

基于多模态异质图网络的专利推荐算法^{*}

赖培源^{1,2,4}, 卢伊虹³, 廖德章¹, 王昌栋^{3,4}, 戴青云^{2,4}, 赖剑煌³



¹(广东省华南技术转移中心有限公司, 广东 广州 511458)

²(广东工业大学 信息工程学院, 广东 广州 510006)

³(中山大学 计算机学院, 广东 广州 510006)

⁴(广东省知识产权大数据重点实验室, 广东 广州 510665)

通信作者: 王昌栋, E-mail: changdongwang@hotmail.com

摘 要: 通过专利推荐将科技创新成果转化为现实生活中的实际应用, 让科学技术实现经济价值, 对社会经济发展具有重大意义. 然而, 现有的专利推荐算法往往忽略了专利本身所包含的多模态信息, 导致推荐结果无法全面反映专利的真实价值与应用潜力, 进而影响专利与企业需求之间的匹配精度. 为此, 提出了一种基于多模态异质图网络的专利推荐算法 (multimodal heterogeneous graph network for patent recommendation, MHGN). 首先, 利用预训练表征模型将专利的多属性文本信息与图像以及企业信息进行初始化表征学习. 随后, 采用图注意力网络学习企业在不同模态下的偏好表征, 在此基础上, 进一步基于偏好表征的相似度学习企业-专利交互的关系权重, 并设计了一个图卷积网络来学习企业和专利的节点偏好表征. 最后, 引入了适配向量, 并使用注意力机制对节点偏好表征与多模态表征进行融合. 在实验验证上, 构建了 4 个真实的高校向企业转让的专利数据集, 并与 7 个先进的基线模型进行了实验对比, 结果表明, 所提模型在各项指标上均显著优于基线模型.

关键词: 专利推荐; 多模态; 异质图; 技术转移; 适配向量

中图法分类号: TP181

中文引用格式: 赖培源, 卢伊虹, 廖德章, 王昌栋, 戴青云, 赖剑煌. 基于多模态异质图网络的专利推荐算法. 软件学报, 2026, 37(5): 1964–1981. <http://www.jos.org.cn/1000-9825/7537.htm>

英文引用格式: Lai PY, Lu YH, Liao DZ, Wang CD, Dai QY, Lai JH. Patent Recommendation Algorithm Based on Multimodal Heterogeneous Graph Network. Ruan Jian Xue Bao/Journal of Software, 2026, 37(5): 1964–1981 (in Chinese). <http://www.jos.org.cn/1000-9825/7537.htm>

Patent Recommendation Algorithm Based on Multimodal Heterogeneous Graph Network

LAI Pei-Yuan^{1,2,4}, LU Yi-Hong³, LIAO De-Zhang¹, WANG Chang-Dong^{3,4}, DAI Qing-Yun^{2,4}, LAI Jian-Huang³

¹(Guangdong South China Technology Commercialization Center Co. Ltd., Guangzhou 511458, China)

²(School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China)

³(School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China)

⁴(Guangdong Provincial Key Laboratory of Intellectual Property and Big Data, Guangzhou 510665, China)

Abstract: The transformation of scientific and technological innovations into practical applications through patent recommendation is of great significance for realizing the economic value of science and technology and promoting socio-economic development. However, existing patent recommendation algorithms often overlook the multimodal information embedded in patents, leading to recommendation results that fail to comprehensively reflect the value and application potential of patents. Consequently, this affects the accuracy of

* 基金项目: 国家自然科学基金 (62276277); 广东省技术转移智能匹配工程技术研究中心项目 (2022A175); 广东省知识产权大数据重点实验室 (2018B030322016)

本文由“多媒体智能理解与生成”专题特约编辑孙立峰教授、闵巍庆副研究员、马占宇教授、蒋树强研究员、彭宇新教授、田丰研究员、黄庆明教授推荐.

收稿时间: 2025-02-27; 修改时间: 2025-07-11, 2025-08-29; 采用时间: 2025-09-05; jos 在线出版时间: 2025-09-23

CNKI 网络首发时间: 2025-12-04

matching patents with the needs of companies. To address this issue, this study proposes a multimodal heterogeneous graph network for patent recommendation (MHGN). The proposed method first utilizes pre-trained models to initialize the representation of multimodal information, including the textual and image attributes of patents as well as company information. Then, a graph attention network is employed to learn the preference representations of companies across different modalities. Based on this, the relationship weights of company-patent interactions are further learned based on the similarity of preference representations, and a graph convolutional network is designed to learn the node preference representations of companies and patents. Finally, to better integrate the multimodal information, an adaptation vector is introduced and an attention mechanism is used to fuse the node preference representations with multimodal representations. In addition, four real-world patent datasets from university-to-company transfers are constructed, and experiments are conducted comparing the proposed model with seven advanced baseline models. The results demonstrate that the proposed model significantly outperforms the baselines across all evaluation metrics.

Key words: patent recommendation; multimodal; heterogeneous graph; technology commercialization; adaptation vector

成果转化是将科研成果转变为实际应用和商业化产品的关键过程, 而专利作为知识产权的重要组成部分, 在这一过程中扮演着至关重要的角色. 向企业推荐专利对于推动科研成果向实际应用转化具有重大意义^[1]. 高校和科研机构存在大量“沉睡专利”, 推动这些专利“下书架、上货架”, 加速转化运用, 对加快发展新质生产力具有重要意义. 专利作为一种法律保护手段, 旨在保护发明创造, 促进技术创新和维护市场秩序, 许多研究将专利数量作为解释或比较国家和地区创新力量的重要标准^[2]. 在此基础上, 一些国家制定政策和采用一系列行动, 如给予专利研发财政奖励, 开发专利交易平台, 以促进专利的研发和成果转化^[3-5]. 此外, 研究表明高校中的专利技术主要集中在企业比较缺乏国际竞争优势的领域^[6], 这说明企业需要通过创新合作或成果转化来弥补其在这些领域的技术和市场竞争短板.

然而, 这一过程面临着诸多挑战, 一方面, 由于专利本身信息的复杂性, 涉及大量技术细节和法律术语, 需要深入理解和专业评估^[7]. 具体地, 与从事专利代理人等知识产权服务人员不同, 企业技术人员通常较少关注专利权利要求的法律表达, 而更关心专利所蕴含的技术实质与可应用性, 包括核心技术方案, 以及如何应用于实际生产、产品或工艺流程. 另一方面, 企业在筛选专利时必须确保其与自身产业结构和发展战略相符, 与企业的现有技术基础和未来发展方向高度契合. 企业技术人员更关注专利技术的技术优势和可行性, 即与企业现有技术相比, 专利技术在性能、成本或效率上能带来改进与差异, 在企业研发或制造环境中落地应用的可能性. 因此, 在专利推荐的过程中, 需要综合考量专利本身的技术复杂性以及与企业的技术适应性. 而纯粹的法律化专利表述往往会掩盖技术贡献, 使得企业技术人员难以快速判断其相关性和可用性, 可见, 通过智能化技术为企业技术人员提供专利推荐服务具有重大意义.

专利转让推荐作为成果转化的关键一环, 已经引起越来越多的研究关注. 以往的研究通常采用基于协同过滤的推荐方法^[8-10]、基于内容的推荐方法^[11-13]以及基于结合协同过滤和内容的混合推荐方法^[14-16]. 这些方法主要基于查询驱动的专利检索推荐场景, 然而在应对专利转让的背景下, 无法深层次建模企业和专利的潜在联系. 针对这一问题, 许多研究探索了新的方法^[17-19], 例如, 文献^[18]提出了一种自监督神经网络, 可以捕捉到企业潜在偏好与专利共分类信息中隐含的语义信息的相似性. 进一步地, 为了应对产业集群的技术创新发展和高效的成果转化需求, 面向产业集群的专利推荐方法被提出^[20,21], 通过获取丰富的历史专利以及集群内各企业的相关性, 可以充分捕获企业的技术需求以及专利的丰富信息, 提高推荐性能.

在现代科技成果转化过程中, 专利推荐的准确性和有效性对于企业创新至关重要. 专利的多模态信息 (如文本和图像) 共同构成了其独特的价值特征. 例如, 一项复杂的技术专利不仅包含详尽的文本描述, 还附有精细的图示, 这些图示可以帮助潜在的技术受让方更直观地理解专利的应用场景和技术原理. 在许多情况下, 整合文本和图片信息通常有助于更精确、更轻松地消除歧义^[22], 提高推荐的准确性. 然而, 传统的方法都没有考虑专利本身的多模态信息, 这使得推荐的准确性和全面性大打折扣.

因此, 使用多模态异质图网络对专利进行建模具有重要意义. 不同模态的信息构成了一个异质图结构, 其中专利和企业之间的关系可以通过复杂的企业-专利交互图来表示. 如图 1 所示, 不同颜色节点代表不同的特征类型, 连接专利 (P) 和企业 (C) 节点的边表示专利和企业之间的转让关系, 通过不同类型的节点和边, 将多模态信息和异

质信息整合在一起, 形成了一个复杂的企业-专利交互图. 如何有效地捕捉企业与专利的高阶特征, 进而从中学习和提取异质图中的有用信息, 是一个巨大的挑战. 多模态信息的存在使得对特征的提取和关系的捕捉变得更加复杂, 而异质图的结构则增加了对企业与专利关系进行建模的难度. 因此, 设计一个能够充分利用多模态信息并有效解析异质图结构的推荐算法, 是解决这一挑战的关键.

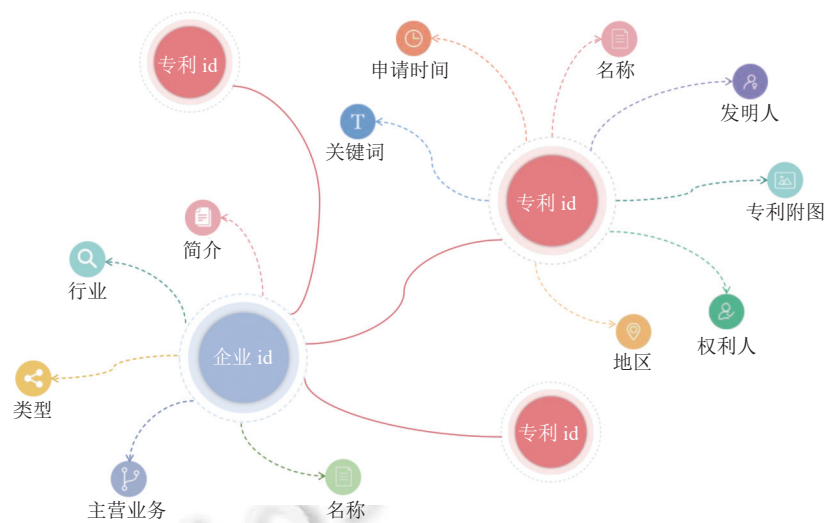


图 1 企业-专利交互的异质信息图

为解决上述问题, 本文提出了一种基于多模态异质图网络的专利推荐 (multimodal heterogeneous graph network for patent recommendation, MHGN) 模型. 在 MHGN 模型中, 首先对企业和专利的多模态信息进行初始特征编码. 具体来说, 企业的信息包括企业介绍、经营范围、所属行业等, 这些信息能够反映企业的潜在技术需求. 通过预训练语言模型对企业信息进行显式建模, 生成企业的表征向量. 与此同时, 专利的文本信息由标题、摘要、关键词、权利要求书等组成, 而专利的图像则包括流程图和设计图等. 同样采用预训练模型对这些多模态专利信息进行表征, 得到初始化的文本和图像表征. 为了进一步编码企业与专利的多模态信息, MHGN 分别在不同模态下构建了企业-专利交互图, 将企业和专利分别作为节点, 节点之间的边表示专利转让给企业的关系. 接着, 本文设计了一个多模态特征编码模块, 利用图注意力网络 (graph attention network, GAT) 来融合企业和专利之间的多模态交互信息, 得到企业的多模态偏好表征, 这种偏好捕捉了企业对专利的潜在需求程度. 在此基础上, 本文进一步构建了一个偏好权重学习的图卷积模块, 基于偏好表征的相似度学习企业-专利交互的关系权重, 然后通过图卷积网络 (graph convolutional network, GCN) 学习企业和专利的节点偏好表征. 随后, 将学习的节点偏好表征与企业和专利的多模态特征进行拼接, 通过引入适配向量和注意力机制进一步对多模态特征进行融合与加权. 最终通过计算企业和专利的相似度来预测企业是否对某一专利有需求.

本文的主要贡献如下.

- (1) 提出了一种基于多模态异质图网络的专利推荐 (MHGN) 方法. 这是结合多模态特征进行专利转让推荐的应用, 是将异质图神经网络应用到多模态专利推荐的一个新方向.
- (2) 为了缓解现有方法中多模态数据带来的噪声问题, 引入了适配向量以增强和融合多模态特征, 有效地减少了不相关特征对模型性能的影响. 此外, 采用注意力机制, 使模型能够关注更加重要的特征, 从而提升推荐精度.
- (3) 提取并处理了 4 个全新的、真实的专利转让数据集, 并在这些数据集上全面验证本文提出的 MHGN 方法. 实验结果表明, MHGN 在所有数据集上均显著优于现有最新方法, 进一步证明了本文方法在专利转让推荐任务中的有效性和优越性.

本文第 1 节介绍相关工作, 包括专利推荐、基于图的多模态推荐. 第 2 节介绍本文构建的基于多模态异质图

网络的专利推荐算法. 第3节介绍实验情况, 包括对比实验、消融实验和参数实验, 验证分析所提模型的有效性. 最后总结全文.

1 相关工作

1.1 专利推荐

随着专利数量的不断增加, 如何有效地推荐合适的专利给企业变得越来越重要. 为应对这一挑战, 已有许多研究提出了各种专利推荐方法, 这些方法主要分为4大类: 基于协同过滤、基于内容的推荐、混合方法推荐以及基于深度学习的序列推荐方法.

基于协同过滤 (collaborative filtering, CF) 的方法通过分析用户的行为偏好进行推荐^[23], 具有出色的推荐性能与良好的可扩展性^[24]. 这类方法不直接依赖于专利的具体内容, 它通过找到与目标用户具有相似行为的其他用户获得的专利进行推荐^[8,9,25,26]. 例如, Trappey 等人^[8]开发了一种基于用户的智能专利推荐系统, 采用专利搜索方法识别具有相似专利搜索行为的用户, 并将相似用户喜欢的专利推荐给目标用户. 对于目标用户, Ji 等人^[9]构建了用户-专利交互矩阵, 根据用户对专利的评分确定相似用户, 并推荐被相似用户评为高分的专利. 基于 CF 的方法简单高效, 然而, 在专利转让的背景下, 基于 CF 的方法具有局限性, 因为大多数转让的专利只能转让给一家企业^[14], 这导致无法充分利用企业的行为数据, 容易出现冷启动问题和数据稀疏性问题.

基于内容 (content based, CB) 的推荐方法通过分析专利的文本信息、技术特征等其他内容属性, 推荐与用户之前所查询过的专利相似的新专利^[27-29]. Chen 等人^[30]通过生成基于国际专利分类的向量空间模型, 构建出目标用户的历史专利表征与候选专利表征, 再计算两者的余弦相似度来进行推荐. Krestel 等人^[31]通过构建的 Dirichlet 分配模型, 对用户查询和专利的主题分布进行比较, 向用户推荐主题类似于查询专利的专利. 基于 CB 方法可以利用丰富的内容信息进行推荐, 但是难以识别专利中的异质信息, 推荐的专利较为单一, 总是与目标公司的历史专利过于相似, 不适合于以技术创新为背景的企业用户.

混合方法推荐结合了协同过滤和基于内容的过滤的优点, 通过多种方式将这两种方法整合, 以提高推荐系统的整体性能. 例如, Oh 等人^[16]提出了一个两阶段的串行推荐框架, 一阶段用基于 CB 方法获取候选专利集; 二阶段采用基于 CF 方法生成最终的推荐列表. 此外, 为了学习到丰富的专利文本信息, 基于异质信息网络的推荐方法被广泛应用. Wang 等人^[14]基于元路径将 CF 和 CB 方法进行线性组合, 最后通过元路径相似度进行专利推荐. Liu 等人^[32]开发了一个专利检索和分析平台, 结合书目耦合和文本挖掘技术的专利检索混合方法进行推荐. 混合方法克服了 CF 和 CB 方法的缺点, 能够充分获取企业和专利的相关信息从而提高推荐精度.

基于深度学习的序列推荐旨在建模行为序列中动态兴趣的演变关系^[33]. 近年来, 基于序列的深度学习专利推荐方法主要是关注于公司的历史专利信息从而进行建模^[20,21], 将公司的历史专利信息等作为表征来捕获与候选专利之间的关系. 为了从公司现有的专利序列中捕获公司的专利偏好, Du 等人^[21]提出了一种基于知识感知注意力的双向长短期记忆网络, 通过知识嵌入来捕获专利的丰富信息, 引入双向长短期记忆网络 (bidirectional long short-term memory, BiLSTM) 来学习公司历史记录的顺序模式. 序列建模取得了很好的效果, 但是比较依赖丰富的历史专利信息. 这些方法面向的都是由多个公司组成的联盟集体, 拥有数量较多的历史专利, 然而对于单个的中小型企业或初创企业来讲, 拥有的专利数量不多, 所以并不适合序列建模.

在现有的专利推荐方法中, 尽管已经取得了一定的进展, 但这些方法都未能有效地利用专利中的多模态信息, 如文本、图像等, 这将导致模型在捕捉企业与专利之间的潜在关联时表现受限. 本研究首次关注专利中的多模态异质信息, 通过使用图神经网络融合专利的多模态特征, 从而提高推荐性能.

1.2 基于图的多模态推荐

近年来, 推荐系统变得越来越重要, 带来大量的商业价值和用户便利^[34]. 早期的多模态推荐模型都是在 CF 的基础上捕获用户的偏好^[35-38]. 得益于深度学习的发展, 图神经网络 (graph neural network, GNN) 凭借其在处理非欧氏空间数据及复杂特征方面的优越性受到了大量关注^[39], GNN 在推荐系统领域取得了显著进展^[40]. 最新的研究

将图神经网络引入到多模态推荐系统^[41], 相比于传统的方法, GNN 能够更好地利用在推荐系统中普遍存在的用户属性和商品属性信息^[42]. 为了挖掘物品的多模态信息, MMGCN^[43]提出了基于 GCN 的多模态推荐方法, 为每个模态分别构造用户-项目二部图, 再通过 GCN 进行特征聚合, 以增强用户和项目的表征. DualGNN^[44]在 MMGCN 的基础上, 利用用户之间的相关性, 基于二分图和用户共现图来捕捉用户对同一项目不同形态特征的偏好. 为了识别并减少用户-项目图中无意的交互, GRCN^[45]构建了图精化层, 通过识别噪声边和破坏虚假边来优化用户-项目交互图的结构. 为了挖掘出项之间的语义信息, LATTICE^[46]构造了项目-项目关系图来发现多模态特征下的潜在物品关系, 最后结合 CF 方法进行推荐. 在 LATTICE 的基础上, FREEDOM^[47]进一步检测到项目-项目图的学习是可以忽略的, 通过冻结该图并同时为用户-项目交互图进行去噪以获得有效和高效的推荐. 为了捕获图中的同质信息和异质信息, MMSR^[48]提出了基于图的方法, 通过自适应顺序融合模态特征, 采用双重注意力机制区分同质和异质邻居节点, 能够更好地识别并学习到模态间的关系程度. 相似地, 为了更好地融合多模态特征, DRAGON^[49]分别构建同质图和异质图来获取用户和项的双重表征. MMFN^[50]通过 GNN 提取跨模态交互信息, 通过门控多因素语义传感器深度提取特征, 提升了推荐系统的性能.

尽管现有的基于图的多模态推荐方法取得了一定的成功, 但它们仍存在一些不足. 首先, 这些方法通常难以有效地去除多模态特征中的噪声, 导致模型容易受到无关或误导性信息的干扰. 此外, 这些方法往往未能充分考虑不同特征的重要性和差异性, 导致特征融合的效果不佳.

本文受到基于深度学习的专利推荐方法和基于图的多模态推荐方法的启发, 提出了可以提取出丰富企业信息和融合专利异质信息的多模态异质图网络专利转让推荐模型, 它可以通过图神经网络来增强企业和专利的表征, 从而提高推荐性能.

2 方 法

本节介绍所提出的多模态专利推荐模型, 表 1 介绍了本文所使用的主要符号, 模型框架图如图 2 所示. 它由 5 个部分组成, 分别是初始特征编码模块、多模态特征编码模块、偏好权重学习图卷积模块、特征融合模块和推荐预测模块. 具体地, 本文首先通过预处理模型对企业和专利的信息进行编码; 然后根据专利和企业的初始表征得到专利和企业的多模态表征; 再通过计算多模态下的偏好权重, 对企业和专利进行图卷积操作, 将得到的节点偏好表征与多模态表征进行拼接, 并通过引入适配向量和注意力机制进一步对多模态表征进行融合与加权. 最后, 在专利推荐预测模块, 本文通过计算企业和专利的相似度来得到最后的推荐结果.

表 1 基础符号和说明

符号	含义
p_{txt}	专利初始文本表征
p_{img}	专利初始图像表征
$c_{(0)}$	企业初始文本表征
$\bar{c}_m$	任意模态下企业偏好表征
$\bar{p}_m$	任意模态下与企业偏好相关的专利表征
e_c	企业节点偏好表征
e_p	专利节点偏好表征
$\hat{p}$	专利多模态融合初始表征
$\hat{c}$	企业多模态融合初始表征
$\tilde{c}$	企业多模态适配融合表征
$\tilde{p}$	专利多模态适配融合表征
e_c^*	企业多模态融合表征
e_p^*	专利多模态融合表征
$\bar{e}_c^*$	企业综合表征
$\bar{e}_p^*$	专利综合表征

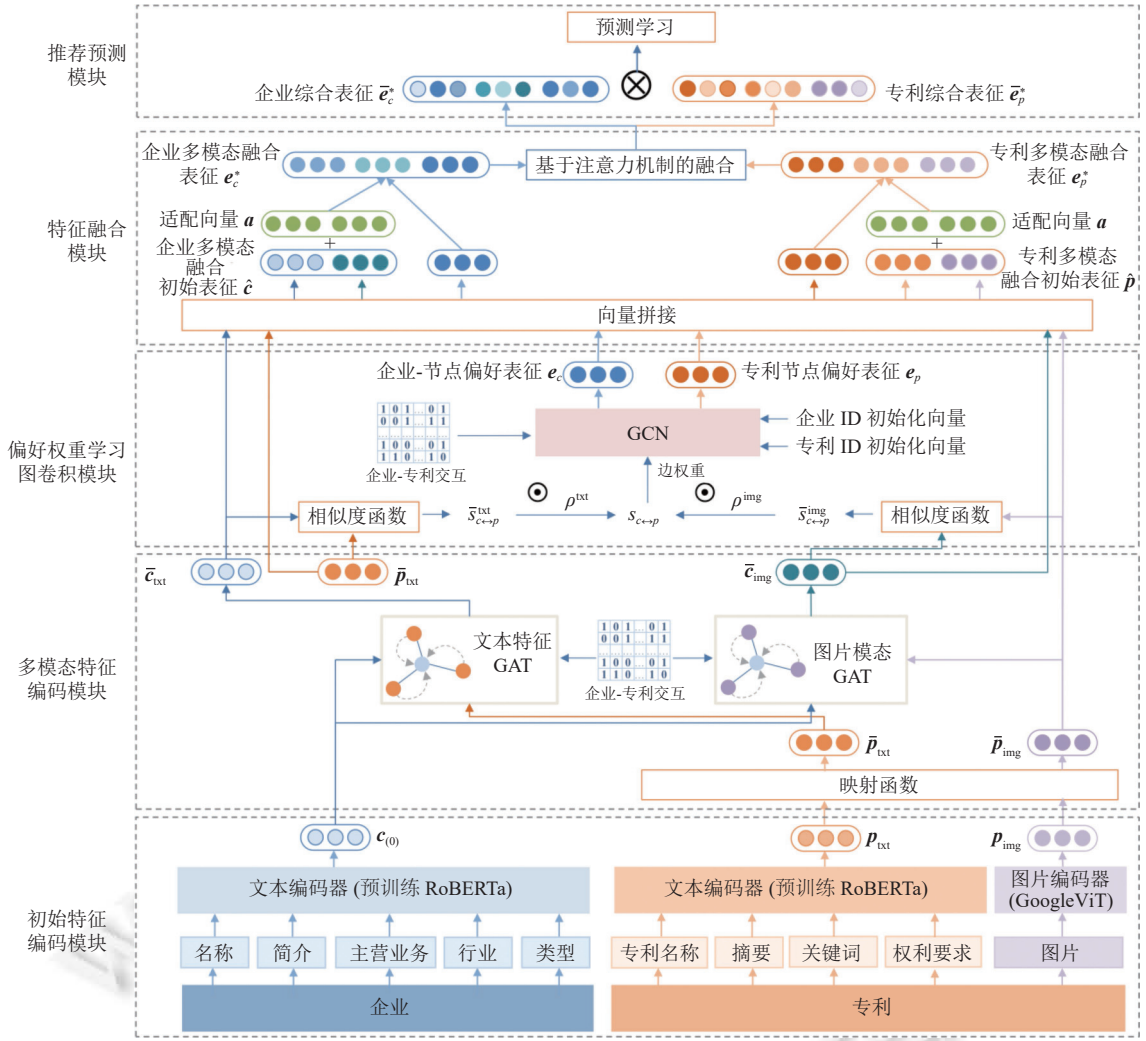


图2 基于多模态异质图网络的专利推荐模型框架图

2.1 初始特征编码

2.1.1 企业初始特征编码

本文提出的模型旨在向企业推荐专利,企业的信息包括企业名称、企业描述、经营范围、企业类型、所属行业等多种类型的文本信息.对于每种类型的企业文本信息先进行预处理,再通过预训练的 RoBERTa 模型^[51]生成每个企业的文本表征,具体表示如下:

$$\mathbf{h}_i^{\text{txt}} = \text{RoBERTa}(T_i)_{[\text{cls}]} \quad (1)$$

其中, T_i 表示第 i 种类型的文本信息, $\mathbf{h}_i^{\text{txt}}$ 表示输出的第 i 种文本信息的表征, txt 代表文本特征, RoBERTa 是文本表征预训练模型, [cls] 代表 RoBERTa 模型输出中一个特殊的标记向量,该向量通常用来表示整个文本的表征.

由于不同类型的文本信息根据语义的丰富程度可能会生成不同维度的特征,这些特征可能位于不同的空间中.对于每个文本表征 $\mathbf{h}_i^{\text{txt}}$,本文使用多层感知机 (multilayer perceptron, MLP) 将其投影到同一维度中,以便后续的特征融合,过程如下:

$$\bar{\mathbf{h}}_i^{\text{txt}} = \mathbf{W}_i \mathbf{h}_i^{\text{txt}} + \mathbf{b}_i \quad (2)$$

其中, $\bar{h}_i^{\text{txt}}$ 表示单个类型的文本表征, W_i 和 b_i 分别表示可训练的投影矩阵和偏差项. 最后将所有类型的文本表征进行拼接, 形成企业初始文本表征 $c_{(0)}$, 过程如下所示:

$$c_{(0)} = \text{concat}[\bar{h}_1^{\text{txt}}, \bar{h}_2^{\text{txt}}, \dots, \bar{h}_n^{\text{txt}}] \quad (3)$$

2.1.2 专利初始特征编码

对于专利数据而言, 其技术方案通常含有丰富的文本信息和图像信息, 文本提供了详细的专利描述和解释, 而图像能够直观地展示专利技术方案的结构、流程和工作原理, 因此可以通过多模态融合的方法来进行专利的表征. 首先, 分别通过单独的预训练模型对专利文本和图像进行表征生成.

(1) 文本表征. 每篇专利的文本信息都可由标题、摘要、关键词、权利要求书等构成, 类似地, 按第 2.1.1 节中企业的编码方式, 首先通过预训练的 RoBERTa 模型得到专利的各个文本属性表征, 然后经过 MLP 和拼接操作得到专利的初始文本表征, 记为 p_{txt} .

(2) 图像表征. 专利的图像信息来自专利的摘要图或技术方案图, 通常含有技术方案的模型结构图和方法流程图. 在文本描述的基础上, 图像信息进一步突出了专利技术方案中各组成部分之间的功能关系和依赖性. 为了更好地进行专利的图像表征, 首先统一对专利图片进行标准化处理, 得到大小一致为 224×224 的图片 I , 接着, 采用了 Google 的 Vision Transformer (ViT) [52] 预训练模型来捕捉专利的图像特征, 再通过 MLP 将其投影到固定维度, 对于专利 P 对应的图像 I , 特征提取过程如下所示:

$$h_{\text{img}} = \text{ViT}(I) \quad (4)$$

$$p_{\text{img}} = W_{\text{img}} h_{\text{img}} + b_{\text{img}} \quad (5)$$

其中, ViT 表示预训练模型, h_{img} 表示 ViT 输出的图像表征. 最终得到了专利的图像初始表征 p_{img} .

采用 ViT 作为图像编码器是由于其能够有效地编码图片中的语义特征, 尤其在处理专利的结构图和设计图时, 能够更好地提取其中蕴含的结构信息. 相较于传统的卷积神经网络, ViT 可以更充分地捕捉专利的图片信息. 后续消融实验部分也进一步验证了 ViT 在图像特征提取中的优越性, 表明其对提取利用图像特征具有显著作用.

2.2 多模态特征编码

为了捕捉多模态的上下文依赖, 本文分别在不同模态下构造企业-专利交互图, 通过图注意力神经网络来学习企业的偏好表征.

根据上述初始特征编码可以得到专利的任意模态表征 $p_m \in \{p_{\text{txt}}, p_{\text{img}}\}$ 以及企业文本表征 $c_{(0)}$, 其中, p_{txt} 和 p_{img} 分别表示文本模态和图像模态表征. 首先, 将专利的原始表征投影到度量空间中, 以提取与企业偏好相关的专利表征, 其计算公式如下:

$$\bar{p}_m = \text{leaky_ReLU}(W_m p_m + b_m) \quad (6)$$

其中, $\text{leaky_ReLU}(\cdot)$ 、 $W_m \in \mathbb{R}^{D \times D_m}$ 和 $b_m \in \mathbb{R}^D$ 分别表示激活函数、可训练权重矩阵和偏置向量. p_m 是指模态 m 下的初始表征, D_m 表示原始模态表征的维度, 而 D 是经过投影处理的表征 $\bar{p}_m$ 的维度.

随后, 引入邻居聚合机制以更好地学习企业偏好. 给定一个企业 c , 通过分析其与邻居节点的相似度, 利用迭代操作逐步调整其表征. 在描述过程中, 为了简洁, 本文以单一模态 m 为例, 其他模态的处理方法相同. 在初始迭代中, 将企业初始文本表征 $c_{(0)}$ 作为初始化表征. 分别在每个模态下, 利用企业初始表征和与企业偏好相关的专利表征 $\bar{p}_m$ 计算其亲和度分数, 具体公式如下:

$$p_{c,p} = \frac{\exp(\bar{p}_m^T c_{(t-1)})}{\sum_{j \in \mathcal{N}(c)} \exp(\bar{J}_m^T c_{(t-1)})} \quad (7)$$

其中, $p_{c,p}$ 表示企业 c 与专利 p 之间的亲和度, $\mathcal{N}(c)$ 表示企业-专利交互图中企业 c 的邻居节点集合. 其中, 企业偏好通过合并企业与邻居专利节点的加权和进行更新, 迭代公式如下:

$$c_{(t)} = c_{(t-1)} + \sum_{p \in \mathcal{N}(c)} p_{c,p} \bar{p}_m \quad (8)$$

通过上述迭代过程, 可以获得企业在模态 m 下的偏好表征, 记 $\bar{c}_m = c_{(t)}$ ($t = 1, 2, \dots, T$) 为企业 c 在模态 m 下的

偏好表征, 具体地, 在文本模态下记为 $\bar{\mathbf{c}}_{\text{txt}}$, 在图像模态下记为 $\bar{\mathbf{c}}_{\text{img}}$.

2.3 偏好权重学习的图卷积

为了进一步度量企业与专利之间的偏好, 本文计算交互图中边的权重, 这些权重为不同模态下的节点之间的相似度分数. 然后, 将每条边在多个模态中的相似度分数进行汇总, 作为企业-专利交互图的边权重, 再通过 GCN 计算企业和专利的节点偏好表征.

首先, 分别计算每个模态中企业和专利的相似度, 具体公式如下:

$$\bar{s}_{c \leftarrow p}^m = \frac{\exp(\bar{\mathbf{c}}_m^T \bar{\mathbf{p}}_m)}{\sum_{j \in \mathcal{N}(c)} \exp(\bar{\mathbf{c}}_m^T \bar{\mathbf{j}}_m)} \quad (9)$$

$$\bar{s}_{p \leftarrow c}^m = \frac{\exp(\bar{\mathbf{p}}_m^T \bar{\mathbf{c}}_m)}{\sum_{l \in \mathcal{N}(p)} \exp(\bar{\mathbf{p}}_m^T \bar{\mathbf{l}}_m)} \quad (10)$$

其中, $\bar{s}_{c \leftarrow p}^m$ 和 $\bar{s}_{p \leftarrow c}^m$ 表示企业 c 与专利 p 在模态 m 中两个方向的相似度, $\mathcal{N}(c)$ 和 $\mathcal{N}(p)$ 分别表示企业和专利在图中的邻居节点集合, $\bar{\mathbf{j}}_m$ 和 $\bar{\mathbf{l}}_m$ 分别表示邻居节点集合 $\mathcal{N}(c)$ 和 $\mathcal{N}(p)$ 中的节点向量. 为了得到整体的边权重, 引入了权重向量对每个模态的相似度进行加权处理, 公式如下:

$$\rho = [\rho^{\text{txt}}, \rho^{\text{img}}] \quad (11)$$

其中, ρ 表示权重向量, 企业权重向量用于衡量企业对不同模态下专利的相对偏好, 专利的权重向量表示专利不同模态的特征对专利表征的重要性. 结合权重向量, 通过如下公式计算企业与专利之间的多模态融合相似度:

$$s_{c \leftarrow p} = \max(\rho_c^{\text{img}} \bar{s}_{c \leftarrow p}^{\text{img}}, \rho_c^{\text{txt}} \bar{s}_{c \leftarrow p}^{\text{txt}}) \quad (12)$$

$$s_{p \leftarrow c} = \max(\rho_p^{\text{img}} \bar{s}_{p \leftarrow c}^{\text{img}}, \rho_p^{\text{txt}} \bar{s}_{p \leftarrow c}^{\text{txt}}) \quad (13)$$

其中, $\max(\cdot)$ 操作表示在各个模态的相似度得分中, 选取最大值作为多模态下的相似度.

在得到企业和专利的多模态相似度得分后, 基于这些相似度分数构建一个企业和专利的带权交互图, 通过图卷积网络来学习节点偏好表征. 通过图卷积操作, 每个节点会聚合其邻居节点的表征, 同时考虑边上的权重信息. 经过卷积操作后, 得到每个节点的多层嵌入, 并将这些嵌入聚合作为企业和专利的节点偏好表征. 具体的计算过程如下:

$$\mathbf{e}_c^{(l)} = \sum_{p \in \mathcal{N}(c)} s_{c \leftarrow p} \mathbf{e}_p^{(l-1)} \quad (14)$$

$$\mathbf{e}_p^{(l)} = \sum_{c \in \mathcal{N}(p)} s_{p \leftarrow c} \mathbf{e}_c^{(l-1)} \quad (15)$$

其中, $\mathbf{e}_c^{(l)}$ 和 $\mathbf{e}_p^{(l)}$ 分别表示在第 l 层中企业和专利的表征, $\mathbf{e}_c^{(0)}$ 和 $\mathbf{e}_p^{(0)}$ 则是企业和专利的 ID 初始化向量.

最后, 将每一层的嵌入表征进行累加, 得到企业和专利的节点偏好表征:

$$\mathbf{e}_c = \sum_{l=0}^L \mathbf{e}_c^{(l)} \quad (16)$$

$$\mathbf{e}_p = \sum_{l=0}^L \mathbf{e}_p^{(l)} \quad (17)$$

2.4 特征融合

2.4.1 基于适配向量的特征融合

为了进一步增强企业和专利的表征能力, 本文引入可学习的适配向量 $\mathbf{a} \in \mathbb{R}^{2D \times 1}$, 并将其融入多模态表征中. 适配向量的引入主要作用于多模态特征的全局调节和对齐, 从而提升最终的表征质量.

具体而言, 首先分别对企业偏好表征 $\bar{\mathbf{c}}_{\text{img}}$ 、 $\bar{\mathbf{c}}_{\text{txt}}$ 和专利偏好表征 $\bar{\mathbf{p}}_{\text{img}}$ 、 $\bar{\mathbf{p}}_{\text{txt}}$ 进行跨模态拼接, 以形成企业和专利的多模态融合初始表征, 表示如下:

$$\hat{\mathbf{c}} = \bar{\mathbf{c}}_{\text{img}} \parallel \bar{\mathbf{c}}_{\text{txt}} \quad (18)$$

$$\hat{\mathbf{p}} = \bar{\mathbf{p}}_{\text{img}} \parallel \bar{\mathbf{p}}_{\text{txt}} \quad (19)$$

其中, $\hat{\mathbf{c}} \in \mathbb{R}^{2D \times 1}$ 和 $\hat{\mathbf{p}} \in \mathbb{R}^{2D \times 1}$ 表示拼接后的企业多模态融合初始表征和专利的多模态融合初始表征.

在此基础上, 本文引入适配向量 $\mathbf{a}$ 为每个企业和专利的多模态融合初始表征进行全局调整学习. 这种调整可以被视为对多模态特征空间的一致性调节, 提高不同模态的特征在同一空间内的协调性. 具体地, 对于一个拼接后的实体表征 $\hat{\mathbf{e}}_i \in \{\hat{\mathbf{c}}, \hat{\mathbf{p}}\}$, 本文分配了一个独立的可学习适配向量 $\mathbf{a}_i$, 添加到 $\hat{\mathbf{e}}_i$ 上, 称为多模态适配融合表征, 记为 $\tilde{\mathbf{e}}_i$:

$$\tilde{\mathbf{e}}_i = \hat{\mathbf{e}}_i + \mathbf{a}_i \quad (20)$$

为了避免由于适配向量数量过多而增大训练难度, 通过训练一组 k 个独立的基向量, 即 $\mathbf{v}_1^b, \mathbf{v}_2^b, \dots, \mathbf{v}_k^b$, 每个适配向量 $\mathbf{a}_i$ 都可以通过使用这组基向量的聚合来生成:

$$\mathbf{a}_i = \sum_{j=1}^k l_{i,j} \mathbf{v}_j^b \quad (21)$$

$$l_{i,j} = \frac{\exp(\mathbf{I}_j^T \hat{\mathbf{e}}_i)}{\sum_{h=1}^k \exp(\mathbf{I}_h^T \hat{\mathbf{e}}_i)} \quad (22)$$

其中, $\mathbf{I}_j \in \mathbb{R}^{2D \times 1}$ ($j = 1, 2, \dots, k$) 是 k 个可学习的线性投影向量. 通过这种方式, 获得了更加协调、一致的多模态适配融合表征 $\tilde{\mathbf{c}}$ 和 $\tilde{\mathbf{p}}$.

为了更好地捕获企业与专利的交互信息, 本文进一步将上述卷积得到的节点偏好表征和多模态适配融合表征连接起来, 并将连接后的企业和专利表征向量分别记为 $\mathbf{e}_c^*$ 和 $\mathbf{e}_p^*$, 形式上表示如下:

$$\mathbf{e}_c^* = \mathbf{e}_c \parallel \tilde{\mathbf{c}} \quad (23)$$

$$\mathbf{e}_p^* = \mathbf{e}_p \parallel \tilde{\mathbf{p}} \quad (24)$$

其中, $\mathbf{e}_c^* \in \mathbb{R}^{3D \times 1}$ 和 $\mathbf{e}_p^* \in \mathbb{R}^{3D \times 1}$ 分别表示企业多模态融合表征和专利多模态融合表征. 这种融合不仅保留了企业与专利之间的协作特征, 还通过适配向量实现了对多模态特征的全局调控, 有助于模型更好地推断用户与项目之间的交互关系.

2.4.2 基于注意力机制的特征融合

为了进一步增强企业和专利的表征, MHGN 引入了注意力机制, 用于捕获特征中的关键信息. 注意力机制通过计算各特征的重要性权重, 使模型能够更加聚焦那些对预测任务最为关键的特征, 从而提升表征的准确性和解释性.

具体而言, 在企业和专利的多模态融合表征 $\mathbf{e}_c^*$ 和 $\mathbf{e}_p^*$ 的基础上, 利用注意力机制进行特征加权. 为了简洁表示, 使用 $\mathbf{e}^*$ 代表企业和专利的多模态融合表征, 注意力机制的过程如下.

首先, 对输入的企业和专利表征进行线性变换和激活函数处理后, 生成中间特征矩阵 $\mathbf{M}_e$, 公式为:

$$\mathbf{M}_e = \sigma(\mathbf{W}_1 \mathbf{e}^* + \mathbf{b}_1) \quad (25)$$

其中, $\mathbf{W}_1 \in \mathbb{R}^{D' \times 3D}$ 表示可学习权重矩阵, $\mathbf{b}_1 \in \mathbb{R}^{D'}$ 是对应的偏置项, $\sigma(\cdot)$ 为激活函数, 例如 $\tanh$ 函数.

通过对特征矩阵进行进一步处理, 计算得到注意力权重矩阵 $\mathbf{A}_e$. 注意力权重反映了各特征在综合表征中的重要性, 具体计算过程如下:

$$\mathbf{A}_e = \mathbf{W}_2 \mathbf{M}_e + \mathbf{b}_2 \quad (26)$$

其中, $\mathbf{W}_2 \in \mathbb{R}^{3D \times D'}$ 和 $\mathbf{b}_2 \in \mathbb{R}^{3D}$ 分别表示一个可学习的权重矩阵和偏置项, 用于对特征矩阵进行线性变换.

计算得到的注意力权重 $\mathbf{A}_e$ 将用于对多模态融合表征 $\mathbf{e}^*$ 进行加权. 具体地, 将权重与多模态融合表征逐元素相乘, 得到加权后的企业和专利表征 $\tilde{\mathbf{e}}^*$, 公式为:

$$\tilde{\mathbf{e}}^* = \mathbf{A}_e \odot \mathbf{e}^* \quad (27)$$

其中, $\odot$ 表示逐元素乘法操作. 至此, 可以得到企业和专利的综合表征 $\bar{\mathbf{e}}_c^*$ 和 $\bar{\mathbf{e}}_p^*$.

2.5 专利推荐预测

最后, 为了预测企业与专利之间的匹配得分, 本文通过内积的方式得到最终的预测输出 $y_{c,p}$, 表示为:

$$y_{c,p} = \bar{\mathbf{e}}_c^{*\top} \bar{\mathbf{e}}_p^* \quad (28)$$

其中, $y_{c,p}$ 表示企业对目标专利的偏好, 得分越高, 代表企业和专利的匹配度越高, 即越倾向于向企业推荐该专利. 为了提高推荐质量和优化模型参数, 采用贝叶斯个性化排序 (Bayesian personalized ranking, BPR) 作为 MHGN 的损失函数, 它可以通过最大化企业对已知正样本和负样本之间的差异来提高推荐质量, 具体定义如下:

$$\mathcal{L}_{\text{BPR}} = - \sum_{(c, p^+, p^-) \in \mathcal{D}} \ln \sigma(y_{c,p^+} - y_{c,p^-}) + \lambda \Theta \quad (29)$$

其中, $(c, p^+, p^-) \in \mathcal{D}$ 代表一组包含企业 c 、正样本 p^+ 和负样本 p^- 的三元组集合, $\sigma(\cdot)$ 为激活函数, λ 是正则化权重, Θ 代表模型可训练参数. 为了清晰起见, 算法 1 总结了 MHGN 算法的主要过程.

算法 1. MHGN 的算法框架.

输入: 专利-企业交互 S , 企业文本信息, 专利文本信息, 专利图像信息, 训练次数 n_epochs ;

输出: 专利推荐结果 y .

1. 通过预训练模型 RoBERTa 和 ViT 分别初始化专利的文本表征和图像表征 $\mathbf{p}_{\text{txt}}$ 和 $\mathbf{p}_{\text{img}}$.
 2. 通过预训练模型 RoBERTa 初始化企业的文本表征 $\mathbf{c}_{(0)}$.
 3. 随机初始化模型参数 ϕ .
 4. **for** $epoch = 1$ to n_epochs **do**
 5. **for each** $batch \subseteq S$ **do**
 6. 通过 GAT 编码生成专利和企业的多模态表征 $\bar{\mathbf{p}}_m$ 和 $\bar{\mathbf{c}}_m$.
 7. 计算企业与专利的多模态表征相似度, 构建企业-专利交互图.
 8. 通过 GCN 学习企业和专利的节点偏好表征 $\mathbf{e}_c$ 和 $\mathbf{e}_p$.
 9. 通过公式 (18) 和 (19) 分别拼接企业和专利的多模态表征得到 $\hat{\mathbf{c}}$ 和 $\hat{\mathbf{p}}$.
 10. 训练 k 个独立基向量 $\mathbf{v}_i^b, i = 1, 2, \dots, k$, 得到适配向量 $\mathbf{a}$.
 11. 通过公式 (20) 生成企业和专利的多模态适配融合表征 $\tilde{\mathbf{c}}$ 和 $\tilde{\mathbf{p}}$.
 12. 通过公式 (23) 和 (24) 连接节点偏好表征和多模态适配融合表征得到企业和专利的多模态融合表征 $\mathbf{e}_c^*$ 和 $\mathbf{e}_p^*$.
 13. 通过注意力机制得到企业和专利的综合表征 $\bar{\mathbf{e}}_c^*$ 和 $\bar{\mathbf{e}}_p^*$.
 14. 预测企业和专利最终评分 $y_{c,p}$.
 15. 通过最小化 $\mathcal{L}_{\text{BPR}}$ 更新模型参数 ϕ .
 16. **end for**
 17. **end for**
-

3 实 验

3.1 数据集

为了确保实验数据的真实有效性, 本文收集了多个高校的专利转让信息, 并根据不同高校的专利转让数量构建了不同规模的数据集. 为便于区分, 数据集以其所包含的专利转让数量命名. 例如, 数据集“专利-666”包含 666 条专利转让记录, “专利-1446”包含 1446 条, “专利-3911”包含 3911 条, “专利-24168”则涵盖了 24168 条专利转让数据. 数据集的详细信息如表 2 所示.

表 2 实验数据集统计信息

数据集名称	企业数量	专利数量	交互数	密度 (%)
专利-666	240	666	333	0.208
专利-1446	265	1 446	725	0.189
专利-3911	539	3 911	1 955	0.093
专利-24168	3 017	24 168	12 084	0.017

本文从专利原文中提取了文本和图像信息,并对已授权的专利,收集了其对应转让企业的信息.为了进一步提取出有效信息,本文对专利数据和企业数据进行了预处理,提炼出专利数据中所包含的专利号、摘要、关键字、权利要求书等文本信息作为文本属性,将专利的摘要图作为预处理图像.另外,本文从企业查询信息平台中获取企业信息,并抽取企业信息中的企业名称、经营范围、简介、行业归属和标签等字段作为企业的属性,根据第 2 节方法中的预训练模型对专利的多模态属性和企业属性进行特征提取.

3.2 对比方法

为了充分验证所提出模型的效果,本文选取了以下几种先进的多模态推荐模型进行实验对比.

(1) DualGNN^[44]: 从用户-项目二部图中构建一个额外的用户-用户相关图,并利用它来融合相关图中邻居的用户表征.

(2) LATTICE^[46]: 该方法着重于项目的多模态特征,对每个模态分别构造项目-项目图,通过融合项目-项目图的多模态特征来挖掘项目之间的潜在结构.接着对项目-项目图 and 用户-项目交互图进行图卷积操作,以学习用户和项目表征.

(3) MGCN^[53]: 该方法首先利用商品行为信息进行去噪以提高特征区分度,同时设计了行为感知融合器来自适应学习不同模态特征的相对重要性,可以全面建模用户偏好.

(4) BM3^[54]: 该方法采用自监督的多模态推荐算法,通过利用简单的随机删除增强方法从用户和商品的表征中提升潜在的对比视图,然后重构用户-商品交互图并调整模态特征,在跨模态和单模态视角下学习用户和商品的表征.

(5) FREEDOM^[47]: 该方法是对 LATTICE 的改进,它认为 LATTICE 的潜在图结构学习是无效的和非必要的.因此冻结了项目-项目图,同时对多模态推荐中的用户-项目交互图进行去噪,增强了用户和项目的表征.

(6) DRAGON^[49]: 该方法基于经常交互的项目构建用户-用户图,并从项目多模态特征构建项目-项目图.然后,利用图学习的用户-项目异构图(二部图)和同构图(用户-用户和项目-项目),以获得用户和项目的双重表征.

(7) GRCN^[45]: 该方法通过改进的 GCN 网络自适应地修正用户-项目交互图,剪枝潜在的误报的边,从而能更好地利用图卷积网络提取有意义的用户偏好信号,提高推荐性能.

(8) MG^[55]: 该方法是一种面向多模态推荐系统的优化方法,采用了一种称为“镜像梯度 (mirror gradient)”的优化策略,在优化过程中通过梯度策略增强推荐模型的鲁棒性,减轻来自多模态信息输入的不稳定影响.

3.3 实验设置

在本文中,所提出方法的实验运行在一个 80 GB 的 NVIDIA A100 GPU 上,使用 PyTorch 框架实现本文所提模型.对于不同的专利数据集,为获得最优的实验效果,会赋予不同的参数设置,学习率在 {0.0001, 0.001, 0.01} 范围内调整,批大小 batch_size 在 {16, 64, 128, 256} 范围内调整,随机失活值 dropout 在 {0.2, 0.3, 0.4, 0.5, 0.6, 0.7} 范围内调整.对于所有基线方法,为获得最佳推荐结果,按照文献的建议和实验配置调整最佳参数.

此外,所有对比模型的多模态特征均由预训练模型得到,为了公平地比较模型的实验性能,本文遵循各模型所用的预训练方法,对专利和企业数据特征提取,其中 DualGNN 的文本预训练模型采用 Sentence2Vector^[56],图像预训练模型采用 ResNet50^[57],其余基线模型均采用文本预训练模型 Sentence-BERT^[58]和图像预训练模型 Caffe^[59].为了保持一致性,本文将所有方法的文本特征和图像特征维度统一设置为 1024.

3.4 评价指标

实验中采用 5 个常用的分类指标来评估本文的实验性能, 即 ROC 曲线下方面积 (AUC)、准确率 (ACC)、精确率 (Precision)、召回率 (Recall)、*F1* 分数 (*F1*-score), 对于这些指标, 更高的指标值代表推荐性能越好.

3.5 对比实验

本实验对比了多种推荐模型在 4 个不同规模专利转让数据集上的表现. 实验结果如表 3 所示, 其中最优秀的结果用加粗字体表示, 第 2 优秀的结果用下划线表示. 可以看出本文提出的模型在所有数据集上均表现出了显著的优势, 在各项指标上均超越了其他对比模型.

表 3 不同方法在 4 个专利数据集上的平均性能

数据集	指标	DualGNN ^[44] (2021)	LATTICE ^[46] (2021)	MGCN ^[53] (2023)	BM3 ^[54] (2023)	FREEDOM ^[47] (2023)	DRAGON ^[49] (2023)	GRCN ^[45] (2020)	MG ^[55] (2024)	MHGN	Improve (%)
专利-666	AUC	0.5061	<u>0.6099</u>	0.5077	0.5423	0.5725	0.4524	0.5401	0.5643	0.7328	20.15
	ACC	0.4925	0.5373	0.5075	0.5448	0.5597	0.4701	0.5590	<u>0.5909</u>	0.6736	14.00
	Precision	0.4938	0.5362	0.5070	0.5357	0.5588	0.4714	0.5029	<u>0.6061</u>	0.6913	14.06
	Recall	0.597	0.5522	0.5373	<u>0.6716</u>	0.5672	0.4925	0.4462	0.5882	0.7309	8.83
	<i>F1</i> -score	0.5405	0.5441	0.5217	0.5960	0.5630	0.4818	0.5762	<u>0.5970</u>	0.6558	9.85
专利-1446	AUC	0.4968	<u>0.8014</u>	0.4969	0.5036	0.7933	0.5160	0.7033	0.4966	0.9438	17.77
	ACC	0.5034	0.6678	0.5000	0.4966	0.6747	0.5137	<u>0.6912</u>	0.5137	0.8676	25.52
	Precision	0.5032	0.6369	0.5000	0.4974	0.6425	0.5149	<u>0.7273</u>	0.4795	0.8800	21.00
	Recall	0.5411	0.7808	0.5685	0.6575	0.7877	0.4726	<u>0.8235</u>	0.5147	0.9706	17.86
	<i>F1</i> -score	0.5215	0.7015	0.5321	0.5664	<u>0.7077</u>	0.4929	0.6512	0.4965	0.8049	13.73
专利-3911	AUC	0.5011	<u>0.8008</u>	0.7561	0.4991	0.7831	0.5508	0.5764	0.5260	0.8125	1.46
	ACC	0.5000	<u>0.6633</u>	0.6297	0.5012	0.6559	0.5461	0.5833	0.5000	0.7917	19.36
	Precision	0.5000	<u>0.6323</u>	0.6028	0.5012	0.6263	0.5457	0.6154	0.5289	0.8333	31.79
	Recall	0.5337	<u>0.7805</u>	0.7606	0.5362	0.7731	0.5511	0.6667	0.5000	0.8000	2.50
	<i>F1</i> -score	0.5163	<u>0.6987</u>	0.6725	0.5181	0.6920	0.5484	0.5714	0.5141	0.7692	10.10
专利-24168	AUC	0.5054	<u>0.8201</u>	0.6883	0.5000	0.8144	0.5110	0.6562	0.4839	0.8681	5.85
	ACC	0.5058	<u>0.6636</u>	0.6307	0.5000	<u>0.6636</u>	0.5116	0.6250	0.4859	0.8333	25.57
	Precision	0.5055	0.6239	0.6006	0.5000	<u>0.6251</u>	0.5115	0.6087	0.4445	0.8462	35.37
	Recall	0.5318	<u>0.8238</u>	0.7804	0.5856	0.8172	0.5136	0.5833	0.4847	0.9167	11.28
	<i>F1</i> -score	0.5183	<u>0.7100</u>	0.6788	0.5394	0.7084	0.5126	0.6364	0.4637	0.7857	10.66

具体来说, 在 4 个不同数据集上, 相比于第 2 优秀的方法, 本文方法在准确率和精确率上分别提升了 14.00%、25.52%、19.36%、25.57% 和 14.06%、21.00%、31.79%、35.37%. 这些提升幅度不仅验证了本文方法的有效性, 也表明其在复杂的专利转让推荐场景中具有更强的泛化能力和稳定性.

MG 聚焦于增强多模态推荐系统的鲁棒性, 在稀疏交互及复杂文本推荐场景下, 在规模较小的数据集上性能仅次于 MHGN, 但随着数据集规模的扩大, 推荐性能则表现一般. DualGNN 和 GRCN 将专利的 ID 嵌入和多模态特征进行混合, 没有区分项目多模态特征的重要性, 这种方式可能会导致不同模态特征对推荐结果的影响无法得到有效控制, 从而在噪声较大的特征上表现不佳. BM3 采用自监督的方式在一定程度上缓解了多模态融合的噪声问题, 但其缺乏对不同模态特征的精细控制, 可能会导致模型在处理复杂多模态特征时表现不稳定. LATTICE 和 FREEDOM 通过一个潜在的项目-项目图增强专利的表征, 学习多模态特征之间的权值, 在一定程度上提升了模型的表征能力.

本文的改进可能归因于引入适配向量的多模态特征增强与融合, 并结合基于注意力机制的特征加权策略. 适配向量不仅提升了模型的表征能力, 使得模型能够更好地捕捉多模态专利特征之间的复杂关系, 还能在实际部署中减少模型的复杂性, 减少模型对大规模训练数据的依赖. 而基于注意力机制的加权策略则允许模型根据企业和专利的具体情况动态调整不同模态特征的重要性, 从而实现更加精准和个性化的推荐. 本文提出的方法在多模态

特征处理和个性化推荐方面展现了强大的优势,并且在实际应用中具有很高的可行性和适用性.它不仅提升了推荐系统的性能,还为复杂场景下的专利转让推荐提供了一种有效的解决方案.

为了进一步分析 MHGN 的推理性能,本文对 MHGN 和各个基线方法完成一个轮次训练的时间进行对比实验,具体结果如表 4 所示.可以看到,推荐性能最接近 MHGN 的基线方法 LATTICE 所需的推理时间最长;GRCN 虽然推理花费的时间最短,但 GRCN 整体推荐性能明显落后于 MHGN. MGCN、FREEDOM、DRAGON 使用的推理时间与 MHGN 最为接近,但推荐性能上均落后于 MHGN,可以进一步看出本文模型在专利推荐场景中的优越性.

表 4 MHGN 及基线模型的推理时间比较 (ms)

模型	专利-666	专利-1446	专利-3911	专利-24168
DualGNN ^[44]	184.731	195.367	201.553	193.551
LATTICE ^[46]	264.807	284.198	787.043	7216.047
MGCN ^[53]	26.697	21.763	23.869	21.589
BM3 ^[54]	102.173	81.853	80.007	74.624
FREEDOM ^[47]	27.768	28.045	27.943	27.514
DRAGON ^[49]	26.352	26.329	26.386	27.472
GRCN ^[45]	9.857	6.667	16.765	50.003
MG ^[55]	132.922	119.79	122.253	109.883
MHGN	26.483	27.201	17.332	108.444

3.6 消融实验

本节主要介绍 MHGN 的消融实验情况,分别研究适配向量、注意力机制、图像编码器、文本模态、图像模态、多模态融合方式等模块组件在模型中的作用.为此,本文构建了 9 个变体模型: MHGN-c1 去掉了适配向量,以评估适配向量在多模态特征融合中的作用; MHGN-c2 去掉了注意力机制,旨在分析注意力机制对模型性能的贡献; MHGN-c3 则同时去掉了适配向量和注意力机制,以探讨这两者在模型中的共同效果; MHGN-c4 将 ViT 图片编码器替换为 ResNet,以评估 ViT 编码器对专利图片的适用性; MHGN-c5 去掉了文本模态的表征,分析文本模态数据对模型的作用,类似地, MHGN-c6 去掉了图像模态数据,分析图像模态数据的重要性.为了分析文本和图像的特征融合方式对 MHGN 模型的影响,设计了 3 种不同融合方式的变体,其中, MHGN-c7、MHGN-c8、MHGN-c9 分别为向量相加、门控融合和注意力机制融合方式.

从表 5 的实验结果可以看出,各变体的性能均较完整模型有所下降,验证了适配向量和注意力机制对模型效果的关键性作用.具体来说,当去掉适配向量时 (MHGN-c1),模型在所有数据集上的 AUC 和 ACC 指标均显著下降,这表明适配向量在减少多模态特征噪声、增强特征融合方面发挥了重要作用.同样,当去掉注意力机制时 (MHGN-c2),模型的整体表现也有所下降,说明注意力机制在突出重要特征、提高推荐准确性方面具有明显优势.当同时去掉适配向量和注意力机制时 (MHGN-c3),模型的性能同样都显著降低.综合来看,适配向量和注意力机制的引入有效地提升了多模态特征的学习能力和模型的整体表现,在专利转让推荐任务中发挥了关键作用.最后,将 ViT 替换为 ResNet 后 (MHGN-c4),模型在 4 个数据集上的性能均显著下降.专利图片通常包含复杂的图形和文本信息,这些信息的长距离依赖特性可能对 ViT 的自注意力机制更为适合.这体现了 ViT 在专利图片处理上表现出的显著优势,因此在专利相关任务中使用 ViT 作为图片编码器更为有效.

当去掉图像模态数据,仅保留文本单模态数据及表征 (MHGN-c5) 时,可以看到推荐性能下降非常明显,在数据集“专利-3911”中性能几乎下降了一半,说明图像数据表征在 MHGN 中具有重要作用.类似地,当去掉文本数据,仅保留图像单模态数据及表征 (MHGN-c6) 时,整体推荐性能有所下降,但下降幅度略优于 MHGN-c5,并没有出现 MHGN-c5 的情况.此外,在文本表征与图像表征的融合方式上, MHGN 采用的拼接方法具有较好的性能.

表 5 比较 MHGN 及其变体的性能结果

变体	专利-666		专利-1446		专利-3911		专利-24168	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
MHGN	0.7328	0.6736	0.9438	0.8676	0.8125	0.7917	0.8681	0.8333
MHGN-c1	0.5302	0.5182	0.8406	0.7500	0.7431	0.7500	0.5625	0.5417
MHGN-c2	0.6675	0.6085	0.9234	0.7206	0.7153	0.625	0.7292	0.7917
MHGN-c3	0.5694	0.6007	0.8713	0.8382	0.6667	0.7083	0.6007	0.5833
MHGN-c4	0.6229	0.5425	0.9332	0.8309	0.7639	0.7083	0.8271	0.7674
MHGN-c5	0.5694	0.6250	0.6663	0.6397	0.4514	0.4167	0.8557	0.7446
MHGN-c6	0.5833	0.6250	0.6539	0.6103	0.6250	0.5833	0.8642	0.8089
MHGN-c7	0.5486	0.5417	0.7485	0.6765	0.5278	0.5000	0.8220	0.7501
MHGN-c8	0.6319	0.5833	0.6447	0.5956	0.5799	0.4583	0.8625	0.8126
MHGN-c9	0.6285	0.6250	0.6601	0.6618	0.4201	0.4167	0.8478	0.7169

3.7 参数分析

对于提出的 MHGN 方法, 主要调整两个参数, 即适配基向量的数量 k 和注意力机制的隐藏层维度 D' . 本节主要进行参数分析, 以探讨这两个参数对各个数据集的影响.

图 3 展示了不同数量的基向量在 4 个数据集上的 AUC、ACC 和 $F1$ -score 的表现. 从图中可以看出, 不同数据集对基向量数量的敏感度存在显著差异. 在较小的数据集 (专利-666 和专利-1446) 中, 基向量数量对模型性能的影响相对较小, 随着基向量数量的增加, 性能表现保持相对稳定, AUC、ACC 和 $F1$ -score 曲线波动较小. 对于较大的数据集 (专利-3911 和专利-24168), 基向量数量对模型性能的影响更加显著. 特别是, 当基向量数量增加到一定程度时, 模型的 AUC、ACC 和 $F1$ -score 都出现了较大的波动, 表明在这些数据集上, 基向量数量的调整对模型的表现有更大的影响. 此外, 不同数据集的最佳基向量数量各不相同. 这表明在实际应用中, 基向量数量需要根据具体数据集的特性进行调整, 以达到最佳效果.

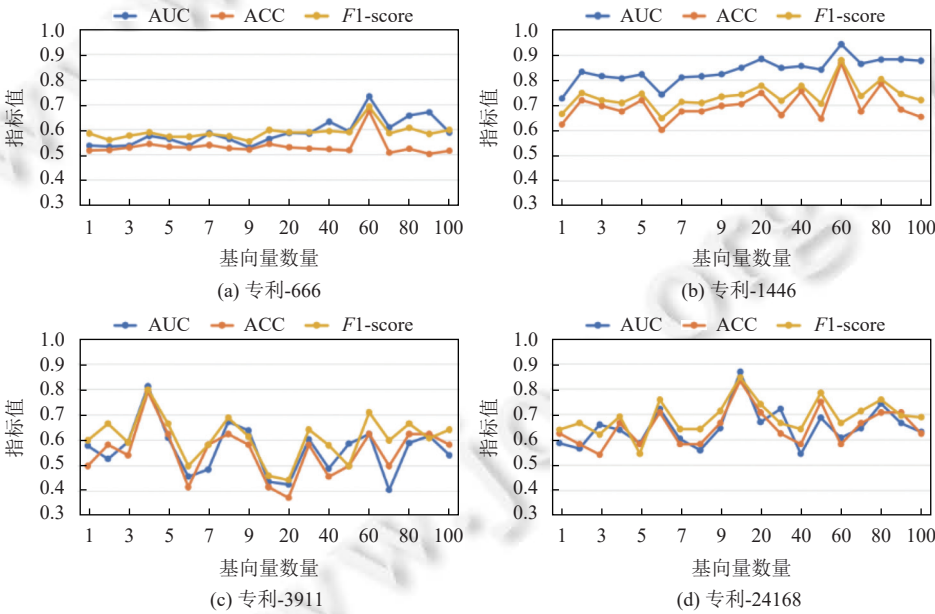


图 3 在 4 个数据集上关于基向量数量的参数分析

本文进一步分析了注意力机制中隐藏层维度对不同数据集的影响, 如图 4 所示. 实验结果表明, 不同的数据集在选择最优隐藏层维度时存在显著差异. 在较小规模的数据集 (专利-666 和专利-1446) 中, 隐藏层维度的增大通常会带来性能的提升, 但随着维度的进一步增加, 性能可能会趋于饱和甚至有所下降. 而在较大规模的数据集 (专利-

3911 和专利-24168) 中, 最佳的隐藏层维度相对较小. 这些实验结果显示了注意力机制的隐藏层维度需要根据数据集的规模和特性进行合理选择, 从而实现性能的最优化.

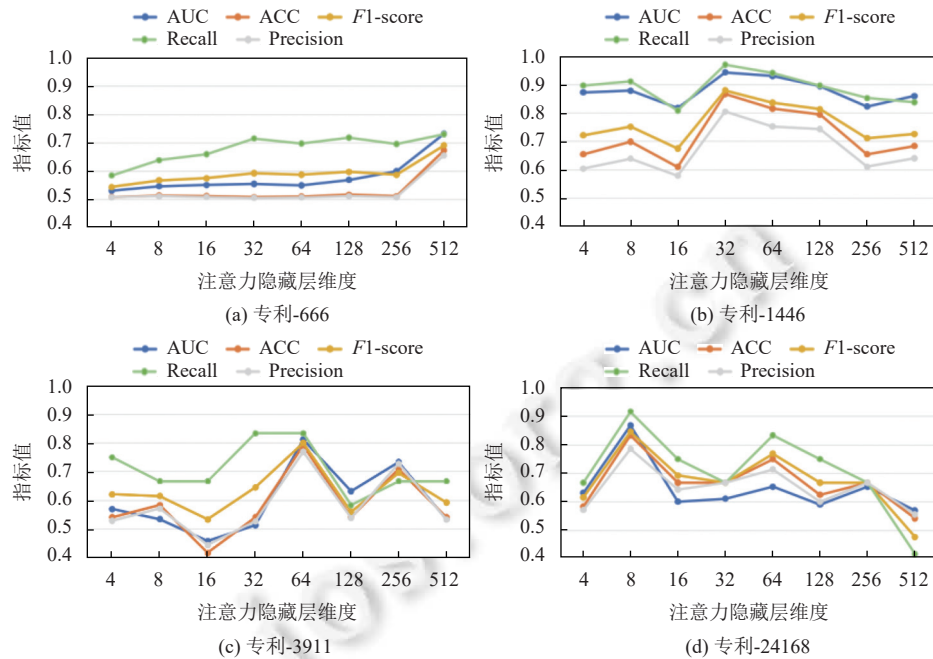


图4 在4个数据集上关于注意力隐藏层维度的参数分析

4 总结与展望

本文提出了一种基于多模态异质图网络的专利转让推荐方法, 通过结合专利的多模态信息 (文本和图像) 以及企业信息, 构建了一个复杂的企业-专利交互图, 再引入图注意力网络来有效捕获和融合多模态特征, 生成企业偏好表征, 并进一步利用图卷积网络学习企业和专利的节点偏好表征. 此外, 本文还引入了适配向量和注意力机制, 以更好地融合与加权多模态特征. 最后, 本文构建了4个专利转让数据集并进行了广泛的实验, 验证了该方法在提升专利与企业需求匹配精度方面的显著效果, 显示出其在专利推荐领域的潜力. 本文的数据和代码将会公开, 便于相关研究者复用这些资源.

然而, 由于数据源的限制, 本文仅在中国的专利数据中开展了验证研究, 在未来的工作中, 还可以进一步研究在更为广泛数据场景中的性能. 此外, 由于实体的多模态表征经过了多层操作, 使得 MHGN 可解释性较弱. 探索引入多元化的方法体系, 例如结合社交推荐机制, 通过对企业信息的深度建模, 捕捉潜在的转让关联性与价值. 同时, 本文还将探索采用先进的大语言模型技术, 利用其强大的文本理解与生成能力, 为专利转让推荐提供更为丰富的建议. 总的来说, 本文为专利推荐领域提供了一个新的视角和有效的解决方案, 为未来的相关研究和应用提供了较好的研究基础.

References

- [1] Liu WD, Qiao WB, Wang Y, Xu Z. Patent transformation opportunity to realize patent value: Discussion about the conditions to be used or exchanged. *Information Processing & Management*, 2021, 58(4): 102582. [doi: [10.1016/j.ipm.2021.102582](https://doi.org/10.1016/j.ipm.2021.102582)]
- [2] Deyle HG, Grupp H. Commuters and the regional assignment of innovative activities: A methodological patent study of German districts. *Research Policy*, 2005, 34(2): 221–234. [doi: [10.1016/j.respol.2005.01.003](https://doi.org/10.1016/j.respol.2005.01.003)]
- [3] Hsu CY, Chen YC, Luo RC, Rau HH, Fan CT, Hsiao BS, Chiu HW. A resource-sharing platform for trading biomedical intellectual property. *IT Professional*, 2010, 12(2): 42–49. [doi: [10.1109/MITP.2010.48](https://doi.org/10.1109/MITP.2010.48)]

- [4] Galasso A, Schankerman M, Serrano CJ. Trading and enforcing patent rights. *The RAND Journal of Economics*, 2013, 44(2): 275–312. [doi: [10.1111/1756-2171.12020](https://doi.org/10.1111/1756-2171.12020)]
- [5] Chen J, Chen JL, Zhao S, Zhang YP, Tang J. Exploiting word embedding for heterogeneous topic model towards patent recommendation. *Scientometrics*, 2020, 125(3): 2091–2108. [doi: [10.1007/s11192-020-03666-4](https://doi.org/10.1007/s11192-020-03666-4)]
- [6] Ye L, Zhang T, Cao XZ, Hu SL, Zeng G. Mapping the landscape of university technology flows in China using patent assignment data. *Humanities and Social Sciences Communications*, 2024, 11(1): 473. [doi: [10.1057/s41599-024-02982-x](https://doi.org/10.1057/s41599-024-02982-x)]
- [7] Muscio A. What drives the university use of technology transfer offices? Evidence from Italy. *The Journal of Technology Transfer*, 2010, 35(2): 181–202. [doi: [10.1007/s10961-009-9121-7](https://doi.org/10.1007/s10961-009-9121-7)]
- [8] Trappey AJC, Trappey CV, Wu CY, Fan CY, Lin YL. Intelligent patent recommendation system for innovative design collaboration. *Journal of Network and Computer Applications*, 2013, 36(6): 1441–1450. [doi: [10.1016/j.jnca.2013.02.035](https://doi.org/10.1016/j.jnca.2013.02.035)]
- [9] Ji X, Gu XJ, Dai F, Chen JX, Le CY. Patent collaborative filtering recommendation approach based on patent similarity. In: *Proc. of the 8th Int'l Conf. on Fuzzy Systems and Knowledge Discovery (FSKD)*. Shanghai: IEEE, 2011. 1699–1703. [doi: [10.1109/FSKD.2011.6019821](https://doi.org/10.1109/FSKD.2011.6019821)]
- [10] Yoon J, Seo W, Coh BY, Song I, Lee JM. Identifying product opportunities using collaborative filtering-based patent analysis. *Computers & Industrial Engineering*, 2017, 107: 376–387. [doi: [10.1016/j.cie.2016.04.009](https://doi.org/10.1016/j.cie.2016.04.009)]
- [11] Trappey A, Trappey CV, Hsieh A. An intelligent patent recommender adopting machine learning approach for natural language processing: A case study for smart machinery technology mining. *Technological Forecasting and Social Change*, 2021, 164: 120511. [doi: [10.1016/j.techfore.2020.120511](https://doi.org/10.1016/j.techfore.2020.120511)]
- [12] Rui XH, Min D. HIM-PRS: A patent recommendation system based on hierarchical index-based MapReduce framework. In: *Proc. of the Advances in Computer Science and Ubiquitous Computing: CSA-CUTE 2016*. Singapore: Springer, 2017. 843–848. [doi: [10.1007/978-981-10-3023-9_130](https://doi.org/10.1007/978-981-10-3023-9_130)]
- [13] Liu XJ, Wan YY, Liu XB, Zhang JH. A Patent recommendation algorithm based on topic classification and semantic similarity. In: *Proc. of the 2021 Int'l Conf. on Wireless Communications and Smart Grid (ICWCSG)*. Hangzhou: IEEE, 2021. 289–292. [doi: [10.1109/ICWCSG53609.2021.00063](https://doi.org/10.1109/ICWCSG53609.2021.00063)]
- [14] Wang Q, Du W, Ma J, Liao XW. Recommendation mechanism for patent trading empowered by heterogeneous information networks. *Int'l Journal of Electronic Commerce*, 2019, 23(2): 147–178. [doi: [10.1080/10864415.2018.1564549](https://doi.org/10.1080/10864415.2018.1564549)]
- [15] Du W, Wang YB, Xu W, Ma J. A personalized recommendation system for high-quality patent trading by leveraging hybrid patent analysis. *Scientometrics*, 2021, 126(12): 9369–9391. [doi: [10.1007/s11192-021-04180-x](https://doi.org/10.1007/s11192-021-04180-x)]
- [16] Oh S, Lei Z, Lee WC, Mitra P, Yen J. CV-PCR: A context-guided value-driven framework for patent citation recommendation. In: *Proc. of the 22nd ACM Int'l Conf. on Information & Knowledge Management*. San Francisco: ACM, 2013. 2291–2296. [doi: [10.1145/2505515.2505659](https://doi.org/10.1145/2505515.2505659)]
- [17] Deng WW, Ma J. A knowledge graph approach for recommending patents to companies. *Electronic Commerce Research*, 2022, 22(4): 1435–1466. [doi: [10.1007/s10660-021-09471-2](https://doi.org/10.1007/s10660-021-09471-2)]
- [18] Lee J, Sohn SY. Recommendation system for technology convergence opportunities based on self-supervised representation learning. *Scientometrics*, 2021, 126(1): 1–25. [doi: [10.1007/s11192-020-03731-y](https://doi.org/10.1007/s11192-020-03731-y)]
- [19] Chen YW, Zhu PH, Ma J, Huang XM, Qin J. A trust-enhanced patent recommendation approach to university-industry technology transfer. *ACM SIGMIS Database: The Database for Advances in Information Systems*, 2024, 55(1): 35–55. [doi: [10.1145/3645057.3645061](https://doi.org/10.1145/3645057.3645061)]
- [20] Liu ZB, Zhang YX, Deng WW, Ma J, Fan X. A deep learning method for recommending university patents to industrial clusters by common technological needs mining. *Scientometrics*, 2024, 129(6): 3089–3113. [doi: [10.1007/s11192-024-05052-w](https://doi.org/10.1007/s11192-024-05052-w)]
- [21] Du W, Jiang GR, Xu W, Ma J. Sequential patent trading recommendation using knowledge-aware attentional bidirectional long short-term memory network (KBiLSTM). *Journal of Information Science*, 2023, 49(3): 814–830. [doi: [10.1177/01655515211023937](https://doi.org/10.1177/01655515211023937)]
- [22] Liu Q, He YY, Lian DF, Zheng Z, Xu T, Liu C, Chen EH. UniMEL: A unified framework for multimodal entity linking with large language models. *arXiv:2407.16160*, 2024.
- [23] Chen BY, Huang L, Wang CD, Jing LP. Explicit and implicit feedback based collaborative filtering algorithm. *Ruan Jian Xue Bao/Journal of Software*, 2020, 31(3): 794–805 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5897.htm> [doi: [10.13328/j.cnki.jos.005897](https://doi.org/10.13328/j.cnki.jos.005897)]
- [24] Wang K, Wang Y, Liu JY, Deng JZ. Graph convolutional collaborative filtering recommendation algorithm based on Rényi differential privacy. *Ruan Jian Xue Bao/Journal of Software*, 2025, 36(3): 1202–1217 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/7165.htm> [doi: [10.13328/j.cnki.jos.007165](https://doi.org/10.13328/j.cnki.jos.007165)]

- [25] Oh S, Lei Z, Lee WC, Yen J. Recommending missing citations for newly granted patents. In: Proc. of the 2014 Int'l Conf. on Data Science and Advanced Analytics (DSAA). Shanghai: IEEE, 2014. 442–448. [doi: [10.1109/DSAA.2014.7058110](https://doi.org/10.1109/DSAA.2014.7058110)]
- [26] Trappey AJC, Trappey CV, Wu CY, Fan CY, Lin YL. Intelligent recommendation methodology and system for patent search. In: Proc. of the 16th IEEE Int'l Conf. on Computer Supported Cooperative Work in Design (CSCWD). Wuhan: IEEE, 2012. 172–178. [doi: [10.1109/CSCWD.2012.6221815](https://doi.org/10.1109/CSCWD.2012.6221815)]
- [27] Mahdabi P, Crestani F. Query-driven mining of citation networks for patent citation retrieval and recommendation. In: Proc. of the 23rd ACM Int'l Conf. on Information and Knowledge Management. Shanghai: ACM, 2014. 1659–1668. [doi: [10.1145/2661829.2661899](https://doi.org/10.1145/2661829.2661899)]
- [28] He Q, Spangler WS, He B, Chen Y, Kato L. Prospective client driven technology recommendation. In: Proc. of the 2012 Annual SRII Global Conf. San Jose: IEEE, 2012. 110–119. [doi: [10.1109/SRII.2012.23](https://doi.org/10.1109/SRII.2012.23)]
- [29] Fu TY, Lei Z, Lee WC. Patent citation recommendation for examiners. In: Proc. of the 2015 IEEE Int'l Conf. on Data Mining. Atlantic City: IEEE, 2015. 751–756. [doi: [10.1109/ICDM.2015.151](https://doi.org/10.1109/ICDM.2015.151)]
- [30] Chen YL, Chiu YT. An IPC-based vector space model for patent retrieval. Information Processing & Management, 2011, 47(3): 309–322. [doi: [10.1016/j.ipm.2010.06.001](https://doi.org/10.1016/j.ipm.2010.06.001)]
- [31] Krestel R, Smyth P. Recommending patents based on latent topics. In: Proc. of the 7th ACM Conf. on Recommender Systems. Hong Kong: ACM, 2013. 395–398. [doi: [10.1145/2507157.2507232](https://doi.org/10.1145/2507157.2507232)]
- [32] Liu SH, Liao HL, Pi SM, Hu JW. Development of a patent retrieval and analysis platform—A hybrid approach. Expert Systems with Applications, 2011, 38(6): 7864–7868. [doi: [10.1016/j.eswa.2010.12.114](https://doi.org/10.1016/j.eswa.2010.12.114)]
- [33] Huang ZH, Zhang JW, Tian CQ, Sun SL, Xiang Y. Survey on learning-to-rank based recommendation algorithms. Ruan Jian Xue Bao/Journal of Software, 2016, 27(3): 691–713 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4948.htm> [doi: [10.13328/j.cnki.jos.004948](https://doi.org/10.13328/j.cnki.jos.004948)]
- [34] Yang H, Wu X, Qiu ZP, Zheng YF, Chen X. Distributional fairness-aware recommendation. ACM Trans. on Information Systems, 2024, 42(5): 131. [doi: [10.1145/3652854](https://doi.org/10.1145/3652854)]
- [35] He RN, McAuley J. VBPR: Visual Bayesian personalized ranking from implicit feedback. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence. Phoenix: AAAI press, 2016. 144–150. [doi: [10.1609/aaai.v30i1.9973](https://doi.org/10.1609/aaai.v30i1.9973)]
- [36] Liu Q, Wu S, Wang L. DeepStyle: Learning user preferences for visual recommendation. In: Proc. of the 40th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Shinjuku: ACM, 2017. 841–844. [doi: [10.1145/3077136.3080658](https://doi.org/10.1145/3077136.3080658)]
- [37] Chen X, Chen HX, Xu HT, Zhang YF, Cao YX, Qin Z, Zha HY. Personalized fashion recommendation with visual explanations based on multimodal attention network: Towards visually explainable recommendation. In: Proc. of the 42nd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Paris: ACM, 2019. 765–774. [doi: [10.1145/3331184.3331254](https://doi.org/10.1145/3331184.3331254)]
- [38] Liu F, Cheng ZY, Sun CC, Wang YL, Nie LQ, Kankanhalli M. User diverse preference modeling by multimodal attentive metric learning. In: Proc. of the 27th ACM Int'l Conf. on Multimedia. Nice: ACM, 2019. 1526–1534. [doi: [10.1145/3343031.3350953](https://doi.org/10.1145/3343031.3350953)]
- [39] Zhao DD, Xu H, Peng SY, Zhou JW. Privacy-preserving graph neural network recommendation system based on negative database. Ruan Jian Xue Bao/Journal of Software, 2024, 35(8): 3698–3720 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/7124.htm> [doi: [10.13328/j.cnki.jos.007124](https://doi.org/10.13328/j.cnki.jos.007124)]
- [40] Gao C, Zheng Y, Wang WJ, Feng FL, He XN, Li Y. Causal inference in recommender systems: A survey and future directions. ACM Trans. on Information Systems, 2024, 42(4): 88. [doi: [10.1145/3639048](https://doi.org/10.1145/3639048)]
- [41] Wu SW, Sun F, Zhang WT, Xie X, Cui B. Graph neural networks in recommender systems: A survey. ACM Computing Surveys, 2022, 55(5): 97. [doi: [10.1145/3535101](https://doi.org/10.1145/3535101)]
- [42] Xu BB, Cen KT, Huang JJ, Shen HW, Cheng XQ. A survey on graph convolutional neural network. Chinese Journal of Computers, 2020, 43(5): 755–780 (in Chinese with English abstract). [doi: [10.11897/SP.J.1016.2020.00755](https://doi.org/10.11897/SP.J.1016.2020.00755)]
- [43] Wei YW, Wang X, Nie LQ, He XN, Hong RC, Chua TS. MMGCN: Multi-modal graph convolution network for personalized recommendation of micro-video. In: Proc. of the 27th ACM Int'l Conf. on Multimedia. Nice: ACM, 2019. 1437–1445. [doi: [10.1145/3343031.3351034](https://doi.org/10.1145/3343031.3351034)]
- [44] Wang QF, Wei YW, Yin JH, Wu JL, Song XM, Nie LQ. DualGNN: Dual graph neural network for multimedia recommendation. IEEE Trans. on Multimedia, 2023, 25: 1074–1084. [doi: [10.1109/TMM.2021.3138298](https://doi.org/10.1109/TMM.2021.3138298)]
- [45] Wei YW, Wang X, Nie LQ, He XN, Chua TS. Graph-refined convolutional network for multimedia recommendation with implicit feedback. In: Proc. of the 28th ACM Int'l Conf. on Multimedia. Seattle: ACM, 2020. 3541–3549. [doi: [10.1145/3394171.3413556](https://doi.org/10.1145/3394171.3413556)]
- [46] Zhang JH, Zhu YQ, Liu Q, Wu S, Wang SH, Wang L. Mining latent structures for multimedia recommendation. In: Proc. of the 29th ACM Int'l Conf. on Multimedia. ACM, 2021. 3872–3880. [doi: [10.1145/3474085.3475259](https://doi.org/10.1145/3474085.3475259)]
- [47] Zhou X, Shen ZQ. A tale of two graphs: Freezing and denoising graph structures for multimodal recommendation. In: Proc. of the 31st

- ACM Int'l Conf. on Multimedia. Ottawa: ACM, 2023. 935–943. [doi: [10.1145/3581783.3611943](https://doi.org/10.1145/3581783.3611943)]
- [48] Hu HC, Guo W, Liu Y, Kan MY. Adaptive multi-modalities fusion in sequential recommendation systems. In: Proc. of the 32nd ACM Int'l Conf. on Information and Knowledge Management. Birmingham: ACM, 2023. 843–853. [doi: [10.1145/3583780.3614775](https://doi.org/10.1145/3583780.3614775)]
- [49] Zhou HY, Zhou X, Zhang LZ, Shen ZQ. Enhancing dyadic relations with homogeneous graphs for multimodal recommendation. In: Proc. of the 26th European Conf. on Artificial Intelligence. Nice: IOS Press, 2023. 3123–3130. [doi: [10.3233/FAIA230631](https://doi.org/10.3233/FAIA230631)]
- [50] Chen YK, Sun TH, Ma YH, Zou HH. Multifactorial modality fusion network for multimodal recommendation. Applied Intelligence, 2025, 55(2): 139. [doi: [10.1007/S10489-024-06038-0](https://doi.org/10.1007/S10489-024-06038-0)]
- [51] Liu YH, Ott M, Goyal N, Du JF, Joshi M, Chen DQ, Levy O, Lewis M, Zettlemoyer L, Stoyanov V. RoBERTa: A robustly optimized BERT pretraining approach. arXiv:1907.11692, 2019.
- [52] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai XH, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929, 2021.
- [53] Yu PH, Tan ZY, Lu GM, Bao BK. Multi-view graph convolutional network for multimedia recommendation. In: Proc. of the 31st ACM Int'l Conf. on Multimedia. Ottawa: ACM, 2023. 6576–6585. [doi: [10.1145/3581783.3613915](https://doi.org/10.1145/3581783.3613915)]
- [54] Zhou X, Zhou HY, Liu Y, Zeng ZW, Miao CY, Wang PW, You Y, Jiang FJ. Bootstrap latent representations for multi-modal recommendation. In: Proc. of the 2023 ACM Web Conf. Austin: ACM, 2023. 845–854. [doi: [10.1145/3543507.3583251](https://doi.org/10.1145/3543507.3583251)]
- [55] Zhong SS, Huang ZZ, Li DF, Wen WS, Qin JH, Lin L. Mirror gradient: Towards robust multimodal recommender systems via exploring flat local minima. In: Proc. of the 2024 ACM Web Conf. Singapore: ACM, 2024. 3700–3711. [doi: [10.1145/3589334.3645553](https://doi.org/10.1145/3589334.3645553)]
- [56] Arora S, Liang YY, Ma TY. A simple but tough-to-beat baseline for sentence embeddings. In: Proc. of the 2017 Int'l Conf. on Learning Representations. OpenReview.net, 2017.
- [57] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778. [doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90)]
- [58] Reimers N, Gurevych I. Sentence-BERT: Sentence embeddings using siamese BERT-networks. arXiv:1908.10084, 2019.
- [59] Jia YQ, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffè: Convolutional architecture for fast feature embedding. In: Proc. of the 22nd ACM Int'l Conf. on Multimedia. Orlando: ACM, 2014. 675–678. [doi: [10.1145/2647868.2654889](https://doi.org/10.1145/2647868.2654889)]

附中文参考文献

- [23] 陈碧毅, 黄玲, 王昌栋, 景丽萍. 融合显式反馈与隐式反馈的协同过滤推荐算法. 软件学报, 2020, 31(3): 794–805. <http://www.jos.org.cn/1000-9825/5897.htm> [doi: [10.13328/j.cnki.jos.005897](https://doi.org/10.13328/j.cnki.jos.005897)]
- [24] 王锬, 王永, 刘金源, 邓江洲. 基于 Rényi 差分隐私的图卷积协同过滤推荐算法. 软件学报, 2025, 36(3): 1202–1217. <http://www.jos.org.cn/1000-9825/7165.htm> [doi: [10.13328/j.cnki.jos.007165](https://doi.org/10.13328/j.cnki.jos.007165)]
- [33] 黄震华, 张佳雯, 田春岐, 孙圣力, 向阳. 基于排序学习的推荐算法研究综述. 软件学报, 2016, 27(3): 691–713. <http://www.jos.org.cn/1000-9825/4948.htm> [doi: [10.13328/j.cnki.jos.004948](https://doi.org/10.13328/j.cnki.jos.004948)]
- [39] 赵冬冬, 徐虎, 彭思芸, 周俊伟. 基于负数据库的隐私保护图神经网络推荐系统. 软件学报, 2024, 35(8): 3698–3720. <http://www.jos.org.cn/1000-9825/7124.htm> [doi: [10.13328/j.cnki.jos.007124](https://doi.org/10.13328/j.cnki.jos.007124)]
- [42] 徐冰冰, 岑科廷, 黄俊杰, 沈华伟, 程学旗. 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](https://doi.org/10.11897/SP.J.1016.2020.00755)]

作者简介

赖培源, 博士, 教授, CCF 专业会员, 主要研究领域为数据挖掘, 推荐算法, 知识图谱, 语义表征.

卢伊虹, 硕士生, CCF 学生会员, 主要研究领域为图机器学习, 推荐算法, 数据挖掘.

廖德章, 工程师, 主要研究领域为数据挖掘, 推荐算法.

王昌栋, 博士, 教授, CCF 杰出会员, 主要研究领域为数据聚类, 网络分析, 推荐算法.

戴青云, 博士, 教授, CCF 专业会员, 主要研究领域为知识产权大数据, 人工智能.

赖剑煌, 博士, 教授, CCF 杰出会员, 主要研究领域为机器学习, 数据挖掘.