

基于大语言模型的高校舆情分析可视化系统

吴瑞, 翟春婕, 王军国

南京警察学院, 江苏南京 210023

DOI:10.61369/EST.2025050041

摘要 : 随着社交媒体在高校的广泛普及, 校园网络舆情呈现爆发式增长态势。传统舆情分析方法在处理海量异构数据时面临实时性差、语义理解不足等多重挑战。本文提出一种基于大语言模型技术的高校舆情分析可视化系统, 该系统融合了多源数据采集、深度语义分析和动态可视化三大核心模块。系统采用 SnowNLP 框架实现基础情感分析, 并结合阿里云情绪识别 SDK 增强细粒度情感分类能力; 通过构建校园专用词典有效解析网络用语; 利用 ECharts 引擎实现舆情时空演进的可视化呈现。实验测试表明, 该系统各功能模块运行稳定, 充分验证了系统设计的可行性和有效性, 并在高校网络舆情监测领域展现出良好的应用前景。

关键词 : 舆情分析; 高校舆情; SnowNLP; 可视化系统; 情感分析

Visualization System for University Public Opinion Analysis Based on Large Language Models

Wu Rui, Zhai Chunjie, Wang Junguo

Nanjing Police College, Nanjing, Jiangsu 210023

Abstract : With the widespread adoption of social media in universities, campus network public opinion has experienced explosive growth. Traditional public opinion analysis methods face multiple challenges, such as poor real-time performance and insufficient semantic understanding, when dealing with massive amounts of heterogeneous data. This paper proposes a visualization system for university public opinion analysis based on large language model technology. The system integrates three core modules: multi-source data collection, deep semantic analysis, and dynamic visualization. The system employs the SnowNLP framework for basic sentiment analysis and combines it with Alibaba Cloud's Emotion Recognition SDK to enhance fine-grained sentiment classification capabilities. It effectively interprets internet slang by constructing a specialized campus dictionary and utilizes the ECharts engine to achieve visual representation of the temporal and spatial evolution of public opinion. Experimental tests demonstrate that the functional modules of the system operate stably, fully validating the feasibility and effectiveness of the system design. It shows promising application prospects in the field of university network public opinion monitoring.

Keywords : public opinion analysis; university public opinion; SnowNLP; visualization system; sentiment analysis

引言

高校作为思想活跃的知识密集区, 其网络舆情状况直接影响校园的稳定与发展。根据《2023中国教育舆情年度报告》的统计数据, 高校相关舆情事件年增长率高达32%, 其中35%涉及学业压力、后勤管理等敏感话题。传统舆情分析方法主要面临以下三重挑战: 首先是在语义理解方面存在明显不足, 如SnowNLP等传统方案依赖静态词典, 难以准确识别反讽等复杂语言表达; 其次是实时性较差, 舆情预警存在明显滞后性; 第三是数据源相对单一, 缺乏对课程论坛、校园墙等高校专属平台的数据整合。这些问题严重制约了高校舆情管理工作的效率和效果。^[1]

大语言模型技术的迅猛发展为高校舆情分析带来了新的机遇。本研究综合运用SnowNLP框架、高校主题分类模型以及主题-情感融合分析方法, 构建了一个面向高校特定场景的舆情分析系统。本系统的主要创新点体现在: 建立覆盖校园主题的专用词典体系; 设计动态语义解析规则库; 开发可视化看板, 实现舆情传播路径的动态追踪。通过这三个方面的创新, 系统能够更好地适应高校舆情分析的特殊需求, 为校园管理提供更精准的决策支持。本研究的实践意义在于为高校提供了一套完整的舆情监测解决方案, 理论意义则在于探索了大语言模型在特定领域应用的新路径。

作者简介:

吴瑞(1992.03—), 男, 汉族, 安徽阜阳人, 工程师, 硕士研究生, 研究方向: 计算机技术;
翟春婕(1988.02—), 女, 汉族, 江苏东台人, 副教授, 博士研究生, 研究方向: 人工智能、公安情报;
王军国(1987.04—), 男, 汉族, 山东潍坊人, 工程师, 硕士研究生, 研究方向: 计算机教育技术。

一、相关技术与方法

(一) 舆情数据采集与清理

本系统采用 Python 中的 requests 库进行数据采集，该库提供了简洁易用的 API 接口，能够通过 HTTP 协议向远程服务器发送请求并获取数据。requests 库支持多种请求方法，包括 GET、POST 等，可以灵活添加 HTTP 头部、查询参数等内容。

在数据采集过程中，系统通过对目标网站进行抓包分析，确定请求参数，然后使用 requests 库模拟浏览器访问数据接口，获取所需的文章和评论数据。系统能够获取文章及评论的发布作者、发布时间、地理位置、点赞数等多维度数据，并将其存储到数据库中。为了提高数据采集的效率，系统采用了多线程技术，同时设置了合理的请求间隔，既保证了数据采集的速度，又避免了对目标网站造成过大的访问压力。^[2]

(二) 语义分析技术栈

1. SnowNLP 技术框架

SnowNLP 是一个基于 Python 的中文自然语言处理库，集成了分词、情感分析、文本分类等功能。情感分析基于朴素贝叶斯分类器，训练数据来源于电商评论、社交媒体文本等。情感分析结果的取值范围为 0 到 1，以 0.5 作为情感倾向的分界点。在实际应用中，系统会对分析结果进行进一步处理，将连续的情感得分离散化为正面、负面和中性三种情感类别，便于后续的可视化展示和统计分析。

在技术实现层面，SnowNLP 采用了先进的自然语言处理算法。中文分词基于 Character-Based Generative Model，能够准确识别中文文本中的词汇边界。词性标注采用隐马尔可夫模型，为每个词汇标注相应的词性。此外，系统还集成了 TextRank 算法用于关键词提取，能够从大量文本中自动识别出重要的关键词。这些功能的组合使用，为后续的舆情分析提供了坚实的技术基础。系统还对 SnowNLP 进行了针对性优化，通过增加高校领域语料的训练，提升了在校园场景下的分析准确性。^[3]

2. 阿里云情绪分析 SDK

为了提升情感分析的准确性，系统引入了阿里云情绪分析服务。该服务支持多种语言的情感分析，能够自动识别文本中的情感倾向，并评估情感强度。通过将 SnowNLP 与阿里云服务相结合，系统构建了多层次的情感分析体系。具体而言，SnowNLP 负责基础的情感倾向判断，而阿里云服务则提供更细粒度的情绪分类，包括喜悦、愤怒、悲伤等多种情绪类型。这种双引擎的设计既保证了分析效率，又提高了情感识别的准确性。在实际部署中，系统还实现了请求批处理机制，将多个文本请求合并发送，显著提高了 API 调用的效率。

(三) 可视化技术

1. ECharts 可视化库

ECharts 是一款功能强大的开源 JavaScript 数据可视化库，支持创建多种类型的交互式图表。它提供了丰富的图表类型，包括折线图、柱状图、饼图等，具有良好的交互性，适合用于舆情数据的可视化展示。系统充分利用了 ECharts 的配置灵活性，针

对不同的数据类型和分析需求，设计了多种可视化方案。例如，使用折线图展示舆情热度的时间变化趋势，使用饼图展示情感分布的占比情况，使用地图展示舆情的地域分布特征。系统还对 ECharts 进行了深度定制，开发了符合高校特点的主题配色方案，使可视化效果更加符合使用场景。

2. WordCloud 词云技术

WordCloud 通过字体大小和颜色展示文本中关键词的重要性分布。高频词汇以较大字体显示，从而快速呈现文本数据的核心主题，帮助用户把握舆论关注焦点。系统对词云的生成算法进行了优化，除了考虑词汇的出现频率外，还结合了词汇的情感属性，使得生成词云不仅能够反映话题热度，还能体现情感倾向。此外，词云支持交互操作，用户可以通过点击特定词汇查看相关的详细数据。系统还实现了动态词云功能，支持按时间维度展示词汇热度的变化过程。

二、系统架构与功能模块

(一) 系统整体架构

本系统采用分层模块化设计，包含五个核心层次：网络爬虫层、数据分析层、数据存储层、业务服务层和前端展示层。网络爬虫层负责采集多源舆情数据，支持微博、论坛等多种数据源；数据分析层集成双情感分析引擎，提供多层次的情感分析能力；数据存储层采用 MySQL 数据库，确保数据的安全性和一致性；业务服务层基于 Flask 框架，提供完整的 RESTful API 接口；前端展示层采用现代 Web 技术结合可视化库，提供友好的用户交互界面。

各层次之间通过定义清晰的接口进行通信，实现了系统功能的有效解耦。这种架构设计不仅提高了系统的可维护性，还便于后续的功能扩展。例如，当需要增加新的数据源时，只需在网络爬虫层进行相应扩展，而不会影响其他层次的功能。同时，系统采用了缓存机制，将频繁访问的数据缓存在内存中，显著提高了系统的响应速度。在部署架构上，系统支持分布式部署，各个层次可以根据负载情况独立扩展，确保了系统在大规模应用时的稳定性。

(二) 系统功能模块

1. 系统首页概览

系统首页提供舆情数据全景概览，采用卡片式布局展示关键指标。通过多种图表展示舆情特征：折线图显示发布趋势，玫瑰图展示文章分布，词云图呈现用户特征。首页还设置了时间范围选择器，支持用户查看不同时间段的舆情数据。所有图表都支持交互操作，用户可以通过点击图表元素查看详细数据，实现数据的下钻分析。此外，首页还提供了预警信息展示区域，实时显示需要重点关注的热点舆情，帮助用户快速把握舆情态势。

2. 热词统计分析模块

该模块对评论文本进行深度挖掘，采用 jieba 分词结合 TF-IDF 算法。包含热词时序分析、情感分析和查询服务，能够识别舆情焦点变化规律。模块还提供了热词对比功能，支持用户选择多个热词进行对比分析，查看不同热词在时间维度上的变化趋势。热词情感分析功能可以展示每个热词对应的情感分布，帮助用户理解

舆论的情感倾向。该模块还创新性地引入了热词关联分析，通过共现分析算法挖掘热词之间的关联关系，揭示更深层次的舆情特征。^[4]

3. 舆情分析统计模块

基于SnowNLP框架构建情感分析流水线，采用朴素贝叶斯算法。系统将情感分析结果以表格形式呈现，支持逐条查看文本内容及情感评分。该模块提供了丰富的数据筛选功能，用户可以按时间范围、情感类别、数据来源等多个维度对数据进行筛选。分析结果支持导出功能，用户可以导出Excel或CSV格式的数据文件，便于后续的离线分析。模块还提供了情感趋势预测功能，基于历史数据建立时间序列模型，预测未来一段时间的情感分布趋势。

4. 文章分析模块

该模块从点赞量、评论量和转发量三个维度评估内容传播效果，为舆情引导策略提供数据支持。模块采用多指标综合评价方法，不仅关注单一指标的表现，还通过构建综合指数来评估内容的整体影响力。系统会自动识别高影响力内容，并通过专题分析功能深入剖析这些内容的传播规律和受众特征。此外，模块还提供了传播路径分析功能，通过追踪内容的转发链条，可视化展示内容的传播过程和关键节点。

5. 地域分析模块

通过ECharts地理可视化组件展示舆情数据地域分布，采用颜色编码机制显示各地区参与度差异。地图支持多级下钻，用户可以从全国视图下钻到省份视图，查看更详细的地域分布数据。模块还提供了地域对比功能，支持用户选择多个地区进行横向对比，分析不同地区的舆情特征差异。为了更好地服务高校管理，模块还特别标注了各个高校校区的位置信息，方便管理者快速定位到具体校区的舆情状况。

6. 评论分析模块

整合情感分析SDK和WordCloud技术，构建多层次的评论分析体系，包含情感分布、词云可视化和情感趋势分析。该模块特别注重评论内容的实时分析，能够快速捕捉评论情感的变化趋势。系统会自动识别评论中的关键意见领袖，并对其评论内容进行专项分析，帮助用户把握舆论导向。模块还提供了评论质量评估功能，通过分析评论的情感强度、内容相关度等指标，自动识别高质量评论和垃圾评论。

7. 热门话题舆情分析模块

支持特定话题的深度监测，用户可通过搜索框输入关键词，系统自动采集相关数据并实现实时追踪。该模块提供了话题传播路径分析功能，可以可视化展示话题的传播过程和关键传播节点。同时，模块还支持话题预警功能，当话题热度或负面情感超

过预设阈值时，系统会自动发出预警通知。模块还提供了话题影响力评估功能，通过构建多维指标体系，量化评估话题在校园内的影响范围和强度。

三、系统测试与验证

我们进行了全面的功能测试和性能验证，覆盖所有核心模块。测试环境采用分布式部署架构，使用多台服务器共同承担系统负载。在压力测试中，系统在并发用户数达到1000时仍能保持稳定运行，平均响应时间控制在2秒以内。情感分析的准确率测试显示，系统在测试集上的准确率达到85.6%，基本满足实际应用需求。特别是在高校特定场景的测试中，系统对校园用语的识别准确率比传统方法提升了15%。

在用户体验测试方面，我们邀请了20名高校管理人员参与测试。测试结果显示，用户对系统界面的易用性和功能的完整性给予了积极评价。特别是在数据可视化方面，用户认为系统的图表设计直观易懂，能够快速把握舆情态势。同时，用户也提出了一些改进建议，包括增加更多维度的数据分析、优化移动端使用体验等，这些建议将在后续版本中逐步实现。系统还在某高校进行了为期三个月的试运行，期间成功预警了5起潜在舆情事件，证明了系统的实用价值。^[5]

四、结论与展望

本文设计并实现了一个基于大语言模型的高校舆情分析可视化系统，通过整合多源数据采集、深度语义分析和动态可视化技术，构建了完整的高校舆情监测解决方案。系统采用模块化架构，具有良好的扩展性。为学校的舆情管理工作提供了有力支持。系统的创新点在于将大语言模型技术与高校特定场景相结合，通过专用词典和语义规则库，有效提升了在校园环境下的分析准确性。

未来的研究工作将集中在提升情感分析准确性、扩展数据采集范围、优化系统性能等方面，进一步增强系统的实用性。具体来说，我们计划引入更先进的大语言模型，提升对复杂语言表达的理解能力；扩展数据采集范围，覆盖更多高校特有的信息平台；优化系统架构，提高处理大规模数据集的效率；探索跨平台数据融合分析，实现更全面的舆情监测。通过这些改进，系统将能够为高校舆情管理提供更全面、更精准的支持，并在更多高校推广应用。

参考文献

- [1] 蔡增玉, 韩洋, 张建伟, 等. 基于SnowNLP的微博网络舆情分析系统 [J]. 科学技术与工程, 2024, 24(13):5457-5464.
- [2] 于路遥, 宋瑾钰. 基于Python的天气信息可视化分析系统的设计与实现 [J]. 软件工程与应用, 2022, 11(6):1394-1403. DOI:10.12677/SEA.2022.116143.
- [3] 韩磊, 夏明亮, 施展, 等. 基于大数据技术的高校舆情分析模型研究 [J]. 智能计算机与应用, 2024, 14(11):194-199. DOI:10.20169/j.issn.2095-2163.241130.
- [4] 张金霜. 基于TF-IDF的微博特征提取网络舆情分析系统设计 [J]. 电子技术, 2025, 54(02):59-61.
- [5] 赵韶雅. 高校网络舆情情感分析与话题发现研究 [J]. 科技传播, 2024, 16(17):91-95. DOI:10.3969/j.issn.1674-6708.2024.17.020.