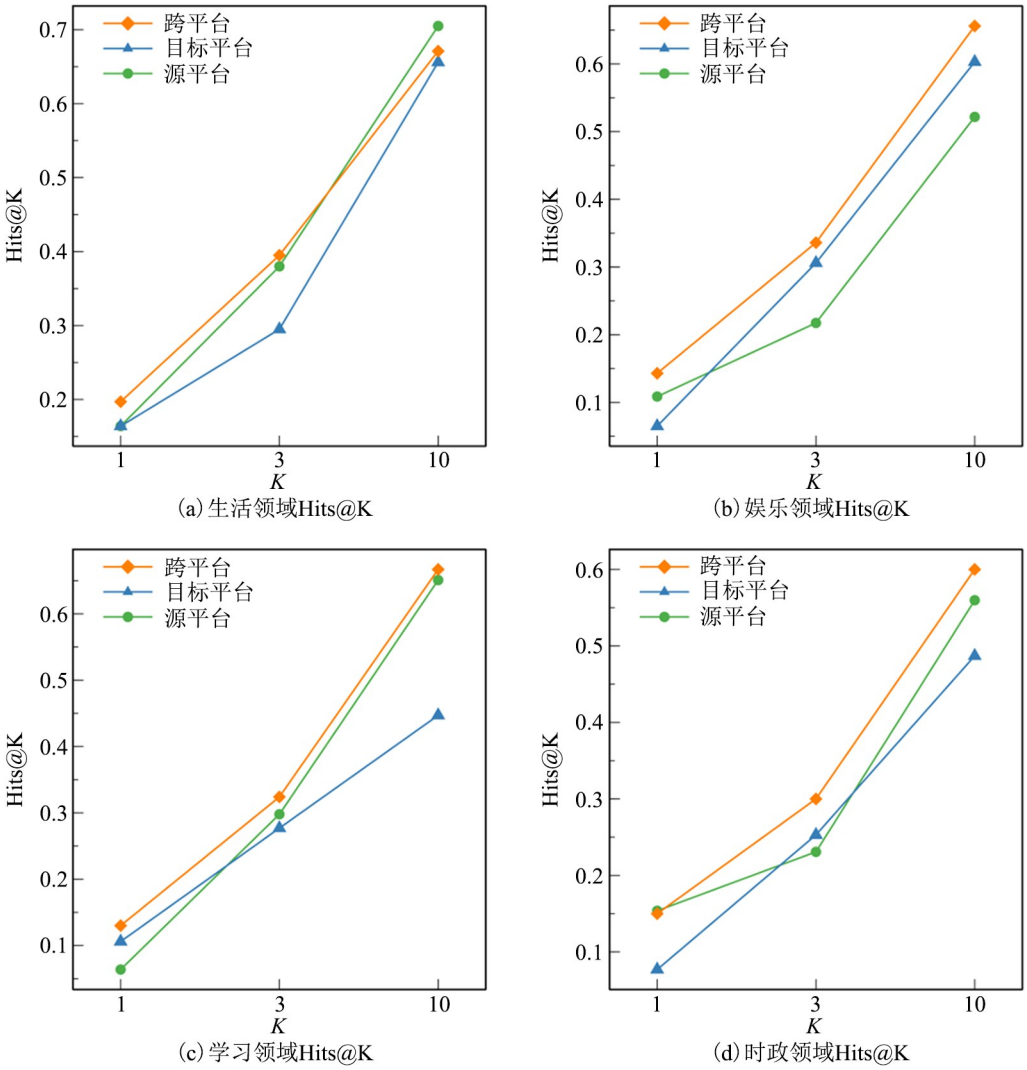

图 2　不同用户偏好建模下的 Hits@K

Fig. 2　Hits@K Under Different User Preference Modeling

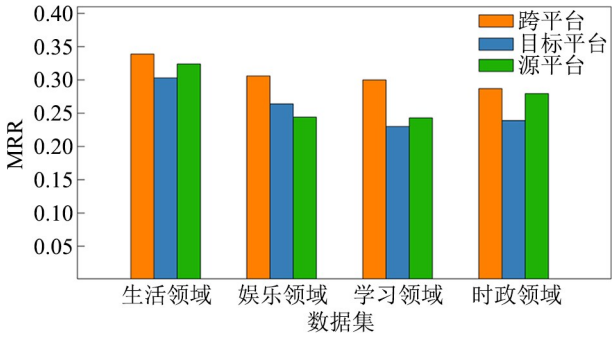

图 3　不同用户偏好建模下的 MRR

Fig. 3　MRR Under Different User Preference Modeling

3.6　消融实验

消融实验进一步探究模型构建的用户跨平台核心兴趣朋友圈以及 HAN 高阶特征聚合模块对模型结果的影响。具体来讲，CPHAR_DU 模型将 CPHAR 模型中的核心兴趣朋友圈替换为用户关注朋友列表，CPHAR_DH 模型移除了 CPHAR 模型中的 HAN 模块，直接使用 TransE 得到的用户和项目向量进行实验，各数据集的消融实验结果如图 4 所示。整体来看，CPHAR 模型的性能要显著优于两个消融模型，证明了 CPHAR 在进行用户核心兴趣挖掘和高阶特征聚合方面的有效性和优越性。CPHAR_DU 使用用户全部的社交结构关系，未考虑到不同朋友的差异性特征以及关键用户产生的重要影响，融合所有具有社交关系的用户在一定程度上干扰了对用户自身特征的识别，且大大降低了模型的运行效率。CPHAR_DH 使用 TransE 进行节点和关系的向量化，

只关注了异质实体之间的直接关系，而无法有效应用异质信息网络中复杂的多跳路径关系，对实体在不同元路径下的特征表达能力有限。CPHAR_DH 模型的推荐性能相对较差，证明了 HAN 高阶特征聚合对提升模型预测能力发挥重要贡献。

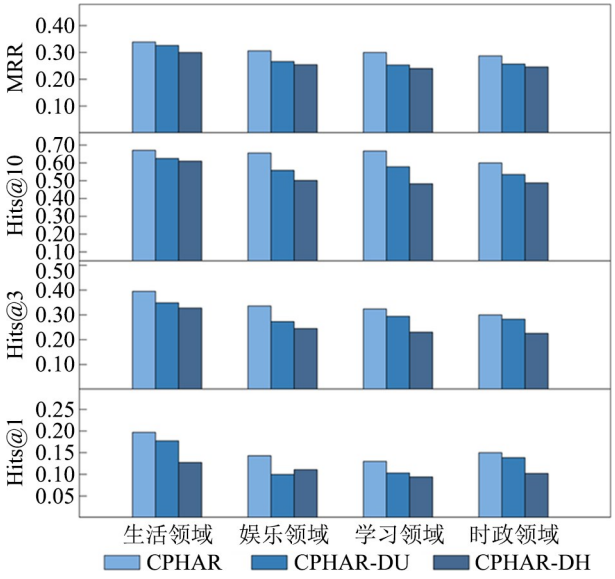

图 4 消融实验结果

Fig. 4 Ablation Study Results

4 结 论

针对当前信息推荐领域存在的数据稀疏和用户冷启动的问题，本文提出一种融合跨平台用户偏好与异质信息网络的推荐模型。该模型整合跨平台多源异构数据识别用户核心兴趣朋友圈，通过卷积神经网络和注意力机制挖掘用户跨平台的信息偏好特征，结合项目语义相似度和 PMI 数值挖掘推荐项目的隐形关联。不仅完成了对跨平台大规模异质信息网络拓扑关系的降噪和完善，也进一步实现了对用户模糊性和多样化偏好的准确识别和迁移融合的优化目标。此外，优化了传统矩阵分解模型，利用神经网络模型将用户跨平台信息偏好和使用 HAN 聚合后的用户和项目高阶特征纳入推荐模型中，较全面地融合了不同元路径上的语义信息，达到了有效利用平台间丰富特征信息以提升模型预测能力的目的。在真实数据集中的实验结果表明，本文模型在各项评估指标上均表现出了显著的优势，对于目标平台中完全冷启动的用户同样具有较好的推荐表现，说明模型在提高推荐效果和优化用户冷启动方面更具优越性和稳定性。消融实验进一步证明了模型构建的跨平台核心兴趣朋友圈和 HAN 高阶特征聚合模块对模型性能的提升发挥重要作用。

在实现上述技术创新的同时，本研究还具有广阔的延伸应用价值。本文所提模型将为多领域多情景下的用户偏好特征建模及推荐应用提供借鉴，为基于场景精细化和跨域关联式的信息资源推荐提供范式拓新。本文模型的普适性和可扩展性较强，可以基于用户在不同场景下的不对称、不均匀的异构数据实现全方位的用户偏好建模，通过充分挖掘和融合多场景下用户和项目的复杂关联，突破单场景下推荐算法的认知局限与偏差，实现跨平台或者跨领域的精准推荐。具体来讲，本文模型可从跨平台的信息推荐应用扩展至图书、专利、科技文献、在线出版物等信息资源的推荐，全面激活与整合数据的价值要素，进一步提升信息资源的利用效率，助力算法技术的革新与信息资源管理的高质量发展。

本文模型也存在一些不足，对于用户跨平台的属性和发布内容数据，模型仅提取了其中的文本特征，忽略了其他相关的多模态数据特征。在后续研究中，将考虑结合图片、视频以及用户的地理位置等信息，更全面地解读用户跨平台信息偏好特征，进一步拓展本研究的内容。此外，未来研究可以进一步结合用户在更多平台和领域中的异质特征信息，在复杂推荐场景下对模型进行进一步优化。

参 考 文 献

[1] Gao H，Wang Y，Shao J，et al. User Identity Linkage Across Social Networks with the Enhancement of Knowledge Graph and Time Decay Function [J]. Entropy，2022，24 (11)：1603.

[2] 马为之. 面向异质环境的用户建模与推荐方法研究 [D]. 北京：清华大学，2019.

[3] 易明，刘明，冯翠翠. 融合异质信息网络表示学习的跨领域推荐研究 [J]. 情报学报，2022，41 (4)：337-349.

[4] 李丹阳，甘明鑫. 基于多源信息融合的音乐推荐方法 [J]. 数据分析与知识发现，2021，5 (2)：94-105.

[5] Xu Z，Wei P，Liu S，et al. Correlative Preference Transfer with Hierarchical Hypergraph Network for Multi-domain Recommendation [C] //Proceedings of the ACM Web Conference 2023，2023：983-991.

[6] Jiang Y，Li Q，Zhu H，et al. Adaptive Domain Interest Network for Multi-domain Recommendation [C] //Proceedings of the 31st ACM International Conference on Information and Knowledge Management，2022：3212-3221.

[7] 汪春播，温继文. 基于异构信息网络的推荐研究综述 [J]. 计算机工程与科学，2023，45 (11)：2047-2059.

[8] Salamat A, Luo X, Jafari A. HeteroGraphRec: A Heterogeneous Graph-based Neural Networks for Social Recommendations [J]. Knowledge-Based Systems, 2021, 217: 106817.

[9] 李广建, 罗立群. 走向知识融合——大数据环境下情报学的发展趋势 [J]. 中国图书馆学报, 2020, 46 (6): 26-40.

[10] Ko H, Lee S, Park Y, et al. A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields [J]. Electronics, 2022, 11 (1): 141.

[11] 张彬, 徐建民, 吴姣. 跨域推荐中的知识融合研究进展 [J]. 现代情报, 2023, 43 (3): 157-166.

[12] 张明星, 张骁雄, 刘姗姗, 等. 利用知识图谱的推荐系统研究综述 [J]. 计算机工程与应用, 2023, 59 (4): 30-42.

[13] 马鑫, 王芳, 段刚龙. 面向电商内容安全风险管控的协同过滤推荐算法研究 [J]. 情报理论与实践, 2022, 45 (10): 176-187.

[14] 丁浩, 胡广伟, 王婷, 等. 基于时序漂移的潜在因子模型推荐方法 [J]. 数据分析与知识发现, 2022, 6 (10): 1-8.

[15] 钱聪, 齐江蕾, 丁浩. 基于用户多重兴趣漂移特征权重的网络出版物推荐研究 [J]. 数据分析与知识发现, 2023, 7 (8): 119-127.

[16] Yang M, Zhou P, Li S, et al. Multi-Head Multimodal Deep Interest Recommendation Network [J]. Knowledge-Based Systems, 2023, 276: 110689.

[17] Das J, Majumder S, Gupta P, et al. Scalable Recommendations Using Decomposition Techniques Based on Voronoi Diagrams [J]. Information Processing and Management, 2021, 58 (4): 102566.

[18] Zang T, Zhu Y, Liu H, et al. A Survey on Cross-domain Recommendation: Taxonomies, Methods, and Future Directions [J]. ACM Transactions on Information Systems, 2022, 41 (2): 1-39.

[19] Zhao C, Li C, Xiao R, et al. CATN: Cross-domain Recommendation for Cold-start Users via Aspect Transfer Network [C] //Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020: 229-238.

[20] Zhang Q, Wu D, Lu J, et al. A Cross-domain Recommender System with Consistent Information Transfer [J]. Decision Support Systems, 2017, 104: 49-63.

[21] Zhang Q, Lu J, Wu D, et al. A Cross-domain Recommender System with Kernel-induced Knowledge Transfer for Overlapping Entities [J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 30 (7): 1998-2012.

[22] 何喜军, 吴爽爽, 武玉英, 等. 基于属性异构网络表示学习的专利交易推荐 [J]. 情报学报, 2022, 41 (11): 1214-1228.

[23] 时倩如, 李贺, 沈旺, 等. 基于高阶和低阶交互关系的深度学习推荐模型研究 [J]. 情报理论与实践, 2024, 47 (4): 189-196.

[24] Shi C, Hu B, Zhao W X, et al. Heterogeneous Information Network Embedding for Recommendation [J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 31 (2): 357-370.

[25] Li L, Gui X, Lv R. Recommendation Algorithm Based on Heterogeneous Information Network and Attention Mechanism [J]. Applied Sciences, 2023, 14 (1): 353.

[26] 熊回香, 唐明月, 叶佳鑫, 等. 融合加权异质网络与网络表示学习的学术信息推荐研究 [J]. 现代情报, 2023, 43 (5): 23-34.

[27] Bi Y, Song L, Yao M, et al. A Heterogeneous Information Network Based Cross Domain Insurance Recommendation System for Cold Start Users [C] //Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020: 2211-2220.

[28] Qiao S, Zhou W, Luo F, et al. Noise-reducing Graph Neural Network with Intent-target Co-action for Session-based Recommendation [J]. Information Processing and Management, 2023, 60 (6): 103517.

[29] Nie Y, Jia Y, Li S, et al. Identifying Users across Social Networks Based on Dynamic Core Interests [J]. Neurocomputing, 2016, 210: 107-115.

[30] Celik M, Dokuz A S. Discovering Socially Similar Users in Social Media Datasets Based on Their Socially Important Locations [J]. Information Processing and Management, 2018, 54 (6): 1154-1168.

[31] 魏玲, 权晨雪. 融合多维特征与兴趣漂移的虚拟学术社区群推荐模型 [J]. 现代情报, 2023, 43 (7): 48-63.

[32] 张佳, 董守斌. 基于评论方面级用户偏好迁移的跨领域推荐算法 [J]. 计算机科学, 2022, 49 (9): 41-47.

[33] Wang X, Ji H, Shi C, et al. Heterogeneous Graph Attention Network [C] //The World Wide Web Conference, 2019: 2022-2032.

[34] 江旭晖, 沈英汉, 李子健, 等. 社交知识图谱研究综述 [J]. 计算机学报, 2023, 46 (2): 304-330.

[35] Meng X, Bai L, Hu J, et al. Multi-hop Path Reasoning over Sparse Temporal Knowledge Graphs Based on Path Completion and Reward Shaping [J]. Information Processing and Management, 2024, 61 (2): 103605.

[36] Yin J, Guo Y, Chen Y. Heterogenous Information Network Embedding Based Cross-domain Recommendation System [C] //2019 International Conference on Data Mining Workshops. IEEE, 2019: 362-369.

[37] 廖宏建, 谢亮, 曲哲. 一种基于隐式信任感知的 MOOCs 推荐方法 [J]. 情报理论与实践, 2021, 44 (2): 128-135, 95.

[38] Wang H, Zhang F, Wang J, et al. Ripplenet: Propagating User Preferences on the Knowledge Graph for Recommender Systems [C] //Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018: 417-426.

[39] Xian Y, Fu Z, Muthukrishnan S, et al. Reinforcement Knowledge Graph Reasoning for Explainable Recommendation [C] //Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019: 285-294.

(责任编辑: 杨丰侨)