

考虑用户隐性反馈行为的社交媒体网络数据挖掘推荐算法

王倩倩¹ 任行学¹

WANG Qianqian REN Xingxue

摘要

在社交媒体网络数据挖掘过程中, 动态社交环境所带来的兴趣漂移和数据稀疏性问题, 使得采用静态用户画像和固定推荐策略的方法难以适应环境的动态变化。由于无法有效捕捉用户的实时行为和偏好变化, 以及及时调整推荐策略, 导致数据挖掘推荐的精度下降。为此, 文章提出一种考虑用户隐性反馈行为的社交媒体网络数据挖掘推荐算法。该算法通过构建隐语义行为反馈矩阵同时捕捉显式反馈(点赞、评论)和隐性反馈(浏览时长、点击次数), 利用高斯分布和低秩隐语义分解从这些反馈中提取用户特征。基于提取的用户特征, 结合用户与所属区域的相似度特征及其与属性种群的关系, 测定用户偏好的全局相似度。将计算得到的用户偏好全局相似度与局部相似度相结合, 共同设定为数据挖掘的推荐目标。将全局相似度与实时隐性行为特征共同输入自适应匹配机制。通过相似度加权调整个性化兴趣度, 并利用特征向量映射动态更新推荐概率, 最终融合显、隐性行为生成精准推荐目标。实验结果表明: 所提方法得出的 RMSE 均控制在 0.40 以下, 显著提高了实际的网络数据挖掘推荐精度, 性能优越。

关键词

用户反馈; 隐形行为; 社交媒体; 网络数据; 数据挖掘; 推荐算法

doi: 10.3969/j.issn.1672-9528.2025.08.037

0 引言

互联网社交媒体平台作为当前信息的载体, 成为日常生活中不可或缺的一部分。用户在平台上分享生活点滴、交流思想情感, 生成了海量的数据。这些数据不仅记录了用户的显性行为, 还隐含了用户的停留时间、浏览路径、重复访问等隐性反馈, 反映了用户的兴趣和偏好, 针对其日常的需求, 给予用户较为真实的推荐。但由于隐性行为具有一定的动态性及不稳定性, 需要多周期捕捉分析, 再加上网络用户之间的关系错综复杂, 一定程度上也增加了数据挖掘推荐的难度。

针对以上问题, 相关领域的专业人员已将其进行体系化梳理。涂剑峰等人^[1]设计的冒充性攻击下无线传感节点信息数据安全推荐算法。通过计算冒充性攻击下无线传感网络邻居节点的信任度, 对备选节点推荐集展开数据去噪、冒充性攻击检测和重构处理。按照信任度等级给用户设置可映射的标签, 建立可躲避攻击节点模型, 运用模型展开多层级的定点推荐测算, 输出推荐结果。仿真结果表明, 所提算法在冒充性攻击下, 查全率高于 88%, 平均响应时间低于 390 ms, MAE 和 RMAE 均低于 1.09% 和 0.95%, 有效提高了无线传感网络通信质量和通信效率, 保证了传输安全性和可靠性。然而, 该算法主要基于静态的信任度计算和预设的

规则, 难以有效适应动态社交环境中的兴趣漂移和数据稀疏性问题, 导致推荐精度下降。丁雨辰等人^[2]融合稀疏隐式反馈数据, 构建双重去偏协同过滤推荐算法。使用双重逆倾向加权方法和对比学习的形式, 辅助任务去除输入双塔自编码器的隐式反馈数据, 根据数据分析对应的特征, 进而映射出用户实际的行为规律和偏好。实验结果显示, 该算法在公开无偏数据集上的 NDCG@K、MAP@K 和 Recall@K 指标优于对比算法。但双塔自编码器需要大量数据进行训练, 在动态社交环境中难以及时捕捉用户行为和偏好的快速变化, 影响推荐结果的准确性。何丹提出的融合时间戳信息的工业大数据 SVD 推荐模型^[3]。融合强化学习 Q-learning 优化技术, 把时间影响添加至 SVD 推荐模型, 预设多个数据处理环节, 通过时间戳对采集的数据整合处理, 提取特征后作为目标展开推荐测算。研究表明, 该模型在数据集推荐测评中全程耗时 18 h, 比分分布式处理节省近 10% 时间, 计算效率显著提升。然而, 预设的数据处理环节难以实时捕捉用户兴趣和行为的快速变化, 可能导致推荐结果过时。陈云云等人^[4]设计的基于挖掘算法的用户大数据周期智能推荐仿真方法。采用协同推荐算法对用户历史行为进行分析, 通过数据相似性衡量智能推荐的效果, 使用 Top-N 算法优化推荐过程, 达到周期智能推荐的目的。融合基于神经网络挖掘长时数据和短时数据, 通过引入灰色均衡算法计算相似度, 达到推荐目的, 缩短推荐时间。实验结果表明, 所提算法在相似度

1. 河南科技职业大学信息工程学院 河南周口 466000

计算准确度方面提升 7%，推荐精确度提升 6%，召回率提升 8%。但基于固定相似性计算模型的方法难以及时反映用户兴趣的动态变化。

针对上述问题，本研究结合用户隐性反馈行为设计一种社交媒体网络数据挖掘推荐算法，通过挖掘用户显性行为和隐性反馈数据构建用户画像，理解用户兴趣和需求，为用户提供个性化推荐服务。

1 设计媒体网络用户隐性反馈行为数据挖掘推荐算法

1.1 基于隐语义行为反馈矩阵的用户特征提取

在社交媒体网络数据挖掘推荐过程中，用户偏好既通过显式反馈（如点赞、评论等直接行为）表示，也隐含在隐性反馈（如浏览时长、点击频次等间接行为）中。显式反馈虽能直接反映用户偏好，但数据稀疏且获取困难；隐性反馈数据丰富但噪声较多。为更全面准确地挖掘用户潜在兴趣，本研究提出构建隐语义行为反馈矩阵^[5]。通过建立语义用户矩阵和项目矩阵，定义用户与项目间的动态反馈关系^[6]。用公式表示为：

$$p(R|U, v, \sigma_t) = \prod_{u=1} [N(R|U, v, \sigma_t) \times \omega u^2]^t \quad (1)$$

式中： $p(R|U, v, \sigma_t)$ 代表反馈条件； N 代表数据分布值； ω 代表正态分布； u 代表项目数量； t 代表低秩隐语义高斯分布。

基于该反馈约束条件，将用户行为数据转化为可计算指令，构建隐语义反馈行为集，并划分为显式和隐式行为两类以便后续匹配分析^[7]。其中隐式行为作为核心分析对象，需在高斯分布假设下进行特征提取与分解^[8]，其计算过程分别为：

$$D = \sum_{k=1} \eta k - \frac{M(R|U, v, \sigma_t) \times \wp}{u} \times \wp k \quad (2)$$

$$O = \xi u \times \left[\frac{\sqrt{\wp}}{\eta(k-1)} \right]^2 \quad (3)$$

式中： D 代表反馈处理； η 代表隐语义服从值； k 代表正态挖掘点； M 代表可识别范围； $\wp$ 代表损失值； $\wp$ 代表低秩隐语义； O 代表反馈处理结果； ξ 代表正则化项。

当前输出的反馈结果为用户隐语义下的基础特征行为，将其作为推荐引导目标展开分析。

1.2 基于全局-局部相似度融合的用户偏好深度挖掘

在社交媒体网络数据挖掘推荐过程中，完成隐语义行为反馈矩阵设计后，用户和项目的潜在特征（如兴趣、行为模式）已被映射到低维空间。然而，这仍不足以全面、精准地理解用户在整个社交媒体网络中的偏好结构以及与其他用户、项目的关联。而隐性行为的反馈能够进一步反映用户的隐藏需求，丰富推荐的内容，完善推荐结构，从而确定更为精准的推荐测算目标。

为了更深入地挖掘用户偏好，需要获取用户和所属区域之间的基本相似度信息，并测定用户与属性种群之间的关系，具体表示为：

$$u_q = \frac{\sum_{l=1} \delta l - (k+1)\zeta}{\sqrt{\xi}} \times \psi \quad (4)$$

式中： u_q 代表变化关系； δ 代表推荐范围； l 代表属性种群点隐藏语义； ζ 代表局部量化值； $\sqrt{\xi}$ 代表反馈异常行为数量； ψ 代表向量评分。

在用户与属性种群之间关系的变化过程中，进行局部相似度的计算，以局部映射全部，测定出用户偏好全局相似度，用公式分别表示为：

$$H = \sqrt{\xi(k+1)^2} - \sum_{m=1} \chi m - \varepsilon \quad (5)$$

$$L = \eta^2 v - \frac{(\lambda + \beta)}{\phi m} \quad (6)$$

式中： H 代表用户偏好局部相似度； χ 代表相似性语义； m 代表同步推荐目标； ε 代表预期设定相似度； L 代表用户偏好全局相似度； v 代表语义项目集群； λ 代表相邻语义； β 代表可匹配语义。

仅依赖局部相似度可能导致推荐结果过于碎片化，无法反映用户的全局偏好，因此将用户偏好局部相似度和全局相似度设定为数据挖掘推荐目标，综合隐性反馈行为，设计出可执行流程，如图 1 所示。

按照图 1 流程执行，将所反馈的隐性行为赋予用户偏好局部相似度和全局相似度，数据整合及挖掘过程中，对符合要求的数据筛选、存储，作为后期的匹配特征。这些结果将为推荐系统提供更精准、全面的用户画像，进而支持更优质的推荐服务。

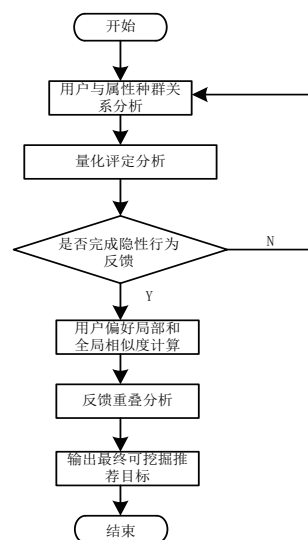


图 1 隐性行为反馈下用户偏好分析流程

1.3 基于动态自适应匹配的个性化推荐

在完成隐语义行为反馈矩阵设计和全局相似度计算后，用户偏好已被有效分解为局部兴趣（反映短期行为）和全局偏好（反映长期兴趣）。但鉴于社交媒体环境固有的动态特性，如兴趣漂移、数据稀疏性等问题普遍存在，传统推荐算法由于依赖于静态的用户画像和既定的推荐策略，难以精准捕捉用户的动态变化与个性化需求。为此，本研究提出动态自适应匹配机制，通过实时调整推荐条件来精准捕捉用户行为变化。

设定自适应的匹配条件：

$$J = \lg \left(\frac{\sum_{c=1} mb_{ij} c \times f}{|Z|} \right) \times \aleph b_{ij} \tag{7}$$

式中： J 代表自适应匹配条件； b_{ij} 代表个性化兴趣度； c 代表特征向量； f 代表特征映射点； $|Z|$ 代表隐性推荐概率； $\aleph$ 代表权重。

依托于设定的自适应匹配条件，以输出的偏好目标作为引导，利用推荐算法展开匹配推荐，具体表示为：

$$J_w = \int (\boldsymbol{T} \times b_{ij} \times \frac{\eta}{Q})^2 - |Z| f \tag{8}$$

式中： J_w 代表匹配推荐测算结果； $\boldsymbol{T}$ 代表隐藏特征目标向量； Q 代表信任度。

通过持续优化特征向量映射和信任度评估，实现了显性行为与隐性反馈的深度融合，最终输出精准的个性化推荐目标，有效提升了推荐系统在动态环境中的适应能力。

2 实验

为验证用户隐性反馈行为的社交媒体网络数据挖掘推荐算法的实际应用效果，采用对比实验的方式进行验证。为确保测试结果的准确性，选取两种现有方法作为对照组，即冒充性攻击下无线传感节点信息安全推荐算法和稀疏隐式反馈数据的双重去偏协同过滤推荐算法。根据测试需求，部署数据集和测试环境。

2.1 数据集与测试环境构建

针对测试算法的有效性和针对性，利用仿真软件 Geph 在真实的社交网络数据集上进行实验。为保证实验的可靠性，选择三组可操作的实验数据集，具体如表 1 所示。表 1 设定的数据集满足测试要求，且包含大量社交用户数据，涉及的用户种类较多。鉴于当前网络数据大多具有动态性，需选择在社交网络中有过 10 次签到以上的活跃用户，剔除签到少于 10 次的僵尸用户。这种筛选方式有助于提取活跃用户的属性，展开实验对比。完成数据集的设定之后，需要对测试的环境进行部署。明确基础的控制条件，确保数据挖掘推荐实验过程中的稳定及可控，具体如表 2 所示。

表 1 实验数据集

数据集	用户数量	项目个数	社交网络记录个数	稀疏度
Imagenet	1 635	2 001	1 855	0.012 5
Filmtrust	1 855	2 351	1 416	0.016 2
Ciao	1 574	1 687	1 354	0.035 4

表 2 数据挖掘推荐实验指标及参数设置

数据挖掘推荐指标	参数标准值
挖掘预测偏差	1.12
目标挖掘耗时 /s	0.12~0.16
R^2 决定系数	11.57
挖掘数据识别覆盖率 /%	98.57~99.55
学习率	0.01
权重	11.2

以表 2 设定的指标为引导，核验当前平台的设备及控制环境是否处于稳定的状态，确保无异常之后，建立联合控制机制后，展开具体的核验及分析。

2.2 媒体网络数据自适应挖掘推荐结果

考虑到隐性反馈行为，从 Imagenet 数据集中随机获取多组数据，并提取可推荐的目标。将这些目标导入实验平台后，采用自适应挖掘推荐的方式输出结果，具体如图 2 所示。

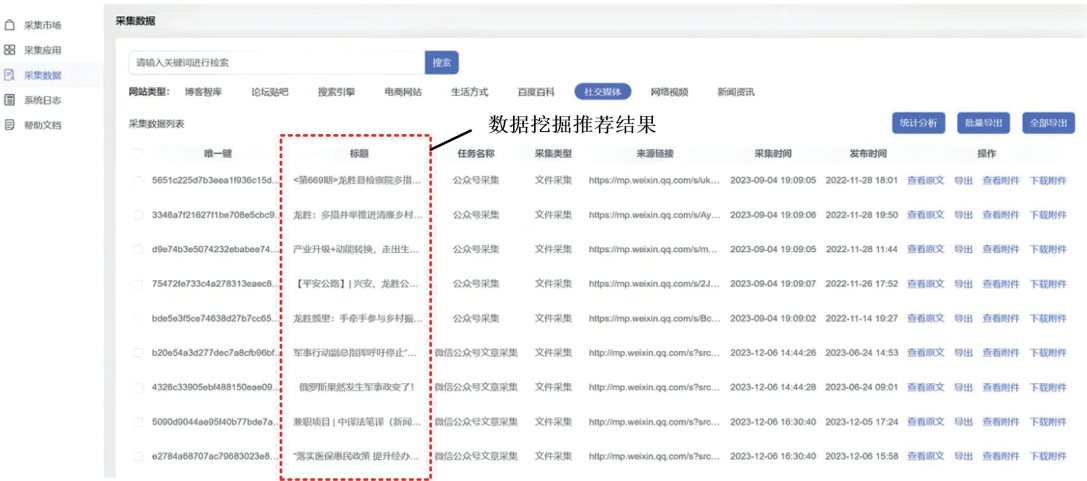


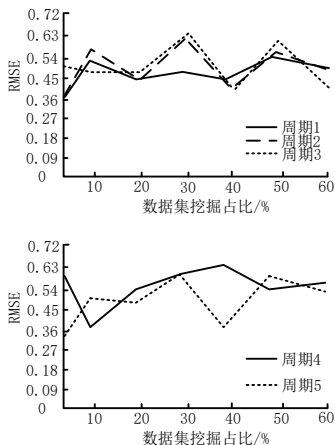
图 2 网络数据自适应挖掘推荐结果

由图 2 结果可知，由于社交媒体网络数据动态化和灵活性等特点，依据提取的特征目标进行数据挖掘推荐，在输出页面观察到，通过层级挖掘和针对性推荐计算，可以精准地输出最佳推荐结果。同时，还为所推荐的数据赋予位置、来源、链接等信息，提供更为完整、具体的推荐结果。该结果反映了所提算法的灵活性与稳定性，在海量数据的高压处理需求下，对提取到的目标展开实时匹配和自适应挖掘处理，在挖掘的过程中不断同步实时更新的推荐目标，确保输出的结果为最佳推荐结果，具有较强的推荐精准度，

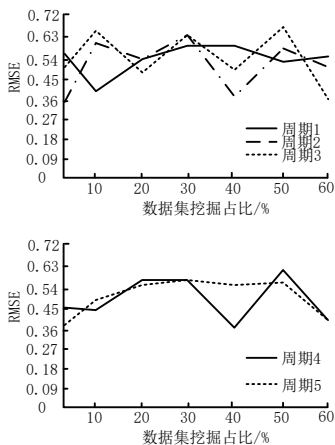
应用效果有所提升。

2.3 数据挖掘推荐算法性能结果对比分析

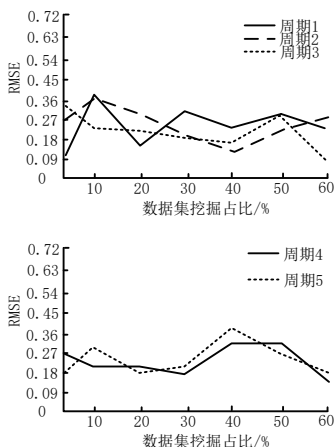
在 Filmtrust 和 Ciao 两个数据集中随机选取多组可用的社交媒体网络数据, 设定 5 个周期进行重复测定。基于隐性反馈行为, 对选定的 3 种算法展开对比, 以验证数据挖掘推荐的性能结果, 如图 3 所示。



(1) 稀疏隐式反馈数据的双重去偏协同过滤推荐算法



(2) 冒充性攻击无线传感节点信息数据安全推荐算法



(3) 用户隐性反馈行为媒体网络数据挖掘推荐算法

图 3 数据挖掘推荐 RMSE 结果

从图 3 的结果可以看出: 所提算法得出的 RMSE 均控制在 0.40 以下, 而文献中选定的算法得出的 RMSE 则在 0.35~0.68 之间, 相对较高。表明所提算法通过隐性行为反馈, 能够在海量数据中更有效地进行挖掘推荐及识别, 以用户偏好全局相似度为推荐匹配标准。通过自适应对比增强了推荐结果的可靠性, 提高了实际的推荐精度, 性能表现优越。

3 结语

结合用户隐性反馈行为, 对社交媒体网络数据挖掘推荐算法进行了验证与分析。鉴于网络数据动态可变的特性, 在高斯分布框架下设定隐性行为反馈条件, 构建反馈矩阵以提取隐语义行为特征。基于这些特征, 深入探究用户与行为种群间的关联, 依据行为偏好变化精准计算用户偏好全局相似度, 从而更有效地映射隐性特征并精准筛选推荐目标。运用自适应匹配处理方法, 将个性化兴趣度与筛选后的推荐目标相融合, 完成数据挖掘推荐测算。本研究为社交媒体数据的处理、整合以及个性化挖掘与推荐提供了全新的思路与方法。

参考文献:

- [1] 涂剑峰, 李芳丽. 冒充性攻击下无线传感节点信息数据安全全推荐算法[J]. 传感技术学报, 2024, 37(8): 1434-1440.
- [2] 丁雨辰, 徐建军, 崔文泉. 适用于稀疏隐式反馈数据的双重去偏协同过滤推荐算法[J]. 计算机系统应用, 2024, 33(8): 145-154.
- [3] 何丹. 融合时间戳信息的工业大数据 SVD 推荐模型建立[J]. 现代工业经济和信息化, 2025, 15(1): 11-12.
- [4] 陈云云, 刘永山. 基于挖掘算法的用户大数据周期智能推荐仿真[J]. 计算机仿真, 2024, 41(4): 456-460.
- [5] 王彬彬. 基于数据挖掘技术的智能仪器数据推荐算法[J]. 成都工业学院学报, 2024, 27(2): 41-46.
- [6] 祝现威, 刘伟, 刘自豪, 等. 基于知识图谱的网络安全事件数据推荐算法[J]. 网络与信息安全学报, 2023, 9(6): 116-126.
- [7] 伍乙生. 基于网络流量感知的个性化推荐算法研究[J]. 长江信息通信, 2023, 36(9): 13-15.
- [8] 侯浩天. 基于数据挖掘的大数据信息推荐算法研究[J]. 信息与电脑(理论版), 2023, 35(16): 76-78.

【作者简介】

王倩倩(1993—), 女, 河南周口人, 硕士, 助教, 研究方向: 智能信息处理。

任行学(1992—), 男, 河南信阳人, 硕士, 讲师, 研究方向: 人工智能。

(收稿日期: 2025-04-01 修回日期: 2025-08-05)