

# 大数据视域下应用型本科双创社交网络与创赛关系研究方法初探

郑翔天<sup>1</sup> 陈昱嘉<sup>1</sup> 张斯淇<sup>2</sup> 袁一飞<sup>1</sup>

(1. 南京工程学院 计算机工程学院, 江苏 南京 211167;

2. 国网江苏南京供电公司, 江苏 南京 211000)

**摘要:** 本文旨在探讨如何利用多模态信息融合技术, 优化应用型本科高校双创社交网络的数据分析与科创关系构建。研究将系统梳理国内外相关文献, 构建多模态信息融合的理论模型与技术框架, 揭示其在科创关系构建中的作用机制。通过从高校双创社交网络、万方系统、校园一卡通等平台采集多模态数据, 进行综合分析 with 实证研究, 提出基于多模态信息融合的科创关系构建策略。本课题期望丰富社交网络分析的理论体系, 促进计算机科学与社会科学的交叉融合, 为应用型本科高校的双创教育提供理论支持和实践指导, 推动多模态信息融合技术在社交网络分析领域的应用与发展。

**关键词:** 双创社交网络; 科创关系; 多模态数据

## 一、引言

近年来, 创新创业(以下简称“双创”)教育逐渐成为应用型高校关注的焦点。然而, 尽管众多高校在推动双创教育方面做出了积极努力, 但在教学资源整合、学科规划、课程设置、教学实践变革、培养计划以及评估标准等方面仍存在诸多不足, 尚未实现全面覆盖。大数据视域下的在线社交网络, 因其个性化的在线社交服务和即时通讯功能, 获得了大量用户的青睐, 为双创教育提供了新的发展机遇。用户在网络上频繁互动, 构筑了丰富的社交通道, 在线社交网络通过精准分析用户行为, 提供个性化的社交推荐服务。

本文旨在探讨如何利用多模态信息融合技术的兴趣点推荐方法, 将应用型高校常用的在线社交网络 APP、校园一卡通等系统中的多模态数据进行有效融合和综合分析, 以揭示学生在校园学习生活中的行为模式, 并进一步构建与科创竞赛活动之间的关联。通过多模态信息融合技术, 优化在线课程设计, 提高教学效果, 并为学校的管理和决策提供支持。本课题的研究不仅具有重要的理论价值, 也为应用型本科高校的双创教育提供了实践指导, 有助于更好地培养创新创业人才, 推动高校双创教育的发展。

## 二、应用型双创社交网络特征

用户在网络上与他人频繁互动, 构筑线上平台的社交通道。在线社交网络为用户提供了信息公开的社交功能, 便于用户更精准、更广泛、更频繁地从事各类在线社交活动, 也有利于在线社交网络对用户开展精准分析, 提供个性化的社交推荐服务。校园双创社交网络符合在线社交网络系列特征, 该网络是一系列特征的集合, 能够反映在校师生的性格、素养、社会背景以及家庭成员、竞争对手、同事伙伴等等。

用户社交特征主要分为两类: 用户属性信息和用户间关系信息。用户属性信息描述用户个体的基本特征, 包括居住地、职业、学校等显性属性信息和用户兴趣等隐性属性信息。用户间关系信息则描述用户个体与个体之间社交关系的基本特征, 包括隐性社交关系和社交关系类型等内容。新浪微博平均日活跃用户数为 2.22 亿, 而月活跃用户数达到 5.16 亿; 微信和 WeChat 的合并月活跃账户数达到约 11.65 亿。在线社交网络和用户的紧密程度日益加深, 用户的社交行为已逐步从原有的真实社会延伸至虚拟在线空间与真实社会活动相结合的方式进行。多模态信息融合技术在社交网络分析领域的应用逐渐受到广泛关注。应用型本科高校作为培养创新创业人才的重要基地, 其双创社交网络中的多模态信息融合

与创赛关系构建研究具有重要的理论和实践价值。诸多学者前期利用文本挖掘、图像处理、语音识别等多种技术手段, 对社交网络中的多模态信息进行提取、分析和融合。在学习通等在线教育平台方面, 研究者们关注学生在上课签到、作业提交、请假申请等方面的行为数据, 以揭示学生的学习态度和习惯。这些数据的综合分析有助于更全面地评估学生的校园生活状态, 为创赛关系构建提供有力支持。一些研究者利用多模态信息融合技术来分析学生不同学习平台上的行为模式, 以优化在线课程设计和提高教学效果; 还有一些研究者利用该技术来挖掘校园生活中的潜在信息和模式, 为学校的管理和决策提供支持。但在在线社交网络在应用型高校的科创、竞赛活动中相关研究, 申请人少有见到类似报道。

在线社交网络未知社交特征信息包含用户基本信息、社交链接关系、社交言论内容、社交互动行为等多个方面, 受到不同身份人群的关注。例如, 对于在线社交网络服务商而言, 新注册用户短期内未填写个人身份及信息说明, 未与其他用户建立社交关系链接, 使得在线社交平台无法对该用户开展有效的信息推荐工作。这些信息对于在线社交网络服务商而言是隐藏的。对于其他获取在线社交网络信息的人群, 仅可通过网络 API、匿名采集或是网络窃取的方式获取在线社交网络数据。对于以上问题, 校内广泛使用的 APP 如学习通、易班和建立在 TIM 以及钉钉、飞书上的课程群及课题组群数据表达存在多种模态。APP 间多模态信息如何融合为科创竞赛有效支撑, 少有相关研究开展。在线社交网络未知社交特征信息在校园创赛领域未知性体现在学生虽然不存在身份隐匿问题但学生的创赛“兴趣点”社交特征信息依然未知。

## 三、研究方法

### (一) 基于协同过滤方法

协同过滤及其改进技术被广泛用于推荐系统中。协同过滤技术是一种利用具有共同兴趣或经历的来向用户推荐感兴趣内容的技术。协同过滤的基本思想就是若两名用户在相似的物品上具有相似的行为(如共同访问同一用户兴趣点), 那么这两名用户就可能在其他的物品也有相似的行为, 即访问其他相同的用户兴趣点。协同过滤技术通常可以分为两类: 基于内存的协同过滤和基于模型的协同过滤。基于内存的协同方法利用了整个用户-项目评分矩阵进行推荐, 已被大量商用推荐系统所采用, 具有高效、易实现的特点, 但却面临着数据稀疏性的问题。

基于模型的协同过滤方法根据训练数据学习相应的模型, 并

将该模型应用至具体的推荐任务,预测用户对不同项目的兴趣,能够适用于用户兴趣点推荐问题。Noulas等人发现矩阵分解的性能要比基于用户或基于项目协同过滤的性能差。AlSuwaidan等人采用兴趣相似性度量来挖掘具有相同兴趣的用户群体。在用户兴趣点推荐问题中,协作过滤技术也是一种常用的方法,受到大量研究人员的关注。Leung等人提出了一种协同地点推荐框架。框架通过考虑GPS的历史轨迹作为主要因素来发现用户可能出现的地理位置。Zhao等人提出了一种基于因式分解项相似度和贝叶斯个性化排名的模型,为用户推荐可能感兴趣的项目。Liu等人运用贝叶斯矩阵分解(BNMF)考虑地理位置影响和文本内容影响,提出了种地理概率因素分析框架CTBNMF。Ye等人提出了一种基于内存的协同过滤模型FCF,根据用户好友关系进行协作过滤推测可能的用户兴趣点。FCF模型受约束于用户关系的影响,仅在朋友中而不是网络内所有用户中搜索最相似的用户。

### (二) 基于社交关系推断方法

已有大量研究表明,用户间社交关系是用户兴趣点推荐中一个重要的方面,具有社交关系的好友相对于陌生人更有可能分享共同的兴趣爱好。Zhang等人将签到频率或评分转化为推荐得分进行推荐。Zhang等人提出了一种基于概率的方法来对用户的社会影响力进行建模,并根据此概率来对用户兴趣点进行推测。Jiang等人采用了一种基于星形结构图的混合随机游动方法,能够融合复杂的异构链接结构。但是,该方法仅考虑了亲密好友的影响,而本文会进一步考虑疏远好友对此方法进行改善。Manotumruksa等人提出了一个具有多抽样标准的个性化排名框架PRFMC(Personalised Ranking Framework with Multiple sampling Criteria),通过贝叶斯个性化排序生成用户兴趣点的排序列表。Li等人提出了一个两步的POI推荐框架,该框架首先根据用户朋友学习潜在的用户兴趣点位置,然后将潜在用户兴趣点合并到加权矩阵进行矩阵分解。Cheng等人采用带社交正则化的概率矩阵分解方法PMFSR,能够将社交影响力与概率矩阵分解方法相融合。Li等人提出了一种基于排名的模型RankGeoFM,该模型通过学习用户对用户兴趣点的偏好排名以降低数据稀疏性的影响。

### (三) 基于地理位置影响推测方法

地理位置影响是用户兴趣点推荐与传统商品推荐不同的重要因素。这是由于用户兴趣点与实际地理位置相关。在基于位置的社交网络中,用户天然地会与用户兴趣点在现实中进行交互,与传统推荐系统有着不同之处。地理位置上的邻近特性会极大影响用户对于用户兴趣点的网络签到行为,同时也能缩小候选用户兴趣点候选集合的规模,从而减轻数据稀疏性的影响。不少研究尝试利用地理位置影响来强化用户兴趣点推荐的性能,主要包括三种具有代表性的模型,即幂律分布模型、高斯分布模型和核密度估计模型。

Ye等人采用幂率分布模型来刻画地理位置的影响。幂率分布已被发现于多种人类活动中,包括ATM机上的取款活动或是在不同城市内的旅行活动。在后续的工作中,Ye等人观察到用户签到行为存在着相似的行为模式。Yuan等人通过不同的假设将地理位置影响整合到用户兴趣点推荐中。作者认为,用户从一个用户兴趣点移动到另一个用户兴趣点的可能性是两个用户兴趣点之间的距离函数Cho等人观察到在线社交网络用户通常会进行短距离的移动或是周期性的距离移动,而这些移动行为通常不会受到社交网络结构的影响,而长途旅行的行为会更易受到社交网络的关系影响。Chen等人采用矩阵分解的思想,提出了一种多中心高斯模型(Multi-center Gaussian Model, MGM)来刻画用户兴趣点的

地理位置影响,推测用户兴趣点的概率。Zhao等人提出了一种基于遗传的高斯混合模型刻画地理位置的影响,其性能结果要优于MGM模型。

第三种地理位置影响模型是核密度估计模型。Zhang等人引入了一种iGSLR框架,认为每名用户关于用户兴趣点地理位置的影响应当个性化处理,而不是服从一个统一的分布。为此,作者有效利用用户在个性化社交和地理位置的影响,采用核密度估计刻画每名用户关于用户兴趣点的位置分布。

### (四) 辅助信息推测方法

众多研究人员利用有关在线社交网络的各种信息来进行用户兴趣点推荐。这些信息包括用户内容、社交标签和用户兴趣点序列等。Liu等人采用用户兴趣和用户兴趣点序列来预测某个时间窗口内的用户兴趣点。Deng等人利用社交标签和双向交互来预测用户兴趣。Levandoski等人设计了一种对空间项目和非空间项目进行评分的新方法。Hu等人利用在线社交网络上的用户生成内容来构建空间话题模型,用来获取用户移动轨迹和用户兴趣点之间的相关性。Zhou等人通过在矩阵分解中融入社交网络结构和用户偏好信息,提出了一种基于社交关系和用户偏好的推荐方法。但该项研究的分析存在着一定的问题,作者仅考虑每个项目的评分,却未考虑每个用户感兴趣项目的评分。

## 四、结论

本文的研究具有重要的理论和实践价值,主要体现在以下几个方面:在理论上,本课题有助于丰富和完善应用型本科高校社交网络分析的理论体系,通过深入研究多模态信息融合的理论基础和技术方法,推动社交网络分析领域的发展,为后续研究提供理论支撑和参考;揭示应用型本科高校社交网络中的信息传播和关系演化的内在规律,为社交网络管理和优化提供理论依据;促进计算机科学、社会科学等多个学科的交叉融合,推动这些学科的交叉融合,为相关领域的研究提供新的思路和方法。在实践上,本课题的研究成果可以为应用型本科高校的双创教育提供有力支持,包括通过多模态信息融合技术准确分析学生在社交网络中的行为特征和科创需求,制定更加针对性的教育策略;为社交网络平台的运营和管理提供技术支持,精准推送内容和服务,提升用户体验和平台价值;帮助校内科创团队更好地利用社交网络中的多模态信息,深入了解社交网络和师生需求,促进科创关系网络的构建,推进团队工作和合作创新。

### 参考文献:

- [1] 汪霖. 高校创新创业教育对大学生创业意愿的影响[J]. 黑龙江科学, 2023, 14(17): 111-113.
- [2] 何巍. 社交网络舆情多模态知识图谱构建框架研究[J]. 情报杂志, 2024, 43(01): 160-166.
- [3] 李婷婷, 楚俊峰, 王燕燕. 基于隐式信任和群体共识的群体推荐方法[J]. 计算机应用, 2024, 44(02): 460-468.
- [4] 鲍彩倩. 社交媒体中基于用户关系扩展的信念网络信息推荐模型[D]. 2023.
- [5] 肖妮. 基于超星学习通的前厅与客房管理课程混合式教育模式改革与实践[J]. 现代商贸工业, 2024, 45(08): 195-197.

基金项目: 1. 南京工程学院教学建设项目 应用型线上线下混合式大数据课程教学实践(JXJS2023051) 2. 江苏省高等教育教改立项研究课题, 新工科背景下“互联网+”双创工程人才校企合作培养的研究与实践(2023JSJG489)