

# 基于 SnowNLP 的比亚迪新能源汽车在线评论情感分析与知识图谱构建研究

郝田佳 李振 华正晓 连腾 张忠浩 宋凯 (通讯作者) 李翠 (通讯作者)

青岛理工大学, 山东省临沂市费县, 273400;

**摘要:** 随着新能源汽车产业的快速发展, 用户反馈成为企业优化产品与服务的重要依据。本研究基于 SnowNLP 模型, 以懂车帝平台 2023-2025 年期间的比亚迪新能源汽车用户评论为研究对象, 运用文本挖掘与知识图谱可视化, 系统分析用户关注焦点与情感倾向。研究发现, 用户对比亚迪的续航、空间及性价比评价积极, 但对智能驾驶、售后服务及价格体系评价消极。本研究旨在为比亚迪产品优化与营销策略制定提供数据支持, 也为文本挖掘技术在汽车行业的应用提供实践参考。

**关键词:** SnowNLP; 文本挖掘; 情感分析; 知识图谱; 比亚迪新能源汽车

**DOI:** 10.69979/3041-0673.26.02.004

## 引言

在全球能源转型与“双碳”战略推动下, 中国新能源汽车产业快速发展。比亚迪作为国内新能源汽车领域的领军企业, 其用户评价对行业发展具有重要参考价值<sup>[1]</sup>。随着社交媒体与专业汽车平台普及, 海量的在线评论为研究消费者真实需求提供了数据规模大、真实性高、时效性强的数据来源。本研究基于 SnowNLP 模型对懂车帝平台用户评论进行情感量化分析<sup>[2]</sup>, 结合词频统计与知识图谱可视化, 系统挖掘用户关注的产品优势与现存问题, 以数据驱动方式为企业产品迭代与服务优化提供决策支持。

## 1 数据获取与研究方法

### 1.1 数据来源与采集

本研究选取懂车帝平台作为数据来源, 该平台凭借其庞大的用户基础、专业的内容和高度活跃的互动环境,

积累了具有重要研究价值的真实用户评价数据<sup>[3]</sup>。

为确保数据的时效性与代表性, 本研究运用 Python 爬虫技术采集了 2023 年 1 月至 2025 年 1 月期间比亚迪新能源汽车相关的用户评论, 覆盖“汉 EV”、“唐 DM-i”、“海豚”、“秦 PLUS”等主要畅销车型。数据采集涵盖用户对各车型在产品性能、使用体验及售后服务等方面的具体评价, 为后续分析提供了多维度、立体化的数据支撑。

### 1.2 数据预处理

#### 1.2.1 数据清洗

本研究对采集的用户评论文本进行了系统清洗, 剔除了标点符号、表情符号等噪声信息<sup>[4]</sup>。经过严格筛选, 最终获得 20679 条有效评论数据。这些数据集中反映了用户对比亚迪新能源汽车的态度倾向, 为后续分析提供了可靠的数据基础。清洗后的部分数据示例如表 1 所示。

表 1 经过清洗后的部分数据展示

用户名	车型	日期	购买地点	评价
彦十 970	2025 款夏 DM-i180KM 卓越型	2025-1-1	湖州	我觉得这款车确实比较实在……
Xpbou 鑫鹏	2025 款夏 DM-i180KM 超越型	2025-1-1	鞍山	好多功能只是宣传, 都没有上线……
范与沫 o o	2025 款唐 DM-i115KM 尊贵型	2025-1-1	宜春	喜提大黑糖一周, 赶在年底最后一天提车……
唐小灰的主人	2025 款唐 DM-i115KM 云辇旗舰型	2025-1-1	海口	25 款唐 DM 造型溜背有点严重……
智者海风 6dO	2025 款唐 DMDM-i115KM 尊荣型	2025-1-1	成都	未央灰唐很好看, 开了几天, 下雨沾灰也不显脏……

#### 1.2.2 停用词处理

本研究对清洗后的评论文本进行了停用词过滤处

理, 构建了包含 306 个词汇的专用停用词表。该词表在通用停用词表基础上, 结合汽车评论的语言特点进行了

优化：在保留“续航”、“智能驾驶”等专业术语的同时，系统剔除了“的”、“是”、“了”等无实际语义的词，以及“车”这一在本研究特定语境下缺乏区分度的高频词。这一优化显著提升了后续词频统计与情感分析的准确性与有效性。

### 1.2.3 分词

为实现评论文本的结构化分析与特征提取，本研究采用严格的处理流程：首先，运用 Jieba 分词工具对清洗后的评论文本进行精确分词<sup>[5]</sup>；其次，基于 306 个停用词表对分词结果进行过滤处理；最后，对处理后的词进行系统的词频统计。

## 2 BYD 新能源汽车相关评论实证研究

### 2.1 高频词统计与可视化

对于采集到的客户评论，使用 Jieba 分词对文本进行分词处理后，利用 Counter 计数器进行词频统计，见表 2。同时，运用 Echarts-wordcloud 绘制高频词积极情感倾向词云图，见图 1。

表 2 正向评价高频词汇统计部分展示

word	count	占比
电	19389	4.15%
空间	18851	4.04%
续航	15872	3.40%
驾驶	14690	3.15%
油	14167	3.03%
充电	11976	2.56%
动力	11564	2.48%

根据表 2 及图 1 可知，用户对比亚迪汽车的“续航能力”、“空间”等方面评价积极。通过这些高频词可以了解 BYD 汽车用户的关注点，同时充分表明比亚迪在电池技术、产品设计与成本控制方面的优势得到了市场认可。



图 1 高频词积极情感倾向词云图

本研究还整理了 2023 年-2025 年期间的消极情感倾向词云图，见图 2。用户反馈的焦点高度集中于“价格

权益”、“智能驾驶”等维度。这些高频词集中反映了当前产品与服务体系中存在的主要不足，为比亚迪精准定位优化方向提供了关键的数据依据。



图 2 高频词消极情感倾向词云图

### 2.2 情感分析

基于 SnowNLP 模型对 20679 条有效评论进行文本挖掘和情感量化分析<sup>[6]</sup>，结果显示整体情感得分均值为 0.68，其中积极情感倾向词 364302 个，消极情感倾向词 102671 个，表明用户对比亚迪新能源汽车的整体评价偏向正面。

深入分析发现，用户情感倾向与产品特性高度相关。积极情感主要集中在“续航能力”、“内部空间”及“性价比”等维度，相关评论情感得分普遍高于 0.7，印证了比亚迪在核心技术方面的优势。消极情感则聚焦于“智能驾驶系统”、“售后服务体验”和“价格权益保障”等方面，这些维度的情感得分多低于 0.4，反映出产品与服务体系中存在的明显短板。

不同车型的情感分布呈现差异化特征：“海豚”车型因经济性获得较高情感评分，而“汉 EV”在智能驾驶方面的负面评价相对集中。这一情感分析结果不仅验证了前期词频统计的发现，更为知识图谱构建提供了关键的情感维度，实现了从用户关注点到情感态度的深化洞察。

### 2.3 知识图谱应用

为克服传统文本分析方法在处理复杂语义关联时的局限性，本研究引入知识图谱技术。知识图谱将不同种类的信息连接成关系网络，提供了从关系角度分析问题的能力。

本研究的知识图谱构建主要包含以下步骤：首先，通过命名实体识别（NER）技术<sup>[7]</sup>，从评论中自动抽取关键实体，主要包括“车型实体”（如汉 EV、海豚等）和“属性实体”（如续航、智能驾驶等）。在此基础上，通过语义关系分析，识别并抽取连接各实体之间的谓语

关系，例如“拥有”“需要改进”等。最终，将这些元素组成“车型-情感-属性”三元组，形成知识图谱。基于知识图谱可视化效果（见图3），可直观清晰地识别各车型的优势特征与待改进领域<sup>[8]</sup>。

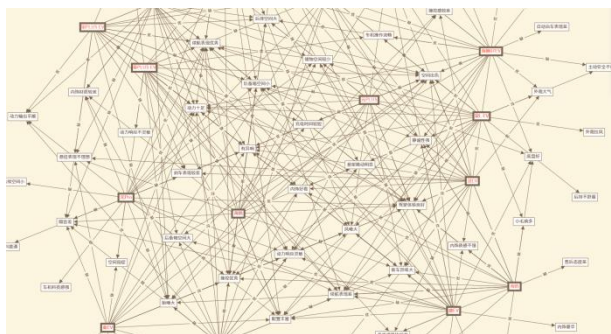


图3 知识图谱可视化效果展示

该知识图谱的应用价值主要体现在以下三个方面：

第一，车型优劣势的精准定位与画像生成。知识图谱能够直观呈现每个车型的用户认知画像。以“宋L EV”为例，其与“续航表现优秀”等属性形成“优势”关联，而与“后备箱空间小”等属性呈现“劣势”连接（见图4），帮助企业精准识别各车型的具体长短板。

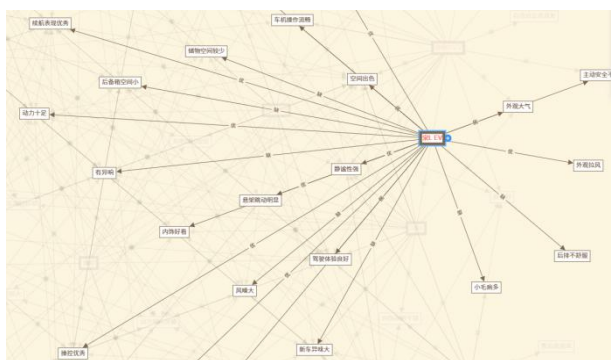


图4 “宋L EV”车型对应的优缺点

第二，产品特征的反向溯源与市场定位。知识图谱支持灵活的反向查询，可从特征角度定位车型集群。如查询“空间出色”特征，可快速定位到海狮07EV、宋L EV等具备该优势的车型（见图5），为企业识别细分市场优势和规划新车型提供参考。

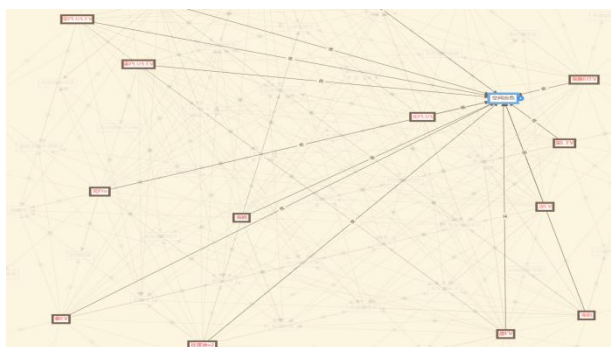


图5 “空间出色”优点对应的车型

第三，竞争态势的结构化洞察与策略制定。通过将竞品车型纳入同一知识图谱，可实现竞争态势可视化。例如对比“汉EV”与竞品在“智能驾驶”等关键属性上的关联强度与情感极性，清晰揭示各车型的差异化定位和相对优劣势，为产品策略制定和市场传播重点调整提供客观依据<sup>[9]</sup>。

## 3 结论与建议

### 3.1 结论

本研究基于SnowNLP模型，对懂车帝平台2023-2025年期间的比亚迪新能源汽车用户评论进行情感分析与知识图谱构建，系统识别用户关注的焦点维度与情感倾向。研究结果显示：用户对比亚迪汽车在续航能力、空间布局、性价比等方面评价积极，情感得分均值为0.68，整体评价偏向正面；而在智能驾驶系统、售后服务体验及价格权益保障等方面存在显著不满。知识图谱进一步揭示不同车型的差异化表现，如汉EV在设计上获认可，而智能驾驶为其弱关联属性；海豚在经济性与操控性上表现突出，内饰材质则评价较低。

需要指出的是，本研究的数据来源集中于懂车帝平台，样本范围和用户群体存在一定局限性，情感分析结果可能无法完全反映整体用户意见。此外，SnowNLP模型在处理汽车领域专业术语和隐含情感时的精度仍有提升空间。因此，在将结论推广为全面判断时需保持谨慎，建议后续研究拓展多平台数据融合与模型优化。

### 3.2 建议

#### 3.2.1 优势强化与传播策略

一是深化技术优势传播。针对用户认可的“续航”、“空间”及“性价比”等核心优势，制作高质量技术内容，通过官网专区与社交媒体矩阵精准传播，重点展示刀片电池、DM-i系统等技术亮点，并开展用户激励计划，鼓励口碑分享<sup>[10]</sup>。

二是优化品牌形象建设。依托用户对“支持国产”的情感认同，将“王朝”、“海洋”系列设计与中华优秀传统文化元素深度融合<sup>[11]</sup>，通过微纪录片、文化沙龙等形式讲述创新故事，提升品牌附加值与情感联结。

#### 3.2.2 短板改进与体验升级

一是优化智能驾驶与售后服务。针对用户反馈的“智能驾驶滞后”、“售后服务响应慢”等问题，加大辅助驾驶研发投入，建立体验中心展示技术优势；同时完善服务网络，推行“24小时响应机制”，构建用户评价闭环提升服务品质。

二是完善价格体系与权益保障。针对“价格权益”



争议,建立透明的价格调整与补偿机制,推出电池终身质保等长效权益以增强用户信任。通过动态定价模型<sup>[12]</sup>实现差异化促销,减少价格波动引发的负面舆情,提升产品竞争力。

#### 4 结束语

本研究基于 SnowNLP 模型与知识图谱技术,通过对懂车帝平台用户评论的分析,明确了比亚迪新能源汽车的优劣势及车型差异,为企业决策提供了数据支撑。但研究存在局限性:其一,数据仅源于懂车帝,数据来源单一;其二, SnowNLP 模型在处理汽车领域专业术语时精度有待提升。未来可通过以下方式改进研究:一是拓展多平台数据源;二是结合企业多维度数据;三是将文本挖掘技术应用于企业舆情监测体系,实现用户需求的动态感知与响应,为比亚迪等新能源汽车企业的可持续发展提供更精准的赋能。

#### 参考文献

- [1]刘云,李乾杰.比亚迪财务绩效评价[J].合作经济与科技,2025,(20):163-165. DOI:10.13665/j.cnki.hzjjykj.2025.20.051.
- [2]奚凡,唐悦珂,奚丽,等.基于网络文本的少数民族村寨旅游地游客感知与情感分析——以西江千户苗寨为例[J/OL].资源开发与市场,1-12[2025-10-01].<https://link.cnki.net/urlid/51.1448.N.20250729.1650.002>
- [3]刘久江,梁晶,鄢章华.新能源汽车评论数据分析与销量预测——基于懂车帝的数据[C]//中国物流学会,中国物流与采购联合会.2024年(第二十三次)中国物流学术年会论文集.哈尔滨商业大学管理学院;,2024:331-350. DOI:10.26914/c.cnkihy.2024.072010.
- [4]许昕.基于文本情感分析的高校学生评教研究——以南京邮电大学为例[J].大学教育,2025,(08):1-4+9. DOI:CNKI:SUN:DXJY.0.2025-08-001.
- [5]田冲.基于文本挖掘的公众绿色消费情感分析[D].河南工业大学,2025.
- [6]宋春燕.基于文本挖掘的新能源汽车消费者行为特征与情感分析研究[D].贵州大学,2024. DOI:10.27047/d.cnki.ggudu.2024.003317.
- [7]程健.基于NER和情感分析的股票交易策略研究与实现[D].西安电子科技大学,2024. DOI:10.27389/d.c

nki.gxadu.2024.001995.

- [8]黄珣,段金利.基于LDA-RoBERTa模型的在线评论主题及情感分析:以京东商城手机为例[J].西安石油大学学报(社会科学版),2025,34(04):69-78. DOI:CNKI:SUN:XASS.0.2025-04-007.
- [9]袁梦秋.印度市场电动汽车消费者偏好[D].大连外国语学院,2025. DOI:10.26993/d.cnki.gslyc.2025.000340.
- [10]赵允天,金玲.“双碳”目标背景下新能源汽车企业财务绩效分析——以比亚迪为例[J].老字号品牌营销,2025,(17):151-154. DOI:CNKI:SUN:LZHP.0.2025-17-050.
- [11]魏葳.中华优秀传统文化元素在汽车设计中的应用[J].文化产业,2024,(05):130-132. DOI:CNKI:SUN:WHCC.0.2024-05-044.
- [12]傅淳新.考虑绿色研发投入水平和市场风险的新能源汽车动态定价研究[D].东华大学,2025. DOI:10.27012/d.cnki.gdhuu.2025.000250.

作者简介:郝田佳,青岛理工大学,女,21岁,物流管理专业,本科生;

李振,青岛理工大学,男,22岁,物流管理专业,本科生;

华正晓,青岛理工大学,女,20岁,会计学专业,本科生;

连腾,青岛理工大学,女,22岁,物流管理专业,本科生;

张忠浩,青岛理工大学,男,22岁,电子商务专业,本科生;

宋凯,青岛理工大学,31岁,女,博士,讲师,大数据挖掘、科技管理、智慧物流,

系本文第一通讯作者;

李翠,青岛理工大学,41岁,女,副教授、管理工程系主任助理,物流与供应链管理方向,系本文第二通讯作者。

基金项目:2025年国家级大学生创新创业训练计划项目“绿擎智算:新能源汽车市场多模型耦合预测与生态优化决策系统”;山东省本科教学改革研究项目重点项目:基于产教融合共同体的商贸物流新形态实践教学体系研究(Z2023145)