

基于 LDA 主题模型的新能源 汽车网络社区用户消费者偏好研究

孙婧钰

(西安汽车职业大学, 陕西 西安 710600)

摘要:以汽车之家新能源汽车活跃度前五的品牌为例,分析汽车网络社区中购车用户评论,对非结构化数据进行分析,得到比较规范的车评数据。从匹配情感词、修正情感倾向、情感分析三部分展开阐述,并将情感词表与评论数据相匹配,得出初始的情感系数,并以此衡量消费者的特征情感。研究新能源汽车领域用户情感特征,并结合新能源汽车产品参数对销量的影响,分析消费者的需求偏好,揭示消费者购买行为背后的动因和决策机制,从而深入分析属性特征背后的消费者情感态度对消费者购买决策的影响。

关键词:新能源汽车;汽车网络社区;数据分析;需求偏好;LDA 主题模型

中图分类号:U46 **文献标识码:**A **文章编号:**1674-957X(2025)05-0135-03

Research on Consumer Preference of New Energy Vehicle Online Community Users Based on LDA Topic Model

Sun Jing-yu

(Xi'an Vocational University of Automobile, Shaanxi Xi'an 710600)

Abstract: Taking the top five active brands of new energy vehicles in Autohome as an example, this paper analyzes the user comments of car buyers in the automobile network community, analyzes the unstructured data, and obtains relatively standardized car review data. This paper elaborates on three parts: matching emotion words, correcting emotion tendency and emotion analysis, and matches the emotion word list with the comment data to get the initial emotion coefficient and measure the characteristic emotion of consumers. This paper studies the emotional characteristics of users in the field of new energy vehicles and combines the impact of product parameters of new energy vehicles on sales volume, analyzes consumers' demand preferences, reveals the motivation behind consumers' purchase behavior and decision-making mechanism, and further analyzes the impact of consumers' emotional attitude behind attribute characteristics on consumers' purchase decisions.

Key words: New energy vehicles; Automobile network community; Data analysis; Demand preference; LDA topic model

DOI:10.19475/j.cnki.issn1674-957x.2025.05.011

0 引言

据国际能源署(IEA)的数据,全球新能源汽车的销量在近五年间增长超过三倍。准确了解消费者的需求和偏好成为新能源汽车企业亟需解决的关键问题。传统的市场调研方法存在周期长、成本高、样本量有限等不足,难以快速响应市场变化;网络社区中的用户评论蕴含着大量有价值的信息,分析这些评论数据,可以更迅速地获取消费者的真实反馈和偏好。因此,基于网络社区评论的消费者偏好研究具有理论意义和实际应用重要价值。通过网络社区中新能源汽车相关评论数据,运用 LDA (Latent Dirichlet Allocation) 主题模型揭示影响消费者购买决策的主要因素。LDA 是一种常用的主题模型,广泛应用于文本数据的挖掘和分析,是通过假设文档由多个主题组成,每个主题又由多个词组成,从而识别出文本中潜在的主题结构,以揭示消费者关注的焦点和偏好,识别出新能源汽车的技术热点和消费趋势。

1 网络社区评论的情感分析研究现状

网络社区评论作为用户生成内容,包含了丰富的情感信息,能够直接反映消费者的使用体验和意见。情感分析,也称意见挖掘,是通过计算机技术对文本中的情感倾向数据进行整理、分析和提取。在新能源汽车领域,现有研究主要通过网络评论的情感分析来了解消费者对产品的看法和态度。例如,有学者通过爬取京东、淘宝等电商平台上的新能源汽车评论,进行情感分类和特征提取,以探讨不同品牌和车型的消费者满意度差异。研究表明,消费者对新能源汽车的关注点主要集中在续航里程、充电便利性、性价比等方面。此外,一些研究还分析了评论文本中的共现词汇和情感变化,揭示了消费者偏好的动态特征。尽管情感分析在新能源汽车领域的应用已经取得了一定的进展,但现有研究仍存在一些不足之处:多数研究主要依赖电商平台的评论数据,忽视了专业汽车论坛和社交媒体上的用户讨论;情感分析算法的精度和适应性有待

基金项目:西安汽车职业大学专项科研计划项目“基于 LDA 主题模型的新能源汽车网络社区用户消费者偏好研究”(2023KJ005)。

作者简介:孙婧钰(1988—),女,山东省济南市,汉族,讲师,硕士研究生,研究方向:人力资源管理。

提高,特别是在处理中文评论时面临更大的挑战;现有研究多集中在单一品牌或车型,缺乏跨品牌、跨平台的比较研究。LDA 是一种常用的主题模型,广泛应用于文本数据的挖掘和分析。

2 基于新能源汽车网络社区用户评论的数据爬取与预处理

2.1 Python 爬虫研究数据采集与数据清洗

(1)汽车网络社区数据的采集。汽车论坛为消费者之间以及消费者和制造商之间搭建了一座沟通的平台,购车用户可以在汽车论坛上分享自己的购车和驾驶体验。通过汽车之家网站提供的车辆信息,获取新能源汽车价格、能源类型、车型、续航里程、充电方式、百公里加速时间、电池类型、电动机功率、车身重量等一系列相关属性参数。以汽车之家新能源汽车活跃度前五的品牌 Tesla、奥迪、比亚迪、蔚来和理想为例,爬取了汽车网络社区当中购车用户的口碑评论及自由评论。

(2)汽车网络社区数据的清洗。汽车之家是一个开放性的网络交流平台,其中的文本评论数据中可能存在部分价值含量低甚至无价值的数据。为避免这些问题对后续研究分析造成干扰,需要对原始数据进行清洗,处理后的数据才能用于分词、词频统计以及进一步的主题挖掘和情感分析等工作,并产生准确可靠的分析结果。数据清洗是进行数据挖掘的前提条件,只有经过清洗后的数据才能产生准确可靠的分析结果。

除爬取用户评论数据外,还需对于消费者购买的新能源汽车产品的部分性能参数进行爬取。通过网络爬虫爬取了汽车之家新能源板块论坛用户的发帖数据共计 1 200 余条,时间为 2022 年 1 月 3 日至 2024 年 1 月 2 日。数据包括用户发帖时间、发帖标题、回复内容等。经过数据处理及数据清洗,爬取的网络评论数据和相关属性参数数据将以.csv 格式保存下来。

2.2 Jieba 分词、分句与数据预处理

评论数据是对非结构化数据,通过使用 Jieba 中文分词库对影评数据进行分词、去掉停用词等处理,得到比较规范的车评数据,并进行可视化分析及语义分析。由于中文表达的习惯不同,消费者在评论中针对评价指标的描述与对其表达情感倾向可能分布在不同部分。运用回归分析等技术手段,将经过中文分词处理过的车评数据通过调用 WordCloud 库进行词云图可视化呈现;按照车评数据的高频关键词和权重不同,使用 TF-IDF 算法对车评数据进行提取和计算,对用户给出的汽车指标评分等数据形成结构化数据,以便分析用户评论及汽车属性的信息特征。

研究对象是汽车之家网站新能源汽车的评论数据,在此部分,依据评论信息自动划分成好评、差评,并对产品的特征、品质的优缺点进行分析,不需要过于精确地计算每条评论的情感指数。在此阶段,从匹配情感词、修正情感倾向、SnowNLP 情感分析三部分展开阐述,将情感词表与 Jieba 分词后的评论数据相匹配,得出初始的情感系数。基于情感词典的 SnowNLP,利用朴素贝叶斯模型,在已知评论样本的信息情形下,推测其为好评的概率。

3 基于 LDA 的主题模型的构建与主题提取

3.1 模型参数设置与优化

3.1.1 模型参数的设置

在研究中,LDA 模型的参数设置至关重要。经过多次实验和验证,最终确定了参数:主题数 K 设置为 10,代表预期从评论数据中提取的主题数量。此设置平衡了模型的复杂性和可解释性,能捕捉到足够的主题模式;最小文档频率阈值设定为 5,意味着在所有文档中出现次数少于 5 次的词语将被过滤掉,减少了噪声词语的干扰,提高了模型的稳定性和准确性;最大迭代次数设为 1 000 次,确保算法有足够的时间进行收敛,避免局部最优解的出现。

3.1.2 模型的训练与验证

使用预处理后的数据集对 LDA 模型进行初始化;通过不断调整模型参数来优化主题的提取效果;每次迭代过程中,模型会根据当前的参数设定重新分配词语到不同的主题,并计算每个主题的词语分布概率。通过对比每次迭代的困惑度指标,监控模型在验证集上的表现,确保其具备良好的泛化能力。一旦困惑度指标不再显著下降或达到最大迭代次数,训练过程停止。此时,模型已充分学习到数据中的隐藏主题结构,可用于后续的解释。

3.2 主题的提取

通过对训练好的 LDA 模型进行分析,得到了 10 个主要的主题分布,涵盖了新能源汽车用户评论的多个方面,包括车辆性能、外观设计、内饰配置、续航里程、充电便利性、售后服务、性价比等。每个主题下的词语分布反映了用户在相应方面的主要关注点和讨论热点。例如,在车辆性能主题下,高频词语包括加速、操控、动力系统等;在外观设计主题下,则有美观、时尚、车身线条等词语频繁出现。为更直观地理解每个主题的内容,从每个主题中提取概率较高的词语作为该主题的代表。例如,在续航里程这一主题中,提取到的关键词包括电池、续航、充电时间等;在售后服务主题中,关键词则有保养、维修、服务态度等。这些关键词,有助于解释各个主题的含义,还能为后续的用户偏好分析和策略制定提供有力支持。

4 利用 LDA 主题模型的文本数据主题分析

利用 LDA 模型得到的主题分布得出 LDA 模型的模型分析结果,包括用户在社区中讨论的主题和话题。LDA 模型是一种无监督的贝叶斯模型,其核心公式可以表示为:

$$P(\text{词}|\text{文档})=P(\text{词}|\text{主题})|P(\text{主题}|\text{文档})$$

无监督 LDA 主题模型算法是提取新能源汽车评论文本中的主题,在 LDA 主题模型中,每个文档主题的概率分布 $P(z|d)$ 和每个主题中单词出现的概率 $P(w|z)$ 被视为按照 Dirichlet 分布概率随机抽取的变量。对于给定文档的集合,假定存在 K 个主题。LDA 主题模型基于超参数 β 的狄利克雷分布可生成 K 个多项式分布,每个分布表示整个词典中的词汇在该主题下的随机生成概率,公式如下:

$$\begin{aligned}\theta_i &\sim \text{Dirichlet}(\beta) \\ i &= 1, 2, \dots, K \\ \theta_i &\in R^w, \beta \in R^w\end{aligned}\quad (1)$$

式中, ω 是整个词典的规模; θ_i 和 β 是维度为的实数向量; θ_i 是主题 K 的概率分布对应的数值向量。

β 包含了各个维度上单词出现概率的先验信息, 通常情况下假定这些概率值相等。不同维度上的绝对值与先验知识的相对重要性相关, 而各个维度上的相对值则决定了单词在每个维度上出现概率的大小。结合人工监督的训练集, 利用 BERT 神经网络对每一条用户评论的细粒度的情感分析, 包含消费者评论中提及的新能源汽车属性

标签以及消费者的情感强度。根据 BERT 神经网络情感分析的结果建立模型计算消费者对各个主题的满意度和关注度, 利用 IPA 模型对结果进行可视化分析。

预设若干个主题, 通过主题模型挖掘新能源汽车用户讨论的话题, 建立对产品销量影响的计量经济模型, 得出相应的结论, 识别用户对产品或服务的看法和需求, 从而帮助新能源车企改善产品定位和市场策略。数据采集情感可视化分析流程如图 1 所示。

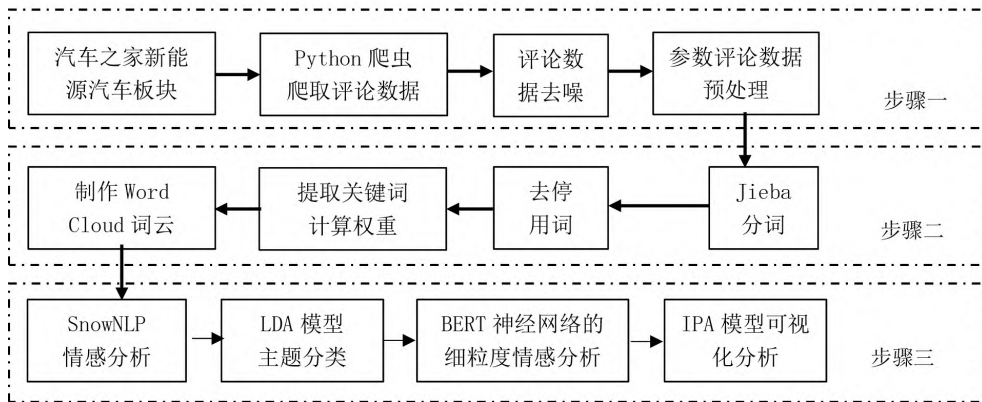


图 1 数据采集情感可视化分析流程

5 用户偏好分析与消费趋势预测

5.1 空间舒适性

新能源汽车的网络社区评论中, 用户普遍关注车内空间的宽敞程度以及座椅的舒适性。车内部空间和较高的座椅位置, 受到了许多家庭用户的青睐。此外, 一些高端品牌通过提供豪华的座椅材质和多种调节功能, 进一步提升了乘坐体验。用户对这些细节表现出高度敏感, 并在评论中给予积极反馈。

5.2 外观与空间设计

车辆的外观设计往往影响用户的初印象和购买决策。流线型车身、独特的前脸设计和高品质的漆面处理是用户关注的重点。例如, 蔚来汽车以精致的做工赢得了许多年轻消费者的喜爱。此外, 车内空间的布局和使用效率也是用户评论的热点话题。

5.3 续航与动力性能

续航里程和动力性能是新能源汽车用户最为关心的技术指标之一。特斯拉凭借其领先的电池技术和高效的能耗管理, 持续占据领先地位, 用户对其续航表现给予了高度评价; 快速加速和强劲的动力输出是许多用户追求的目标。蔚来 ES8 凭借其强大的双电机四驱系统, 提供了卓越的动力体验, 受到了性能爱好者的青睐。

5.4 内饰与智能互联

高品质的内饰材料和精细的做工能显著提升用户的满意度。奔驰 EQC 凭借其豪华的内饰设计和一流的制造工艺, 在用户评论中获得了极高的评价。智能互联功能方面, 用户期望车辆配备大尺寸触控屏、语音助手、在线导航等现代化配置。蔚来汽车在这方面表现尤为突出, 其 NOMI 人工智能助手和智能座舱系统深受用户喜爱。

5.5 经济性与服务体验

经济性包括购车成本, 后期的维护费用和使用成本。

比亚迪通过高性价比的车型和经济实用的维修保养服务赢得了用户的认可; 用户期望经销商提供及时、专业的服务。特斯拉的创新直销模式消除了中间商环节, 为用户提供了全新的购车体验, 但也因服务网络不完善而受到部分用户的批评。

6 结论

通过对新能源汽车网络社区用户评论的数据进行爬取与预处理采用 LDA 主题模型进行深入分析, 揭示了用户在选购新能源汽车时关注的主要偏好及其背后的消费动因, 研究发现空间舒适性、外观与空间设计、续航与动力性能、内饰与智能互联以及经济性与服务体验, 是用户评论中的五大核心主题。续航与动力性能是最受关注的因素, 直接影响用户的购车决策。分析新能源汽车用户的消费偏好, 提供了详实的数据支持, 为企业优化产品策略和市场营销提供了有价值的参考依据。

参考文献:

- [1] 贺畅, 张衡, 张帆. 我国新能源汽车消费者特征和需求洞察[J]. 现代商业, 2021(10): 3-5.
- [2] 吕振杰, 林益成, 陈静娜. 基于消费者偏好的新能源汽车产业发展对策研究[J]. 时代汽车, 2023(23): 163-165.
- [3] 汪宸瑞. 新能源汽车消费者购买意愿影响因素研究[J]. 商场现代化, 2023(24): 14-16.
- [4] 李素丽. 新媒体环境下我国汽车行业营销模式的创新与突破[J]. 科技资讯, 2021(16): 82-85.
- [5] 陈青, 邓皓元, 张仁寿. 基于 LDA 算法的新能源汽车政策舆情分析[J]. 广州大学学报(自然科学版), 2019(05): 34-42.