

绿动中国，双碳先行

——基于济南市新能源汽车市场机会挖掘

陈心雨, 孔悦, 滕文珍, 李茵茵, 宫慧

曲阜师范大学统计与数据科学学院, 山东 曲阜

收稿日期: 2024年3月14日; 录用日期: 2024年4月4日; 发布日期: 2024年4月15日

摘要

发展新能源汽车是保障能源安全和实现双碳目标的重要举措。团队从新能源汽车的市场现状入手, 从新能源汽车的发展现状、消费者对新能源汽车的购买意愿、消费者的主要需求三个方面展开调研。首先, 团队利用大数据文本挖掘, 绘制出新能源汽车博文与评论的词云图; 其次, 基于聚类分析进行目标客群挖掘, 使用对购买分数贡献较大的特征归纳出五类客户群体。分析表明, 女性、35岁左右、高学历、企业职工、拥有城镇户口、家庭年收入较高且有燃油汽车无新能源汽车的群体更倾向于购买新能源汽车。最后, 构建研究新能源汽车购买意愿的结构方程模型, 选择九个潜变量及相应观测变量, 并提出研究假设。研究结果表明各潜变量对新能源汽车购买意愿的影响均显著, 各潜变量与其观测变量之间也是显著的。情境因素、新能源汽车态度和感知行为控制对新能源汽车购买意愿的影响较大。为进一步探究上述三个潜变量的观测变量对购买意愿的影响, 建立XGBoost模型对购买意愿进行回归预测, 得到消费者购买意愿的变化趋势。结果表明, 政府应注重对新能源汽车的宣传推广, 加大对充电桩的建设并增加对新能源汽车的购买补贴。

关键词

新能源汽车, 文本挖掘, 聚类分析, 结构方程, XGBoost

Green China, Double Carbon First

—Based on the Opportunity Mining of New Energy Vehicle Market in Jinan

Xinyu Chen, Yue Kong, Wenzhen Teng, Yinyin Li, Hui Gong

School of Statistics and Data Science, Qufu Normal University, Qufu Shandong

Received: Mar. 14th, 2024; accepted: Apr. 4th, 2024; published: Apr. 15th, 2024

Abstract

The development of new energy vehicles is an important measure to ensure energy security and achieve the goal of dual carbon. The research team starts from the market status of new energy vehicles, and conducts research from three aspects: the development status of new energy vehicles, consumers' willingness to purchase new energy vehicles, and consumers' main needs. First of all, the team uses big data text mining to draw a word cloud map of new energy automobile blog posts and comments; secondly, based on cluster analysis, the target customer group is mined, and five types of customer groups are summarized by using the characteristics that contribute more to the purchased score. The analysis shows that women, about 35 years old, highly educated, corporate employees, urban household registration, high annual household income and fuel vehicles without new energy vehicles are more inclined to buy new energy vehicles. Finally, a structural equation model is constructed to study the purchase intention of new energy vehicles, nine latent variables and corresponding observation variables are selected, and research hypotheses are proposed. The results show that the influence of each latent variable on the purchase intention of new energy vehicles is significant, and the latent variables and their observed variables are also significant. Situational factors, new energy vehicle attitudes and perceived behavioral control have a greater impact on the purchase intention of new energy vehicles. In order to further explore the influence of the observed variables of the above three latent variables on the purchase intention, an XGBoost model is established to predict the purchase intention, and the changing trend of consumers' purchase intention is obtained. The results show that the government should pay attention to the promotion of new energy vehicles, increase the construction of charging piles and increase the purchase subsidy of new energy vehicles.

Keywords

New Energy Vehicle, Text Mining, Cluster Analysis, Structural Equation, XGBoost

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

1.1. 研究背景

世界资源研究所(WRI)指出,交通业的碳排放量占我国碳排放总量的70%以上,仅排在发电供热、制造业用能之后,因此要减少碳排放,新能源汽车的发展迫在眉睫。作为一个负责任的大国,中国提出“力争到2030年前实现碳达峰,到2060年前实现碳中和”。新发展格局下,为达成“双碳目标”,势必要刺激新能源汽车行业的发展,对中国的汽车产业进行重构。当前,新能源汽车产业处在一个“风口”,这将为我国“十四五”期间实现高质量发展奠定良好的基础[1]。

1.2. 文献综述

中华人民共和国国务院办公厅日前颁布新的国家标准:《新能源汽车产业发展规划2021~2035年》,《规划》中指出,在目前至2035年的期间,清洁型能源汽车新车销量将超过传统汽车新车销售总额的百分之二十;力争再经十五年的技术革新与奋斗发展,中国的清洁与能源汽车及相关的主要汽车科技达到

全球的领先科技水准，纯氢电动汽车成为清洁与能源汽车市场的一个重点的新的销售款型。

目前，从新能源汽车国内消费者认知情况看，城镇居民对新能源汽车的认可情况高于乡村居民，对新能源汽车持积极态度的消费者较多，客观来看，消费者对新能源汽车具有可观的购车意向[2]。

但在推广过程中，新能源汽车被群众接受的程度仍然不高，相关电池以及汽车的制造技术水平不高，出现了很多问题。比如新能源汽车的制造成本问题，新能源汽车的配套装备建设问题，充电桩、换电站基础设施建设慢，大约40%的用户没有专属充电桩，存在电池、刹车等安全问题等[3]，这也是发展改进的目标。

1.3. 研究内容

具体思路见图1。

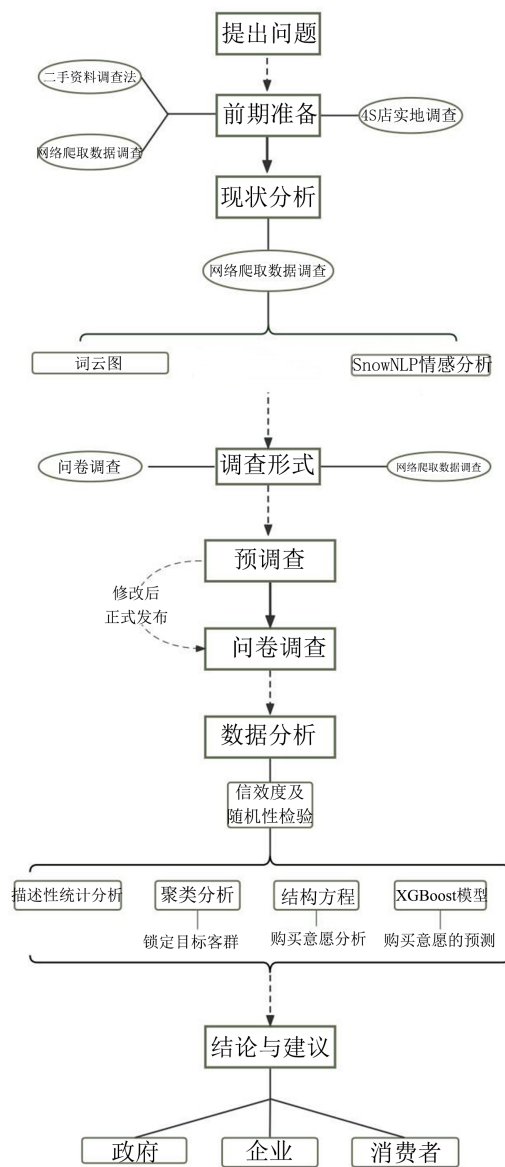


Figure 1. Research idea diagram
图1. 研究思路图

2. 基于文本挖掘的新能源汽车舆情分析

2.1. 数据获取与预处理

1) 数据获取

利用微博的高级搜索功能，以新能源汽车为关键词，获取了从 2022 年 9 月 15 日至 2022 年 12 月 15 日近一万条话题观点和四千条评论观点，用 requests.get 方法获取关于新能源汽车评论观点的 URL，并清除文本中的标签等信息，根据微博 ID 去重。

2) 数据预处理

使用 Jieba 工具进行预处理，得到图 2。

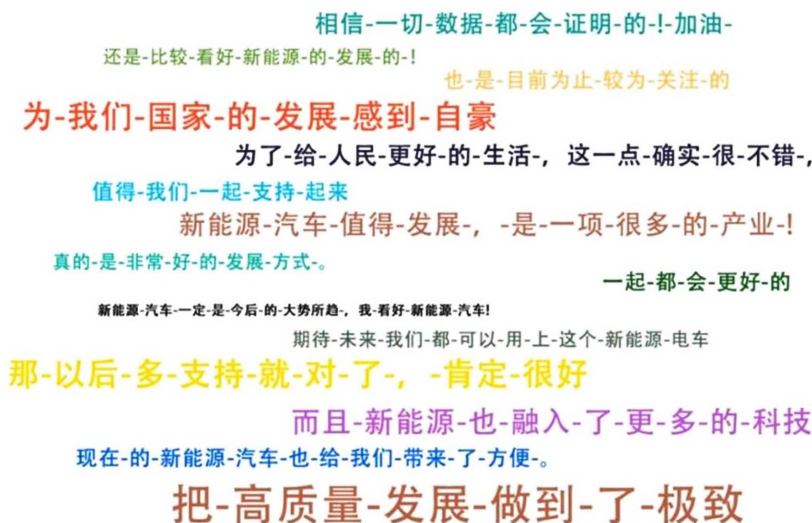


Figure 2. Word segmentation data graph
图 2. 分词数据图

2.2. 词云图

WordCloud 词云图可以帮助快速了解关于新能源汽车的主要内容，显示主要的关键词。用前 500 个高频词绘制词云，为后续问卷的制作提供支撑。

如图 3，话题词云图表明人们较为关注的汽车品牌有：比亚迪，特斯拉，小鹏，蔚来，大众；话题关注问题为：增长，速度，销量，市场，发展；购车关注问题为：电池，公里，续航，充电桩，芯片。



Figure 3. New energy vehicles Weibo inscription cloud map
图 3. 新能源汽车微博话题词云图

如图 4, 评论词云图的高频词汇有发展、充电、支持、建设、期待、比亚迪、性价比、保护环境、发展趋势等。

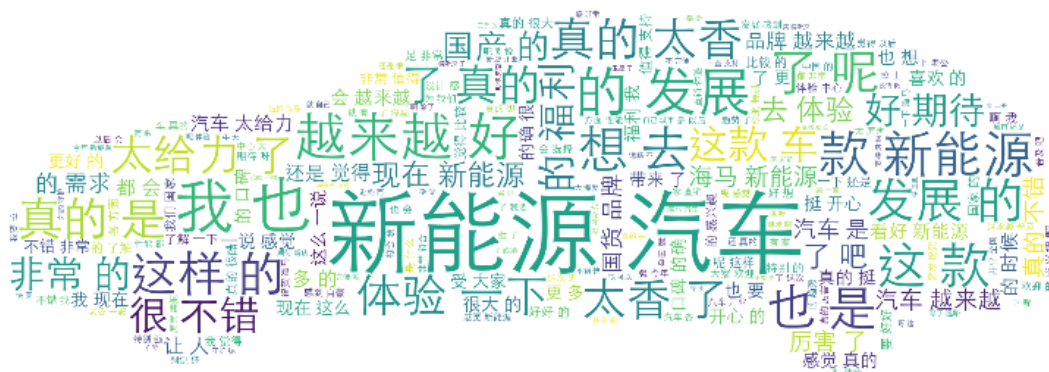


Figure 4. New energy vehicles Weibo comment word cloud map
图 4. 新能源汽车微博评论词云图

2.3. 微博评论情感分析

利用 SnowNLP 进行情感分析。由图 5 知, 新能源市场风评较好。

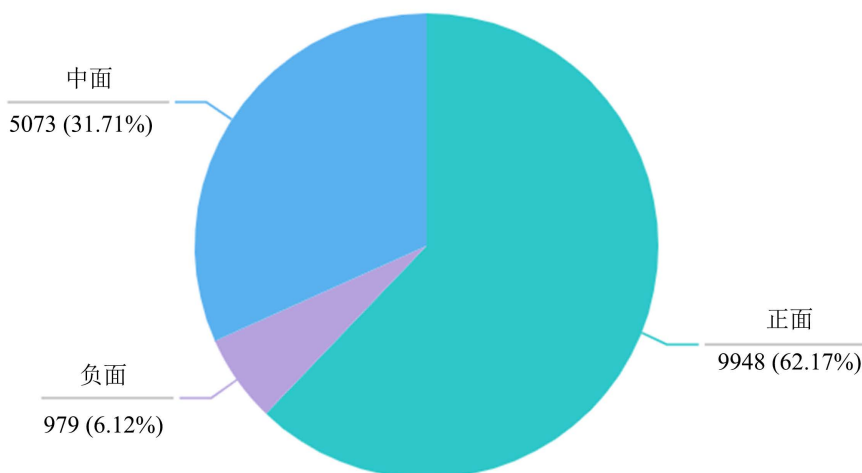


Figure 5. Weibo comment attitude
图 5. 微博评论态度

3. 数据处理

3.1. 数据来源

首先, 基于 Python 的网络爬虫技术, 以新能源汽车为关键词, 在微博获取一万条话题观点和评论观点, 了解人们对于新能源汽车关注的热点问题。其次, 通过 4S 店实地调查和二手资料调查, 确定新能源汽车的发展现状。最后制定调查问卷, 采用多阶段不等概率法在全国范围内抽取 1050 人进行问卷调查。

共回收 1023 份问卷, 使用 SPSS26 对样本数据进行筛选剔除, 如表 1, 最终得到有效问卷 913 份。信效度检验结果显示, 量表的 Cronbach α 系数均大于 0.7, 整体的 KMO 值高达 0.977。因此下文主要利用这 913 份有效问卷进行建模分析。

Table 1. Sample rejection condition**表 1.** 样本剔除情况

	样本量	占比
时长不在 120 s~1200 s 范围内	50	4.9%
量表题全部选择非常不同意	6	0.6%
量表题全部选择非常同意	12	1.2%
量表题全部选择一般	42	4.1%
有效样本数量	913	89.2%
总计	1023	100%

3.2. 数据描述

1) 受访者基本信息

有效样本中男性 437 人，女性 476 人。女性占比大，符合性别比例现状。受访者主要集中在 26~45 岁，约占受访者的 91%。拥有城镇户口的有 590 人，拥有农村户口的有 323 人。学历集中在大学和高中。受访者中企业职员和行政/事业单位居多。受访者的家庭年收入大多在 5~20 万之间。在受访者中只有燃油汽车的人群占比最多，既有新能源汽车也有燃油汽车的受访者最少，有 16.1% 的受访者没有任何汽车。

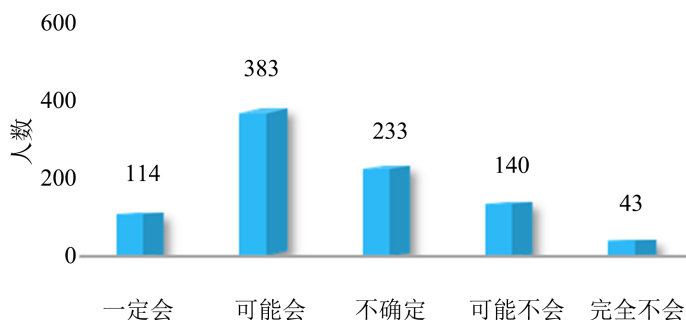
2) 受访者的购车用途及了解途径

大多数人倾向于将新能源汽车应用于日常生活中，而用于载客营运和载货的受访者最少。26~45 岁的青年人倾向于通过网络和新闻报道了解新能源汽车；46~55 岁的中年人倾向通过网络、新闻报道和朋友告知；而 56 岁以上的老年人则倾向于线下宣传。

3) 受访者的购买意愿及需求

① 样本是否会购买新能源汽车分析

由图 6 可以看出，大部分受访者愿意购买新能源汽车。

**Figure 6.** Whether they will buy new energy vehicles distribution map**图 6.** 是否会购买新能源汽车分布图

② 样本购车预算分析

超过八成的受访者可以接受的新能源汽车价格为 9 万~19 万元。

③ 新能源汽车属性重要性分析

由图 7，受访者普遍认为续航能力、充电桩等基础设施建设、安全性能和加速能力更重要，这说明要吸引消费者，最重要的是做好自身性能建设。

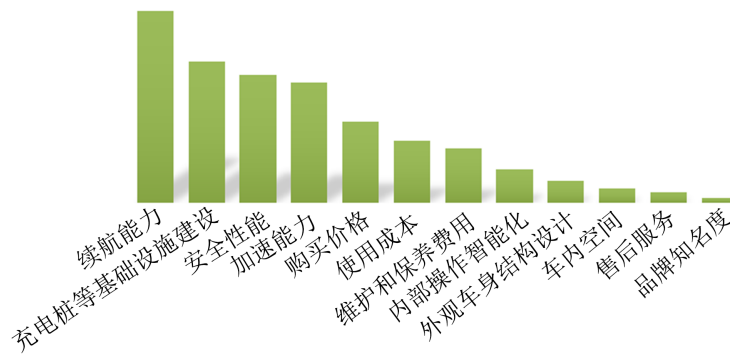


Figure 7. Attribute importance distribution map
图 7. 属性重要程度分布图

④新能源汽车属性改善分析

见图 8，受访者认为目前新能源汽车行业最需要改善的是电桩、充电站等，其次是续航里程、充电时长和市场规范。

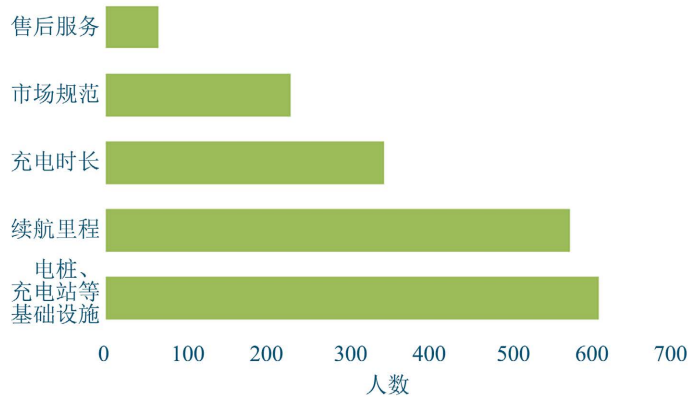


Figure 8. Attribute improvement distribution map
图 8. 属性改善分布图

综上所述，受访者画像图见图 9。



Figure 9. Interviewee portrait
图 9. 受访者画像图

4. 基于聚类分析的目标客群挖掘

根据是否会购买新能源汽车将目标客群分为优质、潜力和一般。客户选择的优先度为：优质客户 > 潜力客户 > 一般客户。

为了找到对购买分数(S)影响最大的特征，采用随机森林算法得出其他特征对购买分数的贡献度，对贡献度较低的因素与购买分数进行相关分析。使用对购买分数贡献较大的特征进行聚类分析，最终得出五类人群。

4.1. 数据预处理

对 913 条数据进行数字型映射编码处理。首先对可能影响购买分数的属性进行相关性分析，相关系数均小于 0.3，无法根据相关性进行降维分析。

随后利用随机森林算法，将购买分数作为类标号进行特征选择。选择重要性大的四个特征(职业、年龄、汽车拥有情况、家庭年收入)为聚类分析的特征，解释度达到 75%。

4.2. K-Means 聚类确定目标客群

上文筛选出了职业、年龄、家庭汽车拥有情况、家庭年收入四个特征，下面采用 K-means 算法进行聚类分析。首先利用手肘法和轮廓系数法求得 K 的最优值为 5。此时轮廓系数为 0.515，效果较好，故选择聚类个数为 5。利用 TSNE 进行数据降维，最终得出五类客户，客户比例分别 0.35、0.22、0.19、0.1、0.14。

4.3. 结论分析

聚类结果显示，客户群体的购买分数分别为：群体 1 = 4.1、群体 2 = 3.1、群体 3 = 1.9、群体 4 = 0.8、群体 5 = 3.5。将购买分数(S)进行均值计算得到全体人群的购买分数 $S = 3.027$ 。群体划分依据如图 10。



Figure 10. Population map
图 10. 群体划分图

结合聚类结果可得结论如图 11。



Figure 11. Five types of customer details
图 11. 五类客户详情图

5. 基于结构方程的购买意愿分析

结构方程模型(SEM)综合了因子分析与路径分析，可用于建立、估计和检验变量之间得因果关系[4]。

5.1. 相关理论分析与回顾

1) 理性行为理论

理性行为理论指出，各种因素通过态度和主观规范作用于行为[5]。具体模型如图 12。

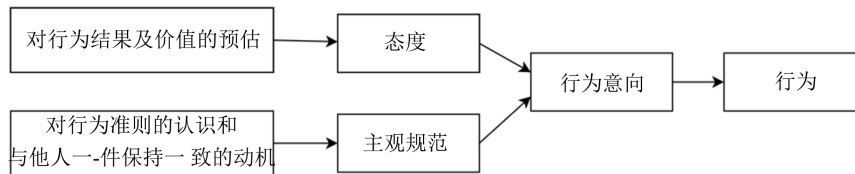


Figure 12. Schematic diagram of a theoretical model of rational behavior
图 12. 理性行为理论模型示意图

2) 计划行为理论

计划行为理论模型示意图见图 13。

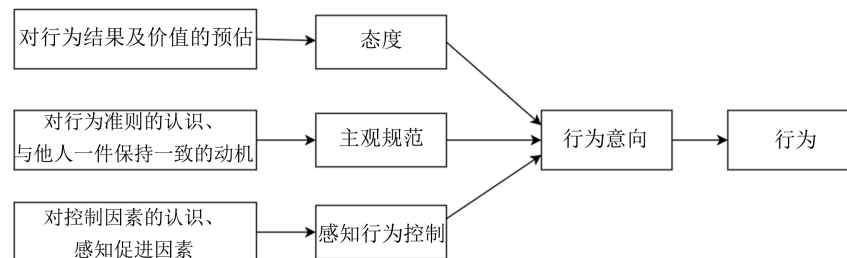


Figure 13. Schematic diagram of a theoretical model of planned behavior
图 13. 计划行为理论模型示意图

3) 顾客感知价值理论

顾客通过权衡获取产品或服务时的付出与回报来判定其具有的价值[6]。

4) 态度 - 情景 - 行为理论

态度 - 情景 - 行为理论认为，当政策不完善或新能源汽车性能不佳时，即使消费者具有非常高的购车意愿，也很难做出购买决策[7]。

5.2. 结构方程模型构建

1) 潜变量的选择

选取绿色观念认知、新能源汽车认知、新能源汽车态度、主观规范、感知行为效力、感知行为控制、情境因素、顾客感知价值、新能源汽车购买意愿这九个潜变量[8]。

2) 研究模型

结合四个理论，构建最终模型如图 14 所示。

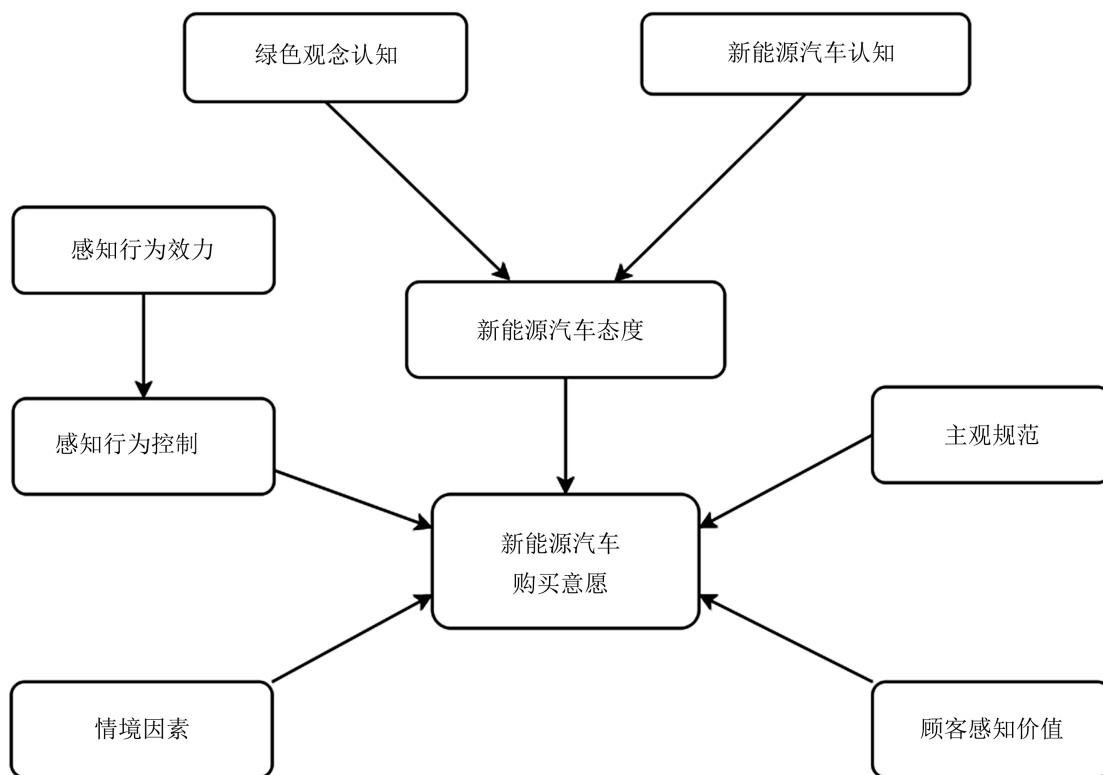


Figure 14. Structural equation model

图 14. 结构方程模型

根据文献[2] [9]和主观制定，为每个潜变量选定几个问题作为观测变量。

5.3. 提出假设

根据以上四个理论，结合已有文献和新发展格局下新能源汽车现状提出假设。

5.4. 问卷数据的结构方程模型

使用 AMOS26 软件对数据建模并计算，得到模型适配度的结果如表 2。

Table 2. Model fitting index table
表 2. 模型拟合指标表

拟合指标	指标值	拟合情况
CMIN (卡方)	716.14	
DF (自由度)	244	
CMIN/DF	2.935	<3 可接受
GFI (拟合优度指数)	0.829	>0.8 可接受
AGFI (修正的拟合优度指)	0.893	>0.8 可接受
RMSEA (近似误差均方根)	0.035	<0.09 可接受

由表 2 可知 X²/df 为 2.935，小于 3，模型的拟合优度指数和修正的拟合优度指数均超过了 0.8，模型近似误差平方根也明显的小于 0.09，这些指标均显示模型适配度较好。最终生成的结构方程路径图如图 15。

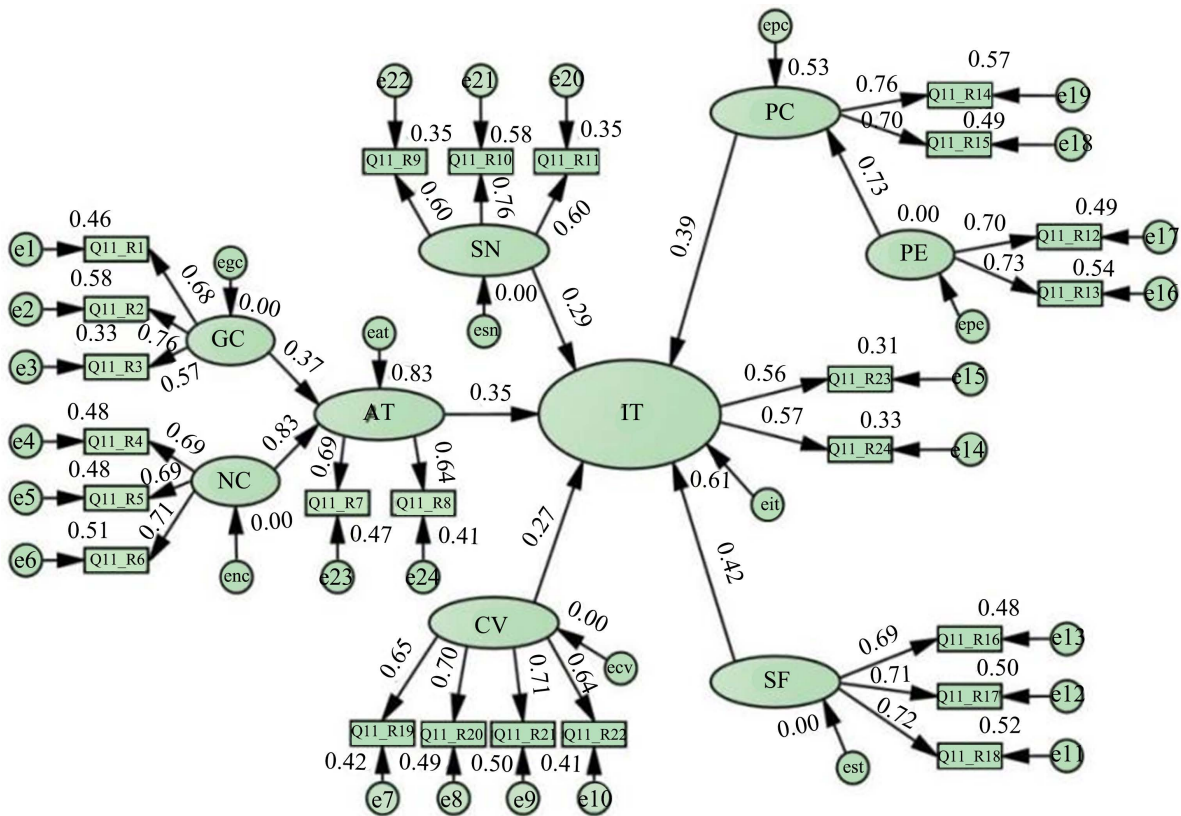


Figure 15. Structural equation path coefficient diagram
图 15. 结构方程路径系数图

该结构方程变量间路径系数均大于 0，表明变量间的影响作用均为正向影响关系，结合研究假设，可得部分结果如表 3。

最终得到模型和简化的假设显著关系如图 16。

Table 3. Model result
表 3. 模型结果

变量 1	变量 2	路径系数	P 值	影响方向	假设 H _i
绿色观念认知(GC)	新能源汽车态度(AT)	0.374	<0.01	正向	H1 通过验证
新能源汽车的认知(NC)	新能源汽车态度(AT)	0.83	<0.01	正向	H2 通过验证
新能源汽车的态度(AT)	新能源汽车购买意愿(IT)	0.348	<0.01	正向	H3 通过验证

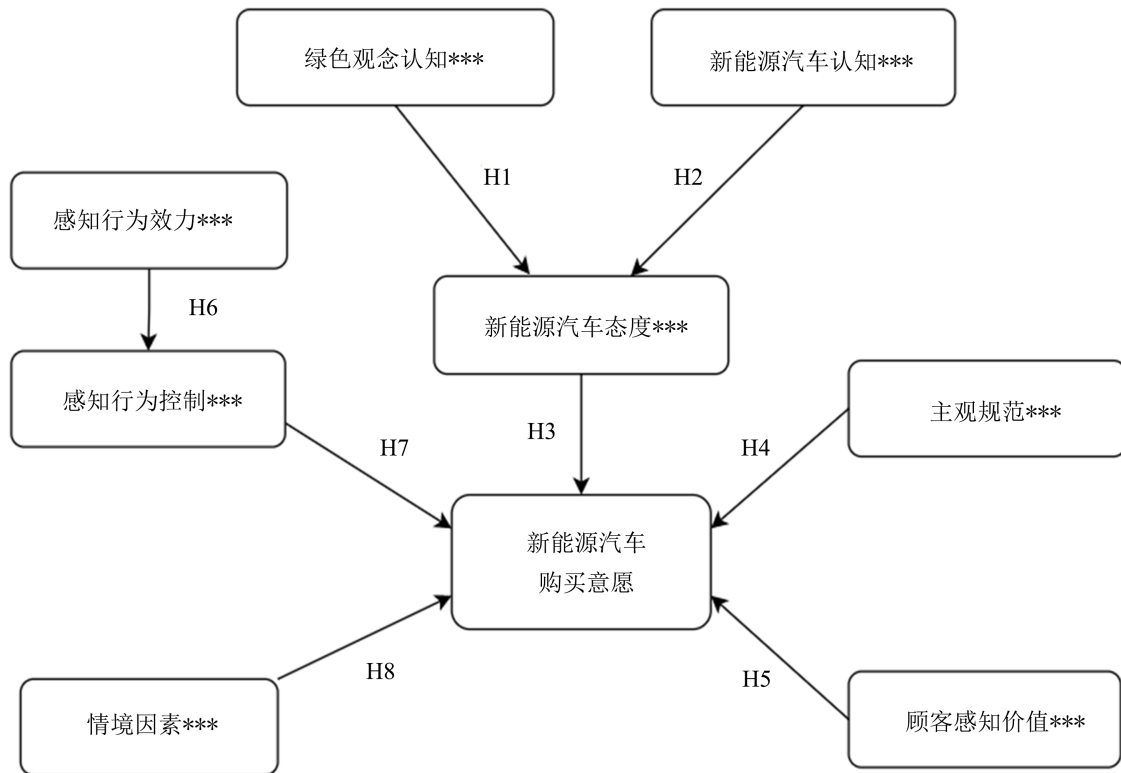


Figure 16. Simplified assumed significant relationship
图 16. 简化的假设显著关系

6. 基于 XGBoost 模型的购买意愿预测

结构方程的研究结果表明，情境因素对新能源汽车购买意愿的路径系数最大，其次是新能源汽车态度和感知行为控制。为了进一步探究在这三个因素对应观测变量对购买意愿的影响，建立 XGBoost 模型对购买意愿进行回归预测。

6.1. 变量说明与数据描述性统计

将问卷选项分别赋值为 1 至 5，对购买意愿进行加权分析，变量均值均在 3.7~3.8 之间，相对较高。

6.2. 模型简介与模型建立

GBDT (梯度提升树)是一个以回归树为基学习器，以 Boost 为框架的加法模型的集成学习，XGBoost 是在 GBDT 的基础上对 Boosting 算法进行的改进。经过参数调优，选定最终模型的基学习器为 gbtree、学习率为 0.12。

6.3. 模型结果与评估

通过建立 XGBoost 模型，得到贡献度最大的四个因素依次为 SF_R16、SF_R18、SF_R17 和 AT_R8。具体特征贡献度如图 17。

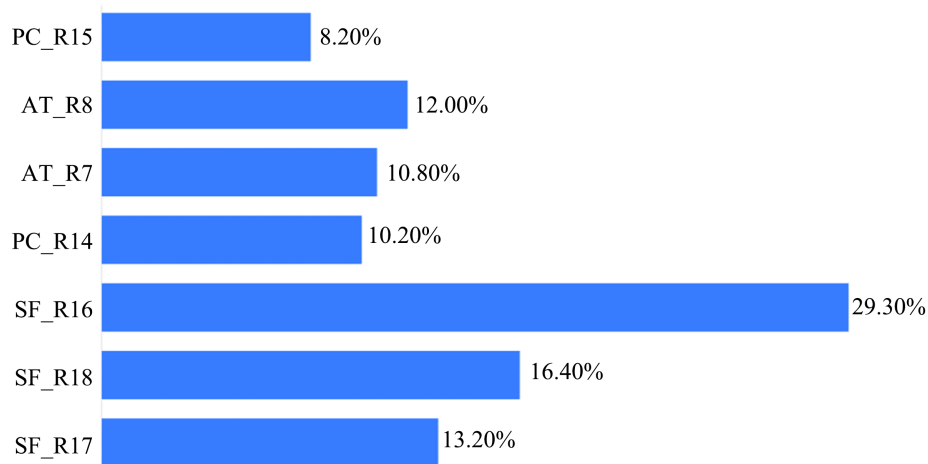


Figure 17. Feature contribution bar chart
图 17. 特征贡献度柱状图

选取 MSE、 R^2 等指标对模型进行评估如表 4，模型对训练集与测试集拟合效果均良好。给出部分预测结果，如图 18。

Table 4. Model evaluation parameter values
表 4. 模型评估参数值

	RMSE	MAE	MAPE	R^2
训练集	0.26	0.142	3.731	0.892
测试集	0.293	0.165	4.268	0.869

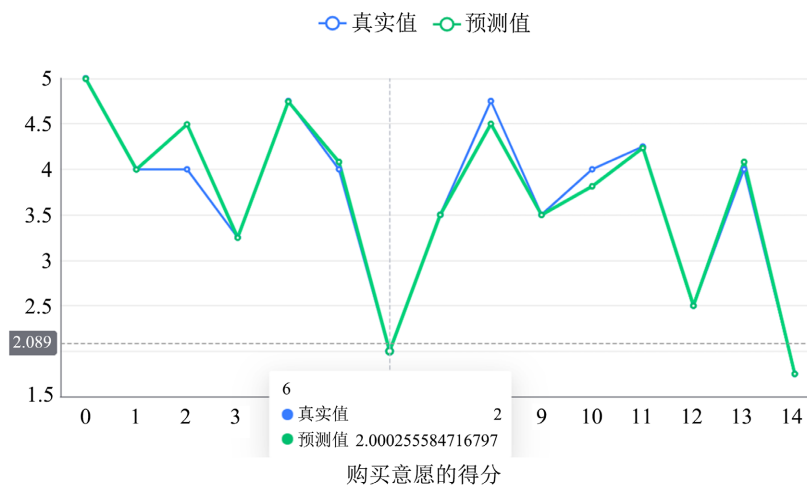


Figure 18. Part of the prediction results are shown
图 18. 部分预测结果展示

6.4. 购买意愿的预测

基于以上模型,采用控制变量的方法对特征贡献度最大的四个特征 SF_R16、SF_R18、SF_R17、AT_R8 分别进行研究。首先固定除研究目标特征外的其余六个特征,并对此六个特征取均值(均值代表大多数受访者在这六个问题上的选择),来研究目标特征变化对购买意愿的影响,找出特征阈值。

当消费者对政府注重宣传推广的态度从一般到同意,购买意愿会显著性增加,从 3.3 增加到 4.1;当消费者对政府重视充电桩建议的态度从一般到同意,购买意愿会显著性增加,从 2.8 增加到 3.9。

在研究单一变量之后,对特征贡献度最大的两个特征进行研究,结果表明,当两者均到达 4.5 时,最高购买意愿为 4.64,及购买意愿相当强烈。详细预测结果如图 19。

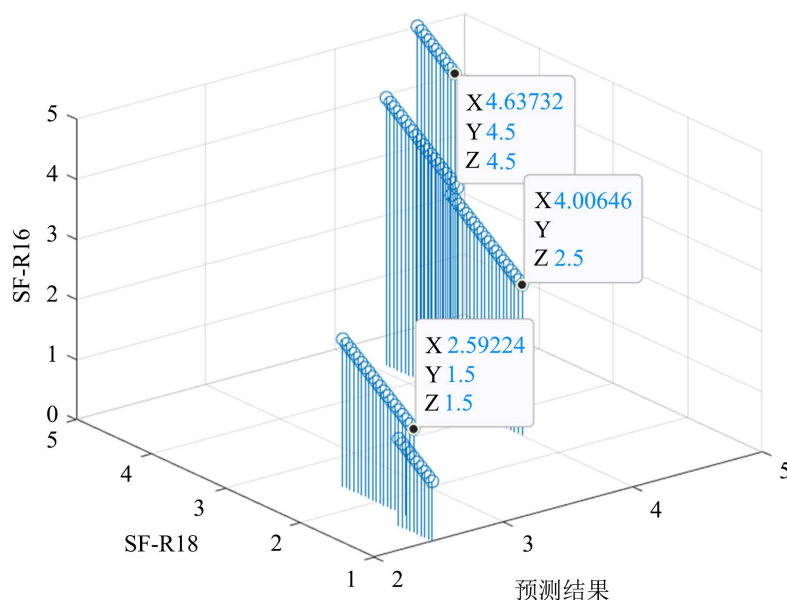


Figure 19. Prediction of purchase intention

图 19. 购买意愿的预测

综上,为了促进消费者购买新能源汽车,政府首先应加大对新能源汽车的宣传推广;其次要重视充电桩的建设并加大对新能源汽车的补贴;最后要增强消费者支持新能源汽车的意识。

7. 结论与建议

7.1. 结论

1) 新能源汽车发展现状及其潜力

依据新能源汽车博文与评论词云图,人们较为关心新能源汽车的发展状况与充电状况,较感兴趣的新能源汽车品牌主要有比亚迪、特斯拉、小鹏、蔚来、大众等,对新能源汽车普遍持有积极态度;新能源汽车的优势在于自身的环保特性与性价比;关注新能源汽车的人主要为青年和中年,且有较大部分分布在广东、北京、江苏、山东、浙江;新能源汽车的话题关注度保持着较高的热度,发展前景良好。

2) 锁定新能源汽车目标客群

① 优质客户

年龄为 35 岁左右的青年群体,职业为企业职员,家庭年收入较高,年收入接近 20 万,家庭汽车拥有情况为有燃油汽车但没有新能源汽车。优先考虑高学历、女性以及具有城镇户口的客户。

②重要潜力客户

年龄为 45 岁左右的中年群体，职业为企业职员，家庭年收入为 17 万左右，家庭汽车拥有情况为没有任何汽车。优先考虑高学历以及拥有城镇户口的客户。

③一般潜力客户

年龄为 33 岁左右，职业为个体经营者，家庭年收入为 16 万左右，家庭汽车拥有情况为有燃油汽车但没有新能源汽车。优先考虑拥有城镇户口的客户。

④一般客户

年龄为 30 岁左右，在事业单位工作，家庭年收入为 15 万元左右，家庭汽车拥有情况为有新能源汽车但没有燃油汽车。

⑤重要潜在客户

年龄为 30 岁以下，年龄较低，大多为学生群体，家庭年收入较低，为 9 万左右，家庭拥有燃油汽车或新能源汽车。

3) 目标客群的购买意愿

通过建立结构方程，我们发现新能源汽车态度、感知行为控制、情景因素和顾客感知价值，对消费者的购买意愿均有显著的影响。其中对购买意愿影响最大的是情景因素，标准化路径系数为 0.419；其次是新能源汽车态度和感知行为控制，标准化路径系数分别为 0.387 和 0.348。说明在我国全力实现双碳减排目标的大背景下，应加大对新能源汽车的宣传推广，充分体现和宣传新能源汽车的绿色环保属性，同时重视充电桩的建设并加大对新能源汽车的补贴，增强消费者支持新能源汽车的意识。

7.2. 建议

1) 企业层面

树立厂销合一新理念，兼顾营销与生产。积极通过线上渠道营销以满足优质客户和潜力客户的消费需求。扩大覆盖圈层，营造热衷环保的氛围。借助高曝光媒介为品牌造势助力，帮助品牌培养“死忠粉”。另外，打铁还需自身硬，企业需要围绕技术堵点进行研发攻关，加大对核心技术与产品的研发力度，实现新能源汽车的网联、智联。

2) 政府层面

在新发展格局下，为实现双碳目标，政府首先应加大对新能源汽车的宣传推广，在社会中打造低碳环保主流；其次要重视充电桩等基础设施的建设和普及，对企业的技术研发给予鼓励和支持；最后要加大对新能源汽车的补贴，让消费者感受到国家对新能源汽车的高度重视，助力消费者打消购车疑虑。

3) 消费者层面

加大对环保低碳观念的认可度，紧跟时事政策，以开放的心态积极尝试新鲜事物，以包容的心态去面对被曝光的消极问题，相信在碳中和时代下，新能源汽车将会慢慢替代燃油汽车。

基金项目

国家级大创项目资助(项目编号：202210446016)。

参考文献

- [1] 孙越. 我国新能源汽车分类及发展现状浅析[J]. 汽车实用技术, 2020, 45(4): 13-15.
- [2] 徐国虎, 许芳. 新能源汽车购买决策的影响因素研究[J]. 中国人口·资源与环境, 2010, 20(11): 91-95.
- [3] 丁少帅. 低碳经济背景下新能源汽车发展对策研究[J]. 内燃机与配件, 2021(6): 181-182.
- [4] 王卫东. 结构方程模型原理与应用[M]. 北京: 中国人民大学出版社, 2009.

- [5] 于丹, 董大海, 刘瑞明, 原永丹. 理性行为理论及其拓展研究的现状与展望[J]. 心理科学进展, 2008, 26(5): 796-802.
- [6] 叶志桂. 西方顾客价值研究理论综述[J]. 北京工商大学学报(社会科学版), 2004, 24(4): 11-15+87.
- [7] Guagnano, G.A., Stern, P.C. and Dietz, T. (1995) Influences on Attitude-Behavior Relationships. *Environment and Behavior*, **27**, 699-718. <https://doi.org/10.1177/0013916595275005>
- [8] 刘腾飞. 北京消费者新能源汽车购买意愿影响因素研究[D]: [硕士学位论文]. 北京: 北京林业大学, 2016.
- [9] 潘鹏, 张娟利, 金高云, 孔丽丽, 陈松, 邹露艳. 贵州省新能源汽车销售现状调查及分析[J]. 时代汽车, 2021, 18(4): 82-84.