

文章编号:1005-9679(2025)02-0043-05

# 基于数据挖掘的新能源汽车用户感知研究

曾若辰<sup>1</sup> 石鼎鑫<sup>2</sup>

(1. 上海大学悉尼工商学院, 上海 201899; 2. 上海浦东发展银行, 上海 200120)

**摘要:**近年来新能源汽车销量呈现爆发式的增长,但是日益增长的新能源汽车行业也导致了诸多问题,例如产品质量参差不齐、用户体验不佳等问题。因此,在竞争日益激烈的市场中如何及时准确地掌握消费者需求,了解消费者对产品各方面的评价信息,为消费者带来更好的产品和服务至关重要。传统的问卷调查等方式获取消费者的需求和质量评价显然已经不能满足当前的需求。本文通过数据挖掘技术分析新能源汽车市场信息,探究用户需求与痛点。研究发现,用户对新能源汽车的关注点主要集中在性价比、内饰、空间等方面。同时,研究也发现用户对于新能源汽车的外观等方面满意度普遍较高,对于舒适性等方面的满意度普遍较低。

**关键词:**新能源汽车;数据挖掘;用户感知

**中图分类号:** F 724.6 **文献标志码:** A

## Research on User Perception of New Energy Vehicles Based on Data Mining

ZENG Ruochen<sup>1</sup> SHI Dingxin<sup>2</sup>

(1. SILC Business School, Shanghai University, Shanghai 201800, China; 2. SPD Bank, Shanghai 200120, China)

**Abstract:** Electric vehicle have shown explosive growth in recent years, but the growing electric vehicle industry has also led to many problems, such as uneven product quality and poor user experience. Therefore, in the increasingly competitive market, how to timely and accurately grasp the consumer demand, understand the consumer evaluation information on various aspects of the product, to bring better products and services for consumers is crucial. Traditional methods such as questionnaire surveys to obtain consumer demand and quality evaluation obviously can no longer meet the current needs. In this paper, the market information of electric vehicle was analyzed through data mining technology to explore user needs and pain points. The study finds that users' concerns about electric vehicle are mainly focused on cost-effectiveness, interior, space and other aspects. At the same time, the study also finds that users are generally more satisfied with the appearance of electric vehicle, and generally less satisfied with the comfort.

**Key words:** electric vehicle; data mining; user perception

## 0 引言

本文利用汽车网站中新能源汽车用户的在线

评论,从消费者的角度出发,基于产品的不同维度获取用户评论。通过聚类算法,将具有相似内容和特征的评论归为一类,从而发现用户关注的主题。

收稿日期:2025-01-18

作者简介:曾若辰(1991—),男,黑龙江哈尔滨人,上海大学悉尼工商学院讲师,研究方向:智能建造;石鼎鑫(2001—),男,湖南吉首人,上海浦东发展银行职员。

进一步针对确定的主题,构建情感分类模型,并得出了用户对新能源汽车感知不同维度的满意程度以及关注程度。同时,通过对满意度和关注度的相关性分析,研究揭示了消费者在购买新能源汽车过程中,可能面临的挑战和痛点。

## 1 文献综述

针对新能源汽车用户偏好和用户感知方面的研究,国内外学者进行了广泛的调查和实证分析。Bennett等(2016)对影响新能源汽车推广的因素进行过探究;Jena(2020)基于印度电动汽车消费者数据,运用机器学习等方法进行了情感分析;Ferguson等(2018)利用加拿大消费者的调查结果,探究了消费者对电动汽车的态度和偏好程度;Singh等(2020)对现有文献进行了元分析,进一步研究了购买新能源汽车的影响因素;余帆(2022)从8个维度出发,针对国内用户评论进行了情感分析,研究发现,消费者对新能源汽车的购买意愿受到多种因素影响,如价格、续航里程、充电设施、环保意识等。

为深入挖掘用户感知,不同研究采用情感分析和用户画像等方法对在线评论进行分析。情感分析主要通过自然语言处理技术,对用户评论中的情感倾向进行量化分析。用户画像是通过挖掘用户评论中的关键信息,构建用户特征模型,以便更精确地了解用户需求和喜好。通过数据挖掘技术,可以从海量数据中获取描述用户的标签,这些用户标签被广泛应用于推荐系统和产品运营(张彩红等,2023)。

传统汽油车用户研究经过多年的积累和发展已经形成了比较完善的评价体系,例如中国汽车技术研究中心有限公司从2012年起已经连续7年开展汽油车消费者满意度调研工作,从产品、销售、售后三个方面,进行评价指标下沉,形成了完备的测量用户满意度的指标体系和研究模型。然而,新能源汽车与传统汽油车在产品特性、使用情境、维修保养等方面存在着显著差异。新能源汽车面临着不同的市场环境和用户群体。根据消费者行为学的理论,不同的用户群体对产品满意度的评判方式也不尽相同。因此,对于新能源汽车用户满意度的影响因素以及满意度的影响机制,仍需要进一步深入研究。

过去的研究较多依赖于传统的问卷调查和访谈等方式,而较少利用在线评论等大数据方法。此外,现有研究主要关注购买前消费者行为的影响因素,而较少对购买后行为和消费者满意度的影响因

素进行研究。相比之下,基于网络评论的文本挖掘研究相对较少,即使在研究消费者满意度的影响因素时,也很少对这些因素进行排序,以找出它们的重要程度。本文基于在线评论挖掘,深入研究新能源汽车用户感知、用户偏好、情感分类等方面。

## 2 研究方法

### 2.1 数据采集与预处理

本文采用的数据来自头部汽车垂直门户网站的汽车用户口碑数据。用户可以在该网站上查找汽车相关的新闻、信息、评测以及购车指南等内容。同时,该网站也提供了用户口碑评论的功能,让用户可以分享自己的购车、使用经验以及对特定汽车的评价和看法。该网站的用户口碑评论有以下特征:用户口碑评论数量庞大,涉及的汽车品牌和车型繁多,覆盖了几乎所有的主流汽车品牌和车型;用户口碑评论包含的内容多种多样,包括购车经验、车辆性能、外观设计、内饰配置、油耗表现、车辆故障、售后服务等方面。

为了收集新能源汽车的用户评论,本文采用爬虫技术,从目标网站收集相关数据。我们设置目标车型ID,以便针对特定新能源汽车品牌和型号抓取用户评论,目标涉及市面主流新能源汽车品牌与型号。

本文共采集到16900余条关于新能源汽车的用户评论,包括用户对于各个新能源车型的各个车型的具体评价。数据内容主要是“车系—满意—评论详情、不满意—评论详情、空间—评论详情、驾驶感受—评论详情、续航—评论详情、外观—评论详情、内饰—评论详情、性价比—评论详情、操控—评论详情、舒适性—评论详情、智能化—评论详情”。其中“最满意—评论详情”与“最不满意—评论详情”代表了用户对于该款车型最强烈的正面与负面评价。而“空间—评论详情、驾驶感受—评论详情、续航—评论详情、外观—评论详情、内饰—评论详情、性价比—评论详情、操控—评论详情、舒适性—评论详情、智能化—评论详情”则是用户对于车系这些具体方面的评价。

对数据进行预处理主要包括分词、去除空值和选取停用词。首先是去除一些常见的停用词,比如数字、符号,同时还包括一些无意义的语气词和一些不包含信息量的词汇,比如“新能源”“车子”等,因此将这些无关的主题词加入停用词。根据采集的数据集特性,数据标注主要分为两类:第一类是

标注为“最满意”和“最不满意”的评论数据;第二类是标注为“空间、驾驶感受、续航、外观、内饰、性价比、操控、舒适性和智能化”这些指标的评论数据。通过使用第二类标签的数据集训练情感分类模型,然后使用训练好的模型对标注“满意”与“不满意”两个标签的口碑数据进行分类,得到用户对新能源汽车各个方面的具体满意程度以及关注程度。

## 2.2 主题聚类及情感分类模型

LDA(Latent Dirichlet Allocation)是一种基于概率分布的主题模型,主要用于从大量文档中提取潜在主题。LDA的基本思想是认为文档是由多个主题组成的,而每个主题则是由多个词汇组成的。通过LDA模型,可以获得文档中的潜在主题。本文使用LDA聚类模型对标注为“满意/不满意”的在线评论数据进行聚类,通过主题词判断需要研究的指标。LDA模型可以有效地帮助发现用户评论中的关键主题,从而指导后续的情感分析研究。

在情感分析中,本文根据主题词筛选出相关评论,然后利用机器学习方法训练模型,以实现评论情感的自动分类。我们采用三种模型包括朴素贝叶斯、TextCNN和MLP(多层感知器)来进行预测,最后通过投票分类的策略得到最终结果。朴素贝叶斯分类器是基于贝叶斯定理的一种分类方法,具有简洁、训练速度快的特性。TextCNN是一种基于卷积神经网络(CNN)的文本分类模型,TextCNN能够自动捕捉文本的局部特征,提取有用的信息,降低了特征工程的复杂性并且其训练速度较快,适用于大规模数据集。MLP是一种多层感知机模型,属于前馈神经网络的一种。MLP能够捕捉数据中的非线性关系,适用于复杂问题,并且通过增加网络层数和神经元数量,可以更准确地挖掘评论中的潜在信息。为了综合利用三个模型的优点,本文采用了投票分类的策略,将三种模型预测结果作为投票分类器的输入得出最后的结果。投票分类是一种集成学习方法,它依赖于不同的机器学习模型进行预测,并根据所有模型的预测结果进行投票。具体步骤是将每一条评论都使用三个模型进行预测,三个模型会分别给出每条评论属于每个分类维度的概率,然后将三个模型给出的概率加权平均,就能得出最终的每条评论属于每个分类维度的概率。通过采用投票分类的方式,我们可以充分利用各种模型的优势,避免依赖单一模型可能带来的风险,从而在新能源汽车用户感知的研究中获得更准确和全面的结果。

## 2.3 模型构建

本文使用Gensim库构建词典和语料库训练LDA模型。为了确定最佳的主题数量,我们采用了一致性评分来评估不同主题数量下LDA模型的表现。尝试了从1到10的主题数量,并计算了对应的一致性评分。这个过程涉及了对每种主题数量下的LDA模型进行训练,并使用CoherenceModel计算对应的一致性评分。通过图1一致性评分与主题数量之间的关系,可发现当主题数量为9时,一致性评分达到最高。

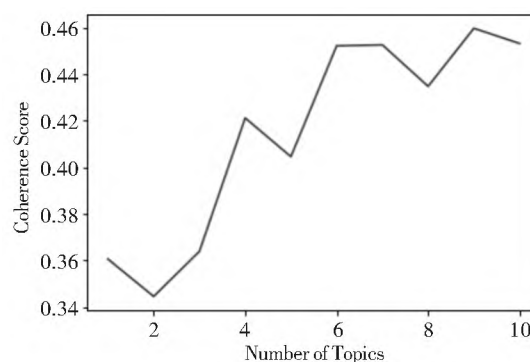


图1 一致性评分曲线

### 朴素贝叶斯分类模型构建

本文采用CountVectorizer将文本转换为词频矩阵表示。CountVectorizer会计算每个单词在文档中出现的次数,并生成一个稀疏矩阵。这种表示方法可以有效地将文本数据转换为适合朴素贝叶斯分类器处理的数值格式。为了评估模型的性能,我们将数据集划分为训练集(80%)和测试集(20%)。为确保训练集和测试集类别分布与总体相似,我们采用了分层抽样的方法进行划分。对于超参数的选取,通过分析训练集和测试集的准确率随alpha变化的曲线,我们确定了最佳的alpha值为0.3。基于选定alpha值,构建MultinomialNB分类器,并使用训练集对其进行了训练。

### TextCNN分类模型构建

通过构建词汇表,将每个词映射为一个唯一的整数值。词汇表的构建是通过统计所有文本中出现的词及其频率来实现的。进一步定义了一个将文本转换为固定长度整数序列的函数,以便将其输入到TextCNN模型中。在本研究中,我们将20%的数据划分为测试集,剩余的80%用于训练。TextCNN模型的关键超参数包括词向量维度、类别数、卷积核数量、卷积核尺寸以及Dropout比率。分别设置这些超参数为128、9、100、[3, 4, 5]和0.5。此外,我们还设置了学习率为0.001、批处理



大小为64以及训练轮数为10。在TextCNN类的构造函数中,我们创建了词嵌入层、卷积层列表、Dropout层和全连接层。根据上述超参数构建了TextCNN模型。图2展示了训练模型的准确率。

#### MLP分类模型构建

使用TF-IDF(Term Frequency-Inverse Document Frequency)特征提取方法将文本转换为稀疏矩阵表示。TF-IDF是一种常用于文本挖掘的特征权重计算方法,它考虑了词汇在文档中的频率以及在整个语料库中的逆文档频率。这种表示方式可以有效地将文本数据转换为适合多层感知机(MLP)分类器处理的数值格式。将20%的数据划分为测试集,剩余的80%用于训练。本文使用sklearn库构建了一个MLP分类器,其中包含两个隐藏层(分别包含64个和32个神经元),激活函数为ReLU,优化器为Adam。在训练过程中,我们还使用了10%的验证集来监控模型性能并防止过拟合。

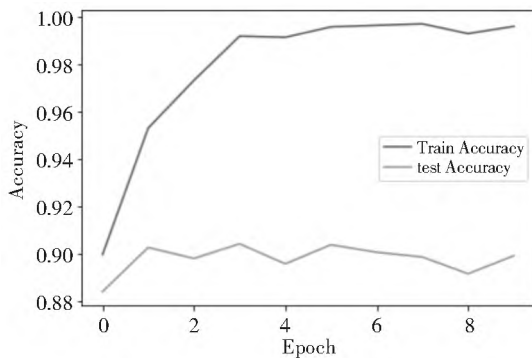


图2 模型准确率曲线

#### 投票分类器构建

首先加载预训练好的朴素贝叶斯、TextCNN和MLP模型,以及它们对应的特征提取器。接着,对于每个输入的文本,我们分别用三个模型进行预测,并得到它们的分类概率。然后将这三个模型的概率结果加权平均,得到最终的概率分布。这种方法能够充分利用不同模型的特点,增强整体分类性能。

此外,在“满意”与“不满意”的评论数据当中,未必只设计“空间、驾驶感受、续航、外观、内饰、性价比、操控、舒适性和智能化”当中的单独一类,同一条“满意”或者“不满意”的评论当中,可能同时涉及“空间、驾驶感受、续航、外观、内饰、性价比、操控、舒适性和智能化”当中的多个类型。而传统的单标签分类方法可能无法充分捕捉这种多样性。为了更准确地反映评论中包含的各个方面,我们决定采用基于概率的分类方法。在分类过程中,我们设置了0.3的阈值。如果数据落在一个类当中的概

率超过0.3,则归为这个类。也就是说一个数据可能被归为多个类,在计算不同类的满意度时,这样的数据会被重复计算。

## 3 结果与讨论

### 3.1 聚类结果

通过构建LDA模型,我们将评论数据聚类为不同的主题并获取了相应的主题词。通过对主题词的研究与归纳,分析了主题词与评论类型的内在逻辑联系,同时抽选了数条评论进行对比。我们发现这些主题词可以与特定类型的评论相对应。表1显示了聚类主题词与对应评论类型。新能源汽车用户在撰写在线评论时,主要关注车辆的空间、驾驶感受、续航、外观、内饰、性价比、操控、舒适性和智能化等方面。这些方面反映了用户在评价新能源汽车时的核心关切点。

表1 聚类主题词与对应评论类型

主题词	对应评论类型
空间、后排、后备箱	空间
高速、动力、驾驶	驾驶感受
续航、充电、公里、油耗	续航
外观、设计、材质	外观
内饰、设计、材质	内饰
优惠、价格、性价比	性价比
底盘、停车、刹车	操控
隔音、座椅、噪声、乘坐、味道	舒适性

### 3.2 分类结果

图3展示了各分类维度的总体满意度。

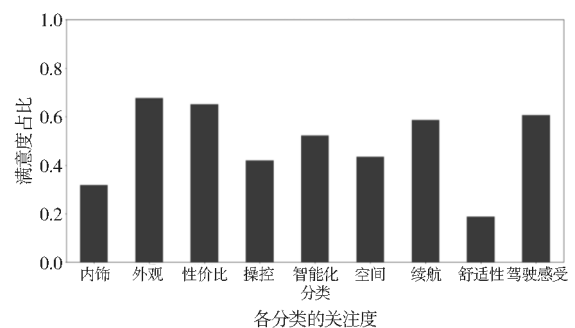


图3 各分类维度总体满意度

在所有车系的综合评价中,外观和性价比方面的总体满意度最高。这意味着消费者对新能源汽车的外观设计大体上感到满意,新能源汽车通常具有独特的设计和形象,相比传统燃油车更显新颖和个性。性价比的高满意度则表明,消费者普遍认为新能源汽车在价格、配置和性能方面达到了较好的平衡。紧随其后的是驾驶感受,得益于新能源汽车的加速性能、稳定性等因素,消费者在驾驶新能源

汽车时体验较满意。续航和智能化的满意度也相对较高,反映了新能源汽车在这些领域取得的进展符合消费者期望。

然而,操控、空间、内饰、舒适性等方面的满意度较低,其中舒适性的满意度最低。这意味着新能源汽车在舒适性方面存在相当大的提升空间。对于新能源汽车来说,座椅舒适度、噪声控制和乘坐空间等因素的表现显然还有待提升。同时,内饰和空间的满意度也相对较低,这可能说明消费者对新能源汽车在这些方面的表现有更高的期待。

图4展示了各分类的关注度,关注度按分类评论的个数来判别。消费者在评价新能源汽车时,关注度的排名如下:性价比、内饰、空间、续航、外观、智能化、舒适性和驾驶感受,最后是操控。

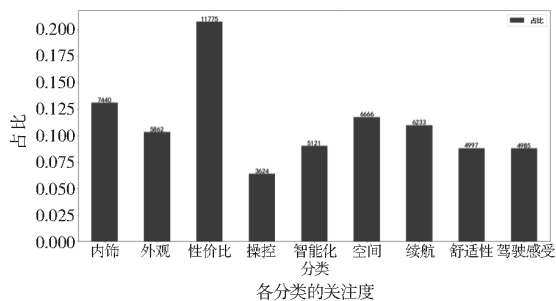


图4 各分类关注度

性价比是消费者在评价新能源汽车时最为重视的部分。其次是内饰,消费者对新能源汽车的内部装饰和设计有着显著的关注。接下来是空间,消费者关心新能源汽车的内部空间,这包括乘坐空间和储物空间。足够的空间可以提供更高的舒适度和实用性,无论是日常通勤还是长途旅行,都能满足消费者的需求;续航能力也是消费者的一大关注点,尤其是对于电动汽车来说,足够的续航里程能减少充电次数,提高使用便利性。同时,消费者也关心充电设施的便利程度,包括充电桩的分布、充电速度等。舒适性和驾驶感受的关注度接近,消费者希望新能源汽车在行驶过程中提供稳定且舒适的驾驶体验。操控性是消费者关注最低的一项。

综合以上分析,我们可以看出,虽然消费者在选择新能源汽车时关注的重点各有不同,但他们共同关心的是汽车的整体性能和价值。

## 4 总结

本文采集并分析了新能源汽车用户在线评论

数据,利用了LDA聚类模型、朴素贝叶斯、TextCNN和MLP等多种情感分类模型,旨在深入挖掘消费者对新能源汽车不同方面的满意程度以及对新能源汽车不同方面的重视程度。

首先,通过LDA主题聚类,研究发现用户的关注度主要集中在新能源汽车的性价比、内饰、空间和续航等方面,而对于操控等方面的关注度相对较低。

其次,通过用户的评价满意度分类,研究发现新能源汽车的各个方面中,外观获得了最高的满意度,而舒适性的满意度最低。这意味着当前的新能源汽车市场在外观设计方面已经满足用户需求,但在舒适性方面仍有很大的提升空间。

本研究揭示了消费者对新能源汽车的真实需求和期望,为汽车制造商提供了有益建议。此外,本研究还提供了解市场动态和消费者行为变化的重要途径。通过对新能源汽车用户评论的情感分析,揭示了消费者对新能源汽车不同方面的满意程度以及对新能源汽车不同方面的重视程度。这些发现对于汽车制造商、政府和企业制定更加合理有效的产品改进策略、市场推广策略以及政策支持措施具有重要指导意义。

## 参考文献

- [1] 涂宇,牛盼强.碳中和背景下上海车展助力汽车产业新能源化路径研究[J].上海管理科学,2023,45(01):50-53.
- [2] BENNETT R, KOTTASZ R, SHAW S. Factors potentially affecting the successful promotion of electric vehicles [J]. Journal of Social Marketing, 2016.
- [3] JENA R. An empirical case study on Indian consumers' sentiment towards electric vehicles: A big data analytics approach [J]. Industrial Marketing Management, 2020, 90: 605-616.
- [4] FERGUSON M, MOHAMED M, HIGGINS C D, et al. How open are Canadian households to electric vehicles? A national latent class choice analysis with willingness-to-pay and metropolitan characterization [J]. Transportation Research Part D: Transport and Environment, 2018, 58: 208-224.
- [5] SINGH V, VAIBHAV S. A review and simple meta-analysis of factors influencing adoption of electric vehicles [J]. Transportation Research Part D: Transport and Environment, 2020, 86: 102436.
- [6] 余帆.基于文本挖掘的新能源轿车用户情感分析[J].物流工程与管理,2022,44(1):137-140.
- [7] 张彩红,李想,周雪侠,等.基于情感分析的突发公共卫生事件微博评论用户画像构建[J].医学信息学杂志,2023,44(3): 18-22.