

基于大规模关联网络学习的稳健并行追踪方法

研究成果: Robust Parallel Pursuit for Large-Scale Association Network Learning

作者: 李文慧, 周信 (共同通讯作者), 董柄朋, 郑泽敏 (共同通讯作者)

发表期刊: *INFORMS Journal on Computing*, 2025, 37(2): 428-445.

稀疏降秩回归在揭示大规模响应与预测变量之间的关联网络中发挥着关键作用, 广泛应用于社交网络和推荐系统等实际场景。然而, 现有方法所恢复的关联网络往往对离群值较为敏感, 或者在大数据背景下缺乏可延展性。因此, 发展一种兼具鲁棒性与可延展性的稀疏降秩回归方法, 对于推动大规模响应-预测网络中结构识别与异常检测的理论研究与实践应用具有重要意义。

中国科学技术大学管理学院和科技商学院郑泽敏教授带领其团队成员李文慧, 周信和董柄朋在国际顶级学术期刊 *INFORMS Journal on Computing* 上发表的论文“Robust Parallel Pursuit for Large-Scale Association Network Learning”对上述问题进行了解答。

本文提出了一种新的统计学习方法——稳健并行追踪 (ROP, Robust Parallel Pursuit), 可在大规模响应-预测网络中实现联合参数估计与离群点检测。该方法通过因子分解技术, 将初始复杂的关联网络学习问题拆解为一系列稀疏单位秩估计的子问题, 进而实现高效的并行求解, 显著提升整体计算效率。此外, 通过将离群矩阵嵌入回归系数矩阵, 该方法将传统的估计问题与异常值识别有机整合至同一建模框架, 并借助并行追踪策略, 有效提升大规模数据分析的可延展性。

我们的方法可应用于社交网络、推荐系统等典型场景。具体而言:

一方面, 在社交网络中, 用户行为可能会受自身历史行为及其他用户行为影响。通过本方法得到的估计参数矩阵, 能够清晰反映用户交互的关联网络。基于恢复的参数矩阵, 可精准识别出那些对他人行为有显著影响的高影响力用户。营销人员与这些关键用户合作开展广告宣传, 便能借助用户的社交网络高效传播信息。不仅如此, 本方法还可依据已建立的模型, 对用户未来行为进行预测, 助力营销人员提前调整策略。

另一方面, 在推荐系统中, 用户对产品的偏好通常与其自身特征 (如社交特征) 紧密相关。本方法所估计的参数矩阵, 可量化用户特征对产品偏好的影响程度。这一量化结果为营销和产品管理带来两方面的积极作用: 其一, 营销人员可利用该信息, 为用户精准推荐其可能感兴趣的产品; 其二, 产品经理通过明确驱动用户偏好的关键特征, 在



新产品开发或现有产品优化过程中，能够精准聚焦核心要素，使产品更好地贴合用户需求，提升市场竞争力。

供稿：周信 编辑：梁鹏 审核：宋捷