

# 基于微博博文情绪分析的话题热度趋势预测研究

鞠屹昂<sup>1</sup> 张天宇<sup>1</sup> 李佳熹<sup>1</sup>

<sup>1</sup>上海科技大学

2023年

## 研究背景

- 微博每日产生数百万条博文反映公众对热点事件的即时反应
- 话题热度生命周期与用户情绪波动密切相关（如明星绯闻事件的情绪演变）

## 研究方法

- 数据挖掘获取微博话题数据
- 建立群体情绪-热度关联模型
- 预测未来1小时热度趋势

- 研究流程:

- ① 数据获取与预处理
- ② 情绪量化分析
- ③ 构建时间序列
- ④ 特征相关性分析
- ⑤ 预测模型构建

图: 情绪-热度关联示意图

## 微博热度预测方法分类

- 基于特征的方法（王晓萌等）
- 基于时序的方法（X Chen等）
- 基于用户行为的方法（H Wang等）
- 传统机器学习（WH Tan等）
- 深度学习（G Chen等）
- 集成学习（J Chen等）

## 本文创新点

- 大语言模型+提示词工程替代朴素贝叶斯
- 基于Russell情绪环形模型的情绪煽动值公式

## 数据处理流程

- 数据爬取：微博API+爬虫工具
- 预处理：去标签、分词、语料库构建
- 情感分析：DeepSeek-R1模型
- 特征提取：
  - 情绪均值/方差
  - 情绪影响力分数(EIS)
  - 关键词词云(TF-IDF)

图：情绪分析流程

## 情绪影响力分数(EIS)

$$EIS = \frac{A \times (1 + |V|)}{1 + A/V} \quad (1)$$

- A: 情绪唤起度(Arousal)
- V: 情绪效价(Valence)

## 特征工程

- 时间特征: arousal\_mean\_ma3(3小时移动平均)
- 统计特征: valence\_std(效价标准差)
- 复合特征: EIS\_mean/EIS\_std

## 数据集

- 话题：《何凯文考研英语造假》
- 时间范围：2025.3.5-3.10
- 数据清洗：
  - 去除博文j10条的时段
  - 排除凌晨1-6点数据
- 训练测试比：6:4

图: 热度与情绪变化趋势

# 实验结果

## 模型配置

- XGBoost分类器
- 关键参数：
  - `n_estimators=200`
  - `max_depth=6`

## 主要发现

- 愤怒情绪高峰与热度高峰时间重合
- 理性情绪在末期占比更高
- EIS特征显著提升预测准确率

图: 情绪热力图



## 研究贡献

- 验证了情绪特征对热度预测的有效性
- 提出情绪影响力分数(EIS)新指标
- 实现基于大模型的细粒度情绪分析

## 未来工作

- 扩展更多情绪维度
- 优化时间序列建模方法
- 构建更大规模基准数据集

感谢聆听!

鞠屹昂	2021533088
张天宇	-
李佳熹	2022533184