

# 基于文本挖掘技术的农村新能源汽车购买决策影响因素分析

苗昌盛

重庆交通大学经济与管理学院, 重庆

收稿日期: 2023年10月30日; 录用日期: 2023年11月17日; 发布日期: 2024年1月4日

## 摘要

为促进新能源汽车产业发展和环境保护, 中国政府实施了新能源汽车下乡政策, 鼓励农村地区消费者购买新能源汽车。本文旨在探讨影响中国农村地区电动汽车采用的因素, 采用了TF-IDF算法对农村消费者购车评论的特征词进行提取和词频统计, 以分析消费者的购车决策因素。通过对农村电动汽车消费者的复合评论进行情感波动的分析, 不仅可以拓展消费者感知价值理论的研究深度, 还可以为厂商和消费者提供内容情报和数据参考, 协助他们调整生产行为和优化购买决策。研究结果表明, 在农村市场, 消费者更注重汽车价格、空间、维修成本、驾驶感受和油耗等因素, 新能源汽车占有量与燃油车价格呈正相关, 并受到羊群效应的影响。研究结果对于推动新能源汽车在农村市场的推广和普及具有重要意义。

## 关键词

新能源汽车, TF-IDF算法, 农村地区

## Analysis on Influencing Factors of Rural New Energy Vehicles Purchase Decision Based on Text Mining Technology

Changsheng Miao

School of Economics and Management, Chongqing Jiaotong University, Chongqing

Received: Oct. 30<sup>th</sup>, 2023; accepted: Nov. 17<sup>th</sup>, 2023; published: Jan. 4<sup>th</sup>, 2024

## Abstract

In order to promote the development of new energy vehicle industry and environmental protection, the Chinese government has implemented a policy of new energy vehicles going to the coun-

tryside to encourage consumers in rural areas to buy new energy vehicles. The purpose of this paper is to explore the factors that affect the adoption of electric vehicles in rural areas of China. TF-IDF algorithm is used to extract the characteristic words of rural consumers' car purchase comments and make word frequency statistics, so as to analyze the factors of consumers' car purchase decision. By analyzing the emotional fluctuation of rural electric vehicle consumers' composite comments, we can not only expand the research depth of consumer perceived value theory, but also provide content information and data reference for manufacturers and consumers to help them adjust their production behavior and optimize their purchase decisions. The research results show that in rural markets, consumers pay more attention to factors such as car price, space, maintenance cost, driving experience and fuel consumption. The research results are of great significance for promoting the promotion and popularization of new energy vehicles in rural markets.

## Keywords

New Energy Vehicles, TF-IDF Algorithm, Rural Areas

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>

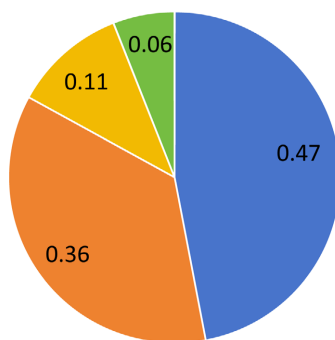


Open Access

## 1. 引言

面对全球气候变化和环境污染的严峻挑战，绿色消费和绿色低碳生活和生产方式受到国际社会的高度重视。在中国共产党第二十次全国代表大会上，中国提出了“碳达峰、碳中和”的目标，并把新能源汽车列为优先发展的战略性新兴产业之一。新能源汽车是指以新型动力系统为主动动力来源并符合节能环保要求的汽车，相比传统汽车，它们具有节能减排、降低油耗、减少噪音等优势。中国过去几年大力推广新能源汽车，以实现减排和高质量发展目标。

■ 一线城市 ■ 二线城市 ■ 三四线城市 ■ 四线以下



**Figure 1.** Sales locations proportion of new energy vehicles in 2022  
**图 1.** 2022 年新能源汽车销售地占比

根据中国工业和信息化部发布的数据，2022 年中国新能源汽车销量达到 300 万辆，占全球市场份额的 40% 以上。然而，在市场规模不断扩大的同时，也存在着一些问题和挑战。其中之一就是新能源汽车市场分布极不均衡，见图 1。从中国 2022 年新能源汽车购买用户地区分布情况来看，其主要集中在东部

沿海地区和一些经济发达城市，而中西部地区和农村地区则占比很低。这种分布不均既反映了农村地区存在着巨大的潜力和机会，也暴露了农村地区面临着诸多困难和障碍。2019年10月1日起，《中华人民共和国促进新能源汽车发展管理办法》正式实施，在全国范围内推行新能源汽车下乡政策，鼓励农村地区消费者购买新能源汽车，并给予相应补贴。

然而，在农村市场上推广新能源汽车面临着诸多挑战。一方面，农村消费者对于新能源汽车的认知和信任水平较低，他们对汽车的评价标准仍然以性价比、安全性、续航里程等因素为主，并对新能源汽车在这些方面的表现持有较大疑虑；另一方面，政策激励不足，补贴制度不规范。因此，在制定有效政策之前，有必要深入了解影响农村地区消费者购买新能源汽车意愿和行为选择的因素。本文基于以上背景与目标，在综合前人研究成果的基础上进行深入研究与分析，并采用以下方法：首先对2022年中国新能源汽车购买用户地区分布情况进行了统计和解读，其次，以农村消费者购买电动汽车的复合评论为样本，采用文本情感分析法探索消费者情感波动特征及其影响因素，有助于拓展消费者感知价值理论研究深度，也可以为厂商调整生产行为、消费者优化购买决策提供内容情报和数据参考。

## 2. 文献综述

与本文密切相关的文献包括文本挖掘技术和农村电动汽车购买行为的影响因素和特征两个方面。

与本文首要相关的主题是农村新能源汽车购买的影响因素。以往关于电动汽车采用的研究通常从确定影响电动汽车采用或渗透的因素开始。在城市地区，影响电动汽车采用或普及的因素已经得到了广泛的研究，包括经济、技术和心理方面的因素。Cecere [1]等人认为，高采购成本是电动汽车普及的障碍，而较低的能源成本是电动汽车使用的优势。此外，Li [2]等人认为每次充电里程有限、充电站不足、电池保修不够等技术问题也是主要障碍之一。Morton [3]等人关注了消费者对环境问题、新产品风险和創新行为等心理因素的影响。然而，在农村地区，影响电动汽车采纳的因素却鲜有研究[4]。城市和农村地区在这方面既有相似之处，也有不同之处。Zhao [5]等人认为农村消费者和城市消费者一样，会考虑买车和用车的花费，这些因素会影响他们是否选择电动汽车。Nazari 等[6]提出采用的增加通常意味着消费者对电动汽车的认可和购买意愿的提高，以及政府和企业对电动汽车的推广和支持措施的有效性。Yongqing 等[7]指出在当前技术下，农村电动汽车采用的主要障碍之一消费者对新能源汽车的认知和购买意愿处于较低水平。上述文献研究了电动汽车在农村推广受到的广泛因素，但少有文献研究农村消费者关注新能源汽车本身的因素。

文本挖掘技术(Text Mining)涉及通过分析评论文本内容来研究商品、服务、人物等对象，从而总结和推导评论者的态度和意见[8]。当前，文本情感分析主要采用基于情感词典与规则的方法、机器学习方法、深度学习方法以及多策略混合方法。基于情感词典与规则的方法利用情感词典和语义规则来获取文本中情感词的情感值，并通过计算语义相似度来确定文本的情感倾向[9]。分析促进农村新能源汽车购买影响因素的一种有效方法是从现有文本数据中发现知识。例如，郭和刘[10]利用社会网络分析，在收集互联网信息的基础上构建了中国新能源汽车产业联盟网络，这一过程有时被称为“网络学”[11]。Wang 等人[12]借助包含车型和生产信息的汽车数据库揭示了政府补贴对新能源汽车市场表现的影响。Wang 和郭的研究处理的都是相对结构化的数据，即数字和预先确定的公司名称。

为了处理结构化程度较低的信息，如自然语言，研究人员使用内容分析或文本挖掘。例如，谢和田[13]从37个政策文本中手动分类了新能源汽车政策工具。

Sung 和 Park [14]利用自动词频和影响分析工具 TF-IDF，在2.65亿篇有关可再生能源的在线新闻和文件中，发现了高影响词。此外，Zeng [11]还利用话题识别方法潜狄利克雷分配(Latent Dirichlet Allocation, LDA)算法作为可再生能源行业的预测工具。更高级的文本聚类方法包括向量空间模型[15]、带有遗传算

子的混合粒子群优化算法[16]和磷虾群算法[17]。

基于新能源汽车研究对中国新能源汽车下乡的重要意义，以及分析由自然语言组成的农村消费者评论数据文本，本研究旨在通过文本挖掘的方法揭示中国农村新能源汽车购买行为影响因素研究。

### 3. 研究设计

#### 3.1. 数据来源及分析

本部分数据来自懂车帝网站口碑栏目中的新能源汽车复合评论和对农村电动汽车问卷调查的评论作为研究数据来源进行消费者情感波动分析。其中懂车帝中获取的数据仅收集懂车帝网站认证通过车主发表的复合评论，并且通过关键词提取出仅在城乡使用的评论词条。本文运用 Python 语言编程采集懂车帝网站新能源汽车消费者复合评论获得 2016 年 4 月至 2022 年 12 月间电动汽车复合评论 1936 条，复合评论的采集字段包括用户名称、购买地、初始评论内容、评论时间等。通过问卷星设计问卷并向四线及以下地区新能源汽车车主发放问卷获得 162 份问卷，问卷内容包括性别、年龄、居住地区以及用车感受等。为提高研究的科学性和准确性，数据清洗主要是删除重复和广告性等无价值评论，最终得到有效评论 1611 条。

#### 3.2. 研究方法

本文通过调用经过训练的 Snow NLP 库来分析文本数据的情感值，得到的分析结果是一个在[0~1]区间连续的实数。在默认情况下，程序将评论情感值大于 0.5 的视为积极情感，小于 0.5 的视为消极情感。然而，研究表明在许多情况下，0.5 并非最佳的临界值。本研究发现，将情感值大于 0.394 定义为正向情感，0.391~0.394 为中性，小于 0.391 为负面情绪。实验证明采用这样的情感阈值划分有助于提高模型的应用准确率。对每条评论进行情感分析统计后，研究还发现各种情感倾向在弹幕评论中的比例如表 2 所示。

词频分析是对文本中关键词出现的频率进行的计算分析。本文使用 Python 语言计算纯电动汽车复合评论中的高频词汇，以了解消费者关注的产品焦点特征及其关联程度。利用 word-cloud 库绘制出高频词汇的词云图，见图 2 所示：



Figure 2. Cloud map of composite comments on electric vehicles for rural consumers

图 2. 农村消费者电动汽车复合评论词云图

TF-IDF 是一种常用的关键词提取算法，主要用来评估某个特征词在文档集中的重要程度[14]，TF-IDF

值越高，则说明该特征词对文档具有越高的重要性。本文将得到的数据进行预处理之后通过改进 TF-IDF 算法来提取新能源汽车评论文本特征。改进 TF-IDF 算法相较于传统词频统计具有更能代表文档特征等优点，研究将改进 TF-IDF 算法提取结果和词频统计结果放在一起分析，结果见表 1 所示。

**Table 1.** Word frequency statistics and the top 50 feature words of TF-IDF weight value

**表 1.** 词频统计和 TF-IDF 权重值前 50 的特征词

序号	TF-IDF	词频统计	TF-IDF	词频统计	
1	价格	价格	26	喜欢	加速
2	驾驶	空间	27	配置	油车
3	空间	驾驶	28	特斯拉	销售
4	维修	动力	29	纯电	时间
5	内饰	公里	30	体验	一点
6	油车	提车	31	够用	表现
7	感受	续航	32	电池	体验
8	满意	感受	33	用车	自动
9	比亚迪	满意	34	刹车	电车
10	续航	高速	35	底盘	车机
11	高速	内饰	36	功能	电池
12	座椅	模式	37	加速	确实
13	动力	不错	38	购车	选择
14	后排	外观	39	方向盘	后备箱
15	外观	比亚迪	40	舒服	试驾
16	油耗	座椅	41	自动	舒服
17	车机	喜欢	42	智能化	底盘
18	小鹏	后排	43	电耗	小鹏
19	试驾	油耗	44	混动	发动机
20	空调	空调	45	辅助	辅助
21	不错	配置	46	操控	购车
22	模式	功能	47	销售	刹车
23	后备箱	价格	48	车子	方向盘
24	电车	情况	49	语音	用车
25	充电	充电	50	行驶	行驶

**Table 2.** The proportion of bullet screen of each emotional tendency

**表 2.** 各情感倾向弹幕所占比例

情感倾向	评论数目统计/条	所占比例
积极文本评论	1342	84.58%
消极文本评论	153	9.57%
中立文本评论	94	5.85%
总计	1611	100



从图 2 可以发现,农村消费者购买新能源汽车时最关注的特征是价格、空间、驾驶、动力,这反映了他们对于使用经济性、舒适性和安全性的高度重视。这与《中国农村地区电动汽车出行研究(2.0 版)报告》中指出的农村消费者在购买新能源车时最先考虑使用经济性的结论相一致。此外农村消费者购买新能源汽车时最常提及的品牌是比亚迪和小鹏,这说明这两个品牌在农村市场具有较高的知名度和认可度,也反映了国产新能源汽车品牌在技术创新、产品质量、服务体系等方面取得了显著进步。农村消费者购买新能源汽车时最常用的评价词是感受、满意、不错和喜欢,这表明他们对于新能源汽车产品有着积极正面的态度和情感,也体现了他们对于美好生活需求的追求,表 2 可以看出大多数农村消费者对新能源汽车体验有着积极的态度。同时,他们也会提及一些存在的问题或建议,如充电难、续航不足等,这为进一步完善新能源汽车产品和服务提供了参考。从 TF-IDF 方法看,油耗、电耗等与传统燃油车进行比较的特征并不突出,而从词频方法看,则相反。这可能说明农村消费者在选择新能源汽车时,并不完全依赖于与燃油车的对比结果,而更多地关注自身需求和实际体验。对于农村消费者而言最在意的还是新能源汽车的价格,其次才是驾驶感受,内饰空间和维修。这反映了农村消费者对于新能源汽车的续航有着较为合理的预期,即在农村环境下,新能源汽车的续航水平已经足以应对他们的日常出行场景。因此,在本文构建的数学模型中,忽略了充电网络对于新能源汽车推广的作用。

#### 4. 结语

本文主要结论如下:① 农村消费者购买新能源汽车时最关注的特征是价格、空间、驾驶、动力,这反映了他们对于使用经济性、舒适性和安全性的高度重视。② 农村消费者购买新能源汽车时最常提及的品牌是比亚迪和小鹏,这说明这两个品牌在农村市场具有较高的知名度和认可度,也反映了国产新能源汽车品牌在技术创新、产品质量、服务体系等方面取得了显著进步。③ 农村消费者购买新能源汽车时最常用的评价词是感受、满意、不错和喜欢,这表明他们对于新能源汽车产品有着积极正面的态度和情感,也体现了他们对于美好生活需求的追求。

基于以上结论,本文提出以下建议:

第一,政府应继续加大对农村市场新能源汽车推广方面的政策支持力度,并考虑区域差异性制定合理有效的补贴标准。

第二,新能源汽车企业应加强在农村市场上品牌宣传与推广工作,并针对农村消费者需求开发更多适合其使用场景与偏好特征的产品。

第三,新能源汽车行业应完善在农村地区充电设施与售后服务网络建设,并加强用户教育与培训工作,提高农村消费者对新能源汽车的认知度和使用技能。

第四,新能源汽车企业应该加强技术创新,降低生产成本,开发更多适合农村市场需求的中低端产品,并缩小电动车与燃油车在同一价位段的差距,消除消费者的价格偏好。

#### 参考文献

- [1] Cecere, G., Corrocher, N. and Guerzoni, M. (2018) Price or Performance? A Probabilistic Choice Analysis of the Intention to Buy Electric Vehicles in European Countries. *Energy Policy*, **118**, 19-32. <https://doi.org/10.1016/j.enpol.2018.03.034>
- [2] Li, L., Wang, Z., Chen, L., et al. (2020) Consumer Preferences for Battery Electric Vehicles: A Choice Experimental Survey in China. *Transportation Research Part D*, **78**, 102185. <https://doi.org/10.1016/j.trd.2019.11.014>
- [3] Morton, C., Anable, J. and Nelson, D.J. (2016) Exploring Consumer Preferences towards Electric Vehicles: The Influence of Consumer Innovativeness. *Research in Transportation Business & Management*, **18**, 18-28. <https://doi.org/10.1016/j.rtbm.2016.01.007>
- [4] Jones, A., Begley, J., Berkeley, N., et al. (2020) Electric Vehicles and Rural Business: Findings from the Warwickshire

- Rural Electric Vehicle Trial. *Journal of Rural Studies*, **79**, 395-408. <https://doi.org/10.1016/j.jrurstud.2020.08.007>
- [5] Zhao, Y. and Zhao, P.J. (2021) The Factors in Residents' Mobility in Rural Towns of China: Car Ownership, Road Infrastructure and Public Transport Services. *Journal of Transport Geography*, **91**, 102950. <https://doi.org/10.1016/j.jtrangeo.2021.102950>
- [6] Nazari, F., Rahimi, E. and Mohammadian, A.(K.) (2019) Simultaneous Estimation of Battery Electric Vehicle Adoption with Endogenous Willingness to Pay. *eTransportation*, **1**, 100008. <https://doi.org/10.1016/j.etrans.2019.100008>
- [7] Xiong, Y.Q. and Wang, L.Y. (2020) Policy Cognition of Potential Consumers of New Energy Vehicles and Its Sensitivity to Purchase Willingness. *Journal of Cleaner Production*, **261**, 121032. <https://doi.org/10.1016/j.jclepro.2020.121032>
- [8] 赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834-1848.
- [9] Chen, L.C., Lee, C.M. and Chen, M.Y. (2020) Exploration of Social Media for Sentiment Analysis Using Deep Learning. *Soft Computing*, **24**, 8187-8197. <https://doi.org/10.1007/s00500-019-04402-8>
- [10] Guo, X. and Liu, Y. (2011) The Analysis of China New Energy Vehicle Industry Alliance Status Based on UCINET Software. *Journal of Computers*, **6**, 1852-1856.
- [11] Zeng, M.A. (2018) Foresight by Online Communities—The Case of Renew-Able Energies. *Technological Forecasting and Social Change*, **129**, 27-42. <https://doi.org/10.1016/j.techfore.2018.01.016>
- [12] Wang, H., Shi, H., Chen, P. and Ouyang, M. (2014) Analysis on the Progress of Energy Saving and New Energy Vehicle Industrialization in China Based on a Database. *Journal of Automotive Safety and Energy*, **5**, 294-297.
- [13] Xie, Q. and Tian, Z.L. (2015) Text Analysis of China's New Energy Vehicle Industry Policy. *Advanced Materials Research*, **1073-1076**, 2499-2502. <https://doi.org/10.4028/www.scientific.net/AMR.1073-1076.2499>
- [14] Sung, B. and Park, S.-D. (2018) Who Drives the Transition to a Renewable-Energy Economy? Multi-Actor Perspective on Social Innovation. *Sustainability*, **10**, Article 448. <https://doi.org/10.3390/su10020448>
- [15] Abualigah, L.M., Khader, A.T. and Hanandeh, E.S. (2018) Hybrid Clustering Analysis Using Improved Krill Herd Algorithm. *Applied Intelligence*, **48**, 4047-4071. <https://doi.org/10.1007/s10489-018-1190-6>
- [16] Abualigah, L.M. and Khader, A.T. (2017) Unsupervised Text Feature Selection Technique Based on Hybrid Particle Swarm Optimization Algorithm with Genetic Operators for the Text Clustering. *The Journal of Supercomputing*, **73**, 4773-4795. <https://doi.org/10.1007/s11227-017-2046-2>
- [17] Abualigah, L.M. and Hanandeh, E.S. (2015) Applying Genetic Algorithms to Information Retrieval Using Vector Space Model. *International Journal of Computer Science, Engineering and Applications*, **5**, 19-28. <https://doi.org/10.5121/ijcsea.2015.5102>