

社会网络视角下民族用户群体行为的识别机制研究

张敏超^{1,2}, 蒲秋梅^{1,2*}

¹中央民族大学, 民族语言智能分析与安全治理教育部重点实验室, 北京

²中央民族大学信息工程学院, 北京

收稿日期: 2023年9月28日; 录用日期: 2023年11月3日; 发布日期: 2023年11月15日

摘要

大数据时代, 社交网络与人们的生活融合在一起, 全民参与网络生活的程度不断提高, 社交平台逐步成为政府与公众进行沟通的重要阵地。社交网络的迅速发展, 带来了大量的流量与商业价值, 同时也吸引了许多通过不正当行为从中渔利的异常用户。如何正确处理网络舆情事件, 精准识别异常用户行为, 减少其带来的不良影响已成为政府工作的一大重要课题。本文通过研究民族地区网络群体用户异常行为, 分析多用户交互行为的特征, 利用大数据范式理论与机器学习算法研究相结合的方法, 通过网络行为数据进行实时计算, 可以实现民族地区热点事件的动态感知, 进而对公众行为的发展趋势进行有效地监测。有助于提高社会稳定性, 维护社交网络环境, 促进民族团结。

关键词

社交网络, 网络舆情, 行为监测, 民族用户, 行为分析

Research on Abnormal Identification and Early Warning of Online Public Opinion Risk of Ethnic Users in Online Social Networks

Minchao Zhang^{1,2}, Qiumei Pu^{1,2*}

¹Key Laboratory of Ethnic Language Intelligent Analysis and Security Governance of MOE, Minzu University of China, Beijing

²School of Information Engineering, Minzu University of China, Beijing

Received: Sep. 28th, 2023; accepted: Nov. 3rd, 2023; published: Nov. 15th, 2023

*通讯作者。

文章引用: 张敏超, 蒲秋梅. 社会网络视角下民族用户群体行为的识别机制研究[J]. 社会科学前沿, 2023, 12(11): 6346-6353. DOI: 10.12677/ass.2023.1211869

Abstract

In the era of big data, social networks are integrated with people's lives, the degree of people's participation in online life is constantly increasing, and social platforms have gradually become an important position for the government to communicate with the public. The rapid development of social networks has brought a lot of traffic and business value, but also attracted many abnormal users who profit from improper behavior. How to correctly deal with online public opinion events, accurately identify abnormal user behavior, and reduce its adverse impact has become a major issue of government work. By studying the abnormal behaviors of network group users in ethnic minority areas, this paper analyzes the characteristics of multi-user interaction behaviors, and uses the method of combining big data paradigm theory and machine learning algorithm research to carry out real-time calculation through network behavior data, which can realize the dynamic perception of hot events in ethnic minority areas, and then effectively monitor the development trend of public behaviors. It helps to improve social stability, maintain the social network environment, and promote national unity.

Keywords

Social Network, Online Public Opinion, Behavior Monitoring, Ethnic Users, Behavioral Analysis

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

一直以来, 舆情都是人们生活的重要组成部分, 它与国家和社会的稳定息息相关, 也影响着一个国家的政治生活, 因此, 舆情一直以来都受到国家与社会管理人员的高度重视。网络舆情是社会需求与以网络为中心的新媒体碰撞的产物, 它是指在社交媒体平台上对特定话题、事件、产品、品牌或个人的公众情感和观点的总体体现。而社交网络中用户的异常行为通常指的是用户在社交媒体平台上表现出的不符合正常社交互动范围或可能引发关注和担忧的行为模式。这种行为不仅挑战着国家的社会和政治稳定, 还对信息生态和公众参与产生深刻影响。所以, 国家持续关注着社交网络中用户异常行为的监管和应对, 维护网络安全、社会和谐以及公共利益的整体平衡。通过分析和监测社交媒体平台上的用户行为, 自动识别出那些不符合正常社交互动行为的模式, 包括虚假信息、仇恨言论、滥用、网络欺诈等, 以维护社交网络的健康和积极性, 保护用户的安全和权益。

在线社交网络中越来越多的民族地区用户持续活跃, 积累了大量媒体用户产生的舆情数据, 伴随着社会网络和基于位置服务的普及, 与公共安全密切相关的社会热点事件在物理空间和网络空间中快速传播与互动。社交网络是现实世界的映射, 其中存在着复杂的用户互动行为。通过对社交网络中用户交互行为建模, 分析其特征, 有利于了解社交网络中用户的交互行为模式。在社交大数据中发现异常用户群体、研究群体的活动规律, 对于抑制网络谣言传播从而维护网络信息安全, 监控突发事件从而做出智能决策等方面具有重要意义。

本文拟对社交网络中用户异常交互行为展开深入研究、科学实验和理论分析, 针对民族地区用户群体特有的文化背景、宗教信仰、地域等属性, 结合网络交互行为, 建立民族用户异常行为发现与预警模

型。旨在通过在线社交网络上民族地区用户的多种信息行为, 掌握少数民族用户的群体分布及特征等信息, 特别是对于社交网络上出现的不安定因素, 谣言以及恐怖势力之类的破坏民族团结的用户, 需要对相关用户群体进行监测, 并进一步对群体的转发行为和舆情发展趋势进行准确分析和预测。通过研究民族地区网络群体用户异常行为, 分析多用户交互行为的特征, 建模单用户具体交互行为, 进一步研究群体交互规律和用户影响力。结合大数据范式理论和机器学习算法的方法, 通过实时计算网络行为数据, 能够实现对民族地区热点事件的动态感知, 进而结合社会治理的目标, 有效监测公众行为的趋势演变。

2. 用户行为研究相关工作

2.1. 国内外研究现状

社交网络的日益普及使得网络舆情的传播十分迅捷, 网络舆情具有自由性、交互性、多元性、偏差性、突发性等特点, 能够左右民众的情感和判断, 能推动和改变事件的发展和走向, 容易被反对分子利用, 已经成为影响社会稳定的重要因素[1]。在大数据时代, 面对海量跨媒体信息, 要实现支持民族地区在线社会网络跨媒体数据分析与挖掘的智能化、精准化, 有必要深入研究以下科学问题: 社会网络分析中的大数据采集与存储问题, 用户行为分析与建模, 群体异常行为检测与可视化。

2.1.1. 社交网络大数据采集与存储

随着社交网络的发展, 大数据采集成为了一个日益重要且复杂的领域。社交网络上产生的海量数据包含了用户的文本、图像、视频、互动记录等多种形式的信息。因此, 有效地采集和管理这些大数据对于了解用户行为、趋势分析、个性化推荐、舆情监测和商业决策等方面至关重要。除新闻网站外, 人们广泛应用的社交媒体平台也是热点事件传播的重要途径。网络大数据的采集与存储是数据应用的重要环节[2]。数据存储方式中, 分布式数据存储也面临着总量超大规模、处理速度要求高和数据类型异质多样等难题。信息采集工具支持网页结构自动分析、元数据解析, 实现对各类型信息的及时、精确采集[3]。对信息采集进行分布式部署, 实现多线程采集, 能够在规定时间内完成监测信息源的采集工作。赵露提出以灵活的软硬件全可编程能力和并行算法作为引擎的数据采集构件, 在处理器的支持下实现对数据采集速度的适应性调节, 可以实现对数据的有效采集[4]。张雁涪采用网络爬虫技术设计企业大数据采集系统, 使企业大数据采集系统更符合时代发展的需求[5]。

本文针对民族地区社会网络媒体大数据的采集和存储, 主要是在指定地域、确定用户目标范围的基础上, 进而对这些数据进行处理, 从中分析和挖掘出有价值的信息。拟采用分布式系统架构方式建立多民族多语言数据采集平台, 提高地域社交网络数据采集效率, 实现多源异构社交网络数据信息提取与集成。

2.1.2. 用户行为分析与建模

社交网络用户行为是用户在对自身需求、社会影响和社交网络技术进行综合评估的基础上做出的使用社交网络服务的意愿, 以及由此引起的各种社会活动的总和, 是在线社交网络研究的重要课题。用户行为分析主要分为用户影响力分析与用户偏好分析两个方向。用户影响力可以通过分析在线少数民族社交用户的行为轨迹(如浏览、创建、转发、评论信息, 建立好友关系等) [6], 能够评估用户的影响力并预测用户可能产生的行为。而用户偏好分析则是从计算用户之间、用户与话题之间的相似性来考虑[7]。张柳等以学院门户网站为研究对象, 建立了基于 Hadoop 的网站用户行为分析系统框架, 进行了用户行为数据的采集、预处理、存储、分析等全过程研究, 并预测了用户行为的发生[8]。殷丽凤等针对网购用户所产生的购物行为进行分析, 通过数据处理构建客户关系管理模型, 并在此模型的基础上采用特征分箱法和聚类两种方法对用户进行细分[9]。马家豪在传播学视角下, 以朋友圈信息传播流程为线索剖析一些典

型的朋友圈用户行为, 引申出了较为深刻的社会化内涵, 为微信应用的持续革新与朋友圈这一功能的修正完善提供借鉴[10]。

民族地区用户行为往往与社会重大事件有着密切的联系, 但很少人对网络舆情传播在社交网络演化过程中所起到的作用进行研究与探索, 基于多语言跨平台的用户行为分析模型也同样有待于深入研究。本文对社交网络中民族用户信息行为构建模型, 分析用户群体的聚集情况, 追踪信息源以及发现有影响力的用户和民族群体感兴趣的话题。

2.1.3. 群体异常行为检测与可视化

群体异常行为检测依托于社交网络模型的构建[11], 对在线社交网络对象网络信息行为和现实活动行为进行分析, 从照片、图片、上下文信息这些社交数据中挖掘各种属性信息, 包括时间、人物、地标等属性信息, 从而分析社交网络对象的现实活动。通过群体行为检测[12], 可以从社交网络跨媒体数据中检测正在发生的事情, 同时分析空间和时间信息, 可以看到群体行为是如何演化发展的, 从而将社交网络对象的网络行为与其现实活动建立关联。为了从大量的社交数据中发现网络用户的异常行为特征, 可以对群体用户进行时空聚类、用户行为特征抽取、行为特征匹配、构建用户异常行为数据库。李忠等按照输入图的类型将工作分为面向静态图的异常检测与面向动态图的异常检测两大类, 并在两类异常检测中, 总结传统的、经典的方法, 以及近年来利用机器学习技术与神经网络技术的一些方法[13]。袁丽欣等对社交网络数据集提取用户内容、行为等特征, 应用梯度增强集成分类器算法进行特征选择, 建立分类模型, 构造非平衡数据集并识别三类垃圾广告发送账号, 提升了社交网络异常用户检测模型的鲁棒性[14]。

当前, 社会网络分析因其结论可视化强等优势, 广泛应用于文献关系和网络舆情分析中[15]。社会网络可视化分析与文本挖掘相结合, 有利于发现公众关注的热点集聚和扩散规律, 从而更准确、更直观地识别政府决策与民族地区网络舆情之间的相互关系。

本文拟对少数民族用户进行关系强度定量分析, 研究其行为特性与规律, 分析兴趣偏好, 建立更有针对性的用户行为分析模型, 推测用户的兴趣、政治倾向等属性, 为民族地区公共策略的制定提供参考依据, 对提高民族地区公众安全与社会服务等多方面具有长远的意义。

以上研究成果显示, 对社交网络中用户行为进行分析与识别, 能够为信息传播、信息检索、信息推荐等应用提供数据和技术支持。大数据环境下, 社交网络的发展使得民族地区网络用户之间的信息交流方式、方法等有了根本性变革, 催生出大量亟待解决的研究问题, 主要包括大数据环境下的民族用户群体行为模式、民族用户群体发现和异常行为检测等, 这都是数据驱动背景下网络信息行为分析的新发展、新变化带来的新问题。

2.2. 挑战与不足

目前, 国内外研究者对于在线社交网络中用户行为分析的关注度依然居高不下, 并且研究成果显著, 但以上提及的相关研究工作存在以下几方面的不足之处。考虑到针对大数据视角下少数民族用户的特点与应用场景, 对这些问题的研究不得不面临一些全新的挑战。

一方面, 由于少数民族的用户涉及到的民族语言以及民族文化的相关背景, 需要采取新的方法。通过构建民族用户信息行为模型, 实现对特定地域、事件、群体或个人的精准分析, 以便对突发事件、热点事件或意见领袖等进行实时监控和跟踪, 提供网络舆情态势预测。根据用户行为模式, 同时结合时空特性和社交特性, 利用机器学习和主题模型等方法, 研究社交网络中的动态社区、群体发现以及突发事件自动检测。

另一方面, 我国社交网络用户具有多民族多元化的特点, 少数民族用户作为一种特殊群体, 应加深该方面的研究。本文拟对少数民族用户进行关系强度定量分析, 研究其行为特性与规律, 分析兴趣偏好,

将具有相同行为特征、兴趣爱好的社交网络信息以及热点突发事件信息自动归类组织在一起，建立更有针对性的用户行为分析模型，这对商业活动以及舆论控制都有着至关重要的意义。

3. 用户异常行为的发现与预警

本文利用采集的社交媒体舆情数据提取用户行为特征，对社会网络中具有民族特征属性的数据进行有效表达。分析社交网络中用户发表的留言或评论、转发、点赞等交互行为，将用户在网络中的搜索、留言等行为看作人类行为和网络资源的交互，并采用提取统计特征、线性回归等方法，根据特征与交互行为对用户群体进行精准划分，并获取用户群体的行为特征，再把个体特征与群体特征进行对比，发现异常行为。拟采用理论研究、算法实现和实验数据验证相结合的研究思路，以理论知识为基础，算法实现为手段，实验验证为支撑。通过社交媒体大数据对信息传播行为的分析，发现民族用户信息行为特征，识别异常社交网络用户。具体而言，将开展如下研究工作，总体框架如图 1 所示。

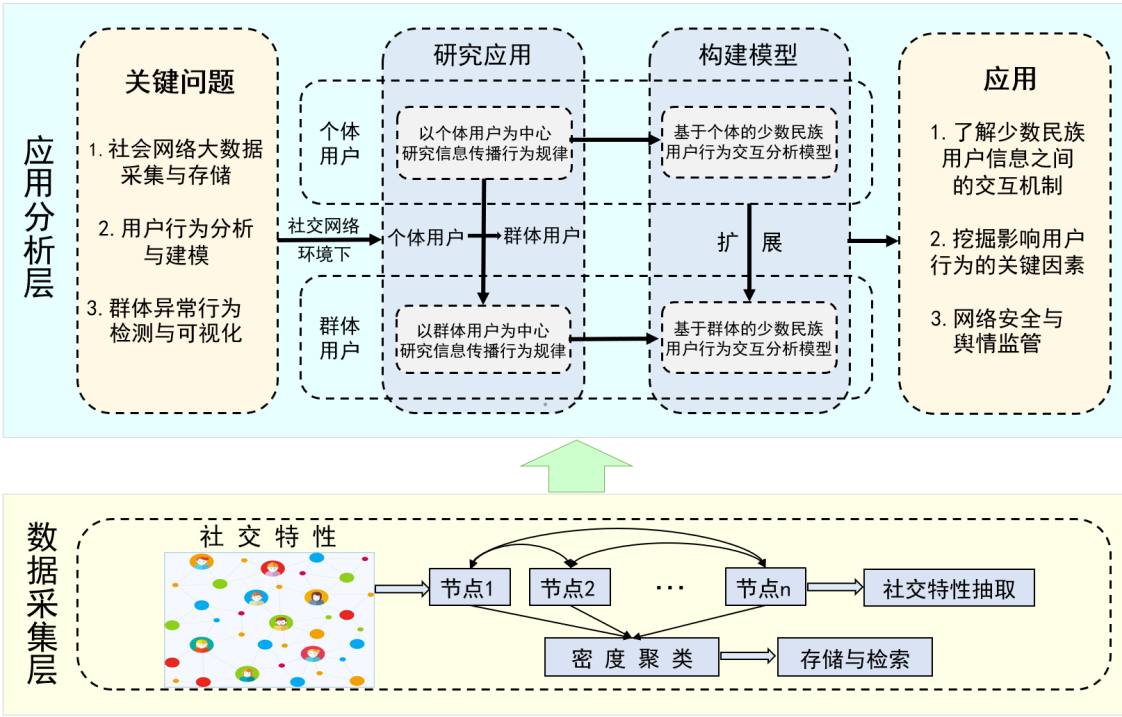


Figure 1. Research content frame diagram
图 1. 研究内容框架图

本文针对少数民族在线社交行为深入研究，从多源异构的社会网络媒体数据中提取用户特征指标，基于用户信息行为规律，构建社区发现模型，通过典型社会安全事件中用户行为的多维度分析，构建民族地区社会网络结构模型，从而促进社会网络安全发展，提升社会安全风险的预测预警能力，具体研究工作主要包含三个方面：1) 研究如何在数据采集过程中，将庞大的数据量化为微观个体特征、介观群体特性和宏观趋势分析；2) 研究如何对用户行为进行分析与建模，从多属性、多维度挖掘不同个体的行为特征；3) 提出异常行为检测的方法，从整体性角度实现网络用户异常行为的评估与分析。

3.1. 社交网络大数据采集

本文面向的是多民族聚居，多语言交流的民族地区，针对少数民族在线社交网络中媒体数据进行分

析和挖掘。数据采集工作一般集中限定在多民族聚居区和民族用户社交圈子进行, 首先向用户发布问卷, 获取社会网络中信息交互行为指标作为标注, 通过分析网络行为数据, 进而提取网络行为特征, 以此作为社区发现的理论基础。其次, 互联网公开信息源的数据格式多样, 社交用户所浏览过的数据, 会被上传至服务器, 为了实现从新闻网站、微博微信等平台提取用户信息交互行为数据, 拟基于开源框架的分布式爬虫技术搭建多源异构社交网络数据采集平台, 采集海量的社交媒体大数据, 并对其进行少数民族用户识别和检测, 确认身份后进行数据分析与处理。利用用户在不同的社交媒体中的行为的相互影响, 解决目前根据用户的历史行为日志数据建模产生的数据稀疏问题, 提高用户建模的准确性以及社交数据采集效率, 实现跨媒体社交网络数据信息提取与集成, 以提供充足的研究语料。最后, 结合线下调查、网络问卷调查等方法得到的数据, 从宏观和微观两个角度对行为特征进行分析。原始数据的异构性与语义缺乏阻碍了数据的深入分析与高效调用, 本文拟利用语义表达模型和元数据实现数据管理和调用, 进行个人数据的分析与建模。

3.2. 网络用户行为分析与建模

在线社交网络中存在大量用户信息行为, 本文计划分别针对区域分布广、人数多的民族如藏族、回族、壮族等用户群体, 以及傣族、苗族、白族等具有地域性较强的民族地区用户进行用户行为数据采集, 剖析社交网络中用户内容创建行为规律、群体互动中的关系选择行为规律和内容选择行为规律, 通过对少数民族虚拟社区用户的行为、兴趣等特征进行分析, 在属性多、复杂性高的网络信息行为数据中发现规律, 提取用户特征, 为进一步进行社区发现提供理论基础。

根据物以类聚人以群分的原理, 拟采用密度聚类方法对采集到的信息进行群体发现, 包括媒体内容与网络用户信息以及其行为数据。针对用户群体特有的文化背景、宗教信仰、地域等属性特征, 结合网络交互行为, 建立群体发现模型。并在用户行为分析与建模的基础上, 对用户聚类、分布和关注倾向等网络结构特征进行形象化的展示, 对网络用户的行为和用户交互进行挖掘, 呈现出用户之间的社交网络关系。

面向社交网络中海量碎片化的个体行为信息进行多属性、多维度的分析和挖掘, 发现不同个体的行为特征, 并通过用户行为分析与建模, 能够得到少数民族虚拟社区中用户自身的行为特征, 再利用用户的特征描述用户之间关系的紧密程度, 由此构建社会网络结构, 将用户节点按照内在的拓扑结构紧密程度划分成若干子图, 发现社会网络存在的民族用户群体。用户行为分析与网络结构建模的基本研究方法如图2所示:

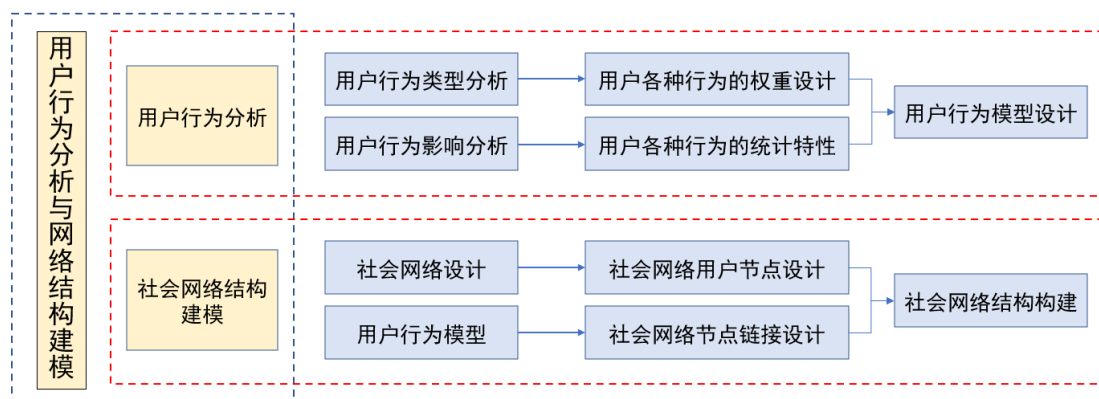


Figure 2. User behavior analysis and network relationship modeling

图 2. 用户行为分析与网络关系建模

3.3. 异常行为检测与可视化

社交网络数据带有时间、地点等时空信息, 基于密度聚类对社交网络数据进行时空聚类, 将具有时间相似性、空间相似性的群体数据进行聚类, 从中可以挖掘发现群体行为特征, 例如群体生活行为规律特征、用户情感特征、用户世界观等, 将抽取的行为特征构建群体行为特征模型, 之后针对个体对象发布的社交媒体数据, 抽取个体特征和群体特征进行匹配, 发现异常行为, 有效关联、分析和挖掘长时间、大规模的社交行为数据, 识别异常社交网络用户, 构建异常行为数据库, 帮助快速发现和定位异常, 及时做出判断和响应。

基于在线社交网络大数据的时空特性和社交关系特性, 结合民族地区网络用户现实活动特征进行关联分析, 分析用户网络行为特征与现实行为特征的关联, 从时空数据中找出严重偏离正常模式的对象, 识别异常社交网络用户, 构建异常用户行为特征数据库。

本文为更好地在社交网络平台中进行群体特征提取和异常行为检测, 通过构造用户关系网络以表示用户之间的关系, 利用复杂网络的相关理论基础, 分析在线社交网络的拓扑结构, 确定社交网络结构的构建模式, 并结合用户行为分析, 构建能够合理有效表达用户行为特征的社交网络拓扑结构。

4. 结语

任何事物都有两面性, 网络时代在普惠人民的同时也带来了随时可能爆发的负面网络舆情, 为了社会发展和国家善治, 网络治理成为不可忽视的部分, 本文结合当前国内外研究现状, 对在线社会网络中的用户行为规律进行研究, 揭示特定区域用户群体背景下的社会网络用户行为模式, 并进行异常行为分析与检测, 同时针对民族地区社会网络结构和用户信息行为建立有效的互联网民族地区用户行为研究体系和社会关系网络模型, 为用户的信息行为识别及信息传播能力的控制提供方法和依据, 同时监控和维护国家网络安全, 争取营造积极向上的网络文化, 建设一个和谐的网络环境。

基金项目

国家民委民族研究项目重点课题“基于社交网络的各民族交往交流交融现象多维动态分析”(2022-GMA-012)。

参考文献

- [1] 付琳, 张媛. 文本数据事件检测的研究热点及趋势分析[J]. 计算机技术与发展, 2023, 33(2): 24-31.
- [2] 梁艳华. 社交网络中大数据非结构稳定性检测仿真研究[J]. 计算机仿真, 2019, 36(5): 325-328+336.
- [3] 陈为东, 王萍, 王益成, 黄新平. 面向 Web Archive 的社交媒体信息采集工具比较研究[J]. 图书馆学研究, 2017(13): 10-16. <https://doi.org/10.15941/j.cnki.issn1001-0424.2017.13.003>
- [4] 赵露, 李庐. 网络用户行为大数据采集分析引擎设计[J]. 绵阳师范学院学报, 2022, 41(11): 83-87. <https://doi.org/10.16276/j.cnki.cn51-1670/g.2022.11.013>
- [5] 张雁沔. 基于网络爬虫技术的企业大数据采集系统设计[J]. 信息与电脑(理论版), 2023, 35(12): 154-156.
- [6] 陈雄逸, 许力, 张欣欣, 等. 社交网络基于意见领袖的谣言抑制方案[J]. 信息安全研究, 2023, 9(1): 57-65.
- [7] 周法国, 叶远胜, 张静, 等. 社交网站热点话题与用户分析——以知乎为例[J/OL]. 计算机工程与应用: 1-12. <http://kns.cnki.net/kcms/detail/11.2127.TP.20201125.1343.006.html>, 2023-09-22.
- [8] 张柳, 林德智. 高职院校门户网站用户行为分析系统构建研究[J]. 信息与电脑(理论版), 2022, 34(12): 60-62.
- [9] 殷丽凤, 路建政. 基于特征分箱和 K-Means 算法的用户行为分析方法[J/OL]. 云南民族大学学报(自然科学版): 1-9. <http://kns.cnki.net/kcms/detail/53.1192.N.20220510.1404.005.html>, 2023-09-22.
- [10] 马家豪. 社交 APP 中的个体用户行为分析——以微信朋友圈为例[J]. 新闻传播, 2021(18): 44-45.
- [11] 张广胜, 康昭, 田玲. 面向网络安全治理的用户身份识别技术与挑战[J]. 电子科技大学学报, 2023, 52(3): 398-412.

-
- [12] 曲强, 于洪涛, 黄瑞阳. 社交网络异常用户检测技术研究进展[J]. 网络与信息安全学报, 2018, 4(3): 13-23.
 - [13] 李忠, 靳小龙, 庄传志, 孙智. 面向图的异常检测研究综述[J]. 软件学报, 2021, 32(1): 167-193.
<https://doi.org/10.13328/j.cnki.jos.006100>
 - [14] 袁丽欣, 顾益军, 赵大鹏. 基于 XGBoost 方法的社交网络异常用户检测技术[J]. 计算机应用研究, 2020, 37(3): 814-817. <https://doi.org/10.19734/j.issn.1001-3695.2018.08.0651>
 - [15] 王一慈, 顾桂芳, 曹钰椒, 谢丽玲. 社交网络可视化分析对电商企业网络口碑营销的影响[J]. 江苏商论, 2020(2): 34-37+41. <https://doi.org/10.13395/j.cnki.issn.1009-0061.2020.02.008>