

汽车消费者购买意愿的决策树分类研究

张利青 张毅

广东东软学院 信息管理与工程学院

DOI:10.12238/ej.v7i2.1354

[摘要] 识别汽车消费者购买意愿能辅助有效预测汽车消费者的购买行为,构建基于决策树的汽车消费者购买意愿分类模型,有助于提高汽车消费者购买意愿的准确性。采用问卷调查获取638份数据,运用SPSS Modeler 18软件和C5.0决策树方法,利用特征选择识别关键预测变量,并建立汽车能源类型选择和汽车品牌类型选择的决策模型,研究结果显示:能源类型的分类准确性为86.85%,品牌类型的分类准确性为88.28%。

[关键词] 决策树; 汽车消费者; 购买意愿; 分类

中图分类号: F407.471 文献标识码: A

Research on Decision Tree Classification of Automobile Consumers' Purchase Intention

Liqing Zhang Yi Zhang

School of Information Management and Engineering, Guangdong Neusoft University

[Abstract] Identifying automobile consumers' purchasing intentions can assist in effectively predicting their purchasing behavior, and constructing a decision tree based classification model for automobile consumers' purchasing intentions can help improve the accuracy of automobile consumers' purchasing intentions. A questionnaire survey was conducted to obtain 638 pieces of data. SPSS Modeler 18 software and C5.0 decision tree method were used to identify key predictive variables using feature selection, and a decision model for selecting automotive energy types and automotive brand types was established. The research results showed that the classification accuracy of energy types was 86.85%, and the classification accuracy of brand types was 88.28%.

[Key words] Decision tree; Automotive consumers; Purchase intention; Classification

引言

根据中国汽车工业协会数据显示,2022年汽车产销分别完成2702.1万辆和2686.4万辆,其中乘用车的产销分别完成2383.6万辆和2356.3万辆。让人们的出行变得十分便捷,也让汽车变为人们出行的重要工具之一。这也使得市面的汽车行业的竞争变得愈发激烈。改革开放已是这个时代的主题。汽车购买对于大部分人来说是一件需要慎重选择的事,而汽车能让人看中的方面有很多,消费者在购买汽车时,会考虑到许多因素,比如品牌、价格、配置等,以便选择出最符合自身需求的汽车。

研究发现,影响汽车的选择的因素有政治、文化、个人、价格等多种因素^[1],消费者不再只满足汽车的基本使用功能,对产品的价值和情感诉求越来越激烈。杨一翁等认为品牌首先影响消费者态度,再进一步影响消费者购买倾向^[2]。郑先锋认为汽车外观设计能够抓住消费者的眼球,并且引发人的积极的情感体验,消费者才会花更多的时间去了解汽车的其它性能,并影响最终的购车决策^[3]。

在购买决策中,用户一般会有对某种特征的偏好,特征的筛

选对购买意愿分析有很大意义,有利于汽车经销商形成针对用户特点制定有效营销策略的管理思路^[4]。决策树算法是最为广泛的数据分类技术应用^[5]。

1 文献综述

1.1 汽车消费者购买意愿研究

收入水平、汽车品牌、汽车的价格和油耗等因素影响汽车消费者的购买意愿。肖欣莹在基于综合熵权法和VAR的模型研究中表明各个影响因素之间有相互联系、牵制,其中GDP和钢材市场行情对汽车销售量的影响更明显^[6]。汽车的定价对于人们的购买意愿也是有着很大的影响,刘道东实证研究发现汽车的销量会随着汽车的价格发生变化,且汽车价格是制约着汽车消费的主要因素^[7]。汽车的油耗往往也是人们关注的焦点,李全森证明从消费者情感价值上看油耗对消费者购买意愿影响最大^[8]。

1.2 消费者购买意愿的分类研究

在汽车消费者购买意愿分析领域,决策树可以有效地分析消费者的购买行为,从而了解消费者的购买意愿。陈家齐通过Cart决策树算法为例,通过判断商品用户的属性信息,以达到对

购买行为的预测,从而服务商家,同时也为用户带来便利^[9]。根据其特点选择合适的决策树分类算法,往往能起到事半功倍的效果,从而提高决策的准确性。Mou等在对50个数据的数据集的分析中应用了决策树分类技术,其中,支持向量机(SVM)的预测准确率最优,能达到86.7%的准确率。此外,还有学者通过新冠疫情前后消费者汽车购买情况的不对称研究,梯队提升决策树(GBDT)对疫情前后消费者购买意愿情况有较强的解释作用。

1.3 文献总结

在影响因素的研究中,目前对汽车消费者购买意愿的研究,从多个不同的角度来考虑,如汽车消费者的属性、汽车消费者的购买行为等。在属性方面,如汽车消费者的年龄、性别、收入水平等属性对消费者购买意愿的影响进行分析,从而更好地了解汽车消费者的购买行为。在消费心理和购买行为方面,经济发展水平、消费者情感感知、家庭环境等因素也被认为会对汽车购买产生影响。因此,本研究对汽车消费者购买影响因素的分析研究中,将可能存在的影响分为消费者基本信息、消费者购车的情感需求影响因素两类,将消费者购车的目标变量设置成能源类型与品牌类型。

2 研究设计

2.1 问卷设计

调查汽车消费者购买意愿,属于典型的意愿调查(Stated Preference),在意向调查中需要考虑的影响因素较多,容易出现不同因素交叉、调查表繁杂的情况。为避免出现对消费者调查不全、误导决策的情况,问卷的题项设计参考了陈美云等^[13]的研究成果,问卷的内容具体如表2.1所示。

表2.1 汽车消费者购买意愿调查问卷

类别	名称	选项
消费者基本信息	性别	男;女
	年龄	30岁以下;30-45岁;45-60岁;60岁以上
	家庭收入情况	10万以下;10-20万;20-40万;40-60万;60万以上
购车影响因素	安全性	5 非常在意;4 比较在意;3 一般;2 不是很在意;1 完全不在意
	内饰/装配	
	排气量/性能	
	外观/设计感	
	维修/售后服务	
	油耗	
	品牌偏好/口碑	
价格预算		
购车行为偏好	能源类型	新能源;传统油耗;油电混动
	品牌类型	国产品牌:长城、五菱、比亚迪、吉利、荣威、长安、红旗、奇瑞等 国外品牌:特斯拉、丰田、马自达、捷豹、奔驰、雷克萨斯、本田、沃尔沃、宝马、三菱

2.2 问卷数据回收

预调查完善后,问卷主要通过汽车之家论坛、豆瓣汽车交流社区、购车交流群等渠道进行发放与收集。共回收765份问卷,筛选后共有638份有效问卷,回收率83.39%。消费者购车影

响因素量表的信度CA值是0.913,影响因素量表的KMO值为0.894(KMO>0.6),问卷有效。

2.3 描述性统计分析

2.3.1 消费者基本信息

从调查数据中的性别来看,男性占比为51.73%,女性占比为48.27%。其中30-45岁的人群最多,占比为52.36%;30岁以下和45-60岁的人群分别占比17.08%、24.76%,60岁以上的高龄人群占5.80%。家庭收入情况中大部分集中10万以下、10-20万、20-40万、40-60万,分别占比16.01%、30.09%、31.66%、13.32%,60万以上的高收入人群占比8.91%。

2.3.2 购车行为偏好

在汽车类型选择方面,34.64%的人群选择新能源汽车,48.43%的人群选择传统油耗车,16.93%的人群选择油电混合车。油耗车技术发展成熟,更受消费者青睐。

选择国内汽车品牌时,五菱、长城、比亚迪和吉利这四个品牌的汽车选取比例超过30%,荣威、长安和红旗的选取比例分别为27.12%、19.44%、10.66%,奇瑞的选取比例为10.66%。

选择国外汽车品牌时,特斯拉的选取比例最高,占28.21%;其次丰田、马自达和捷豹均在20%以上;三菱的选取比例最低,只有9.40%。

3 影响因素的决策树分析

3.1 数据预处理与划分

3.1.1 数据预处理

在构建决策树模型之前,需要对原始数据进行预处理,主要包括数据清洗、缺失值处理、问卷数据处理。数据清洗是去除数据中的异常值和重复值,以保证数据的有效性和准确性;缺失值处理是填补数据中的缺失值,以保证数据的完整性;问卷数据处理需要对变量的数据类型进行细分,明确定量与定类变量,以便于后续的决策树分析。

首先,导入问卷收集的数据,并对连续数据、分类数据的数据类型进行详细设置,同时将“能源类型”和“品牌类型”设置为目标变量;接着进行特征选择,选择能源类型决策和品牌类型决策过程的关键预测变量;基于特征选择结果,进行决策树模型构建,构建前先进行分区,按照训练集70%、测试集30%的标准进行C5.0决策树分析;最后,对决策树模型的构建结果的精度进行分析。

3.1.2 数据集划分

在构建决策树模型之前,需要将收集的数据集进行划分,分为训练集和测试集。训练集用于建立决策树模型,测试集用于检验模型的准确性。本文采用70%的数据作为训练集,30%的数据作为测试集,以更好地反映模型的性能。在数据划分的过程中,本文利用SPSS Modeler 18中的“分区”完成,采用随机划分的方法。由于汽车消费者购买意愿数据集中的每一条数据是独立的,因此,采用随机划分的方法可以使训练集和测试集的数据分布更加均匀,更能体现模型的准确性。

3.2 确定关键预测变量

本研究问卷调查中充分考虑了汽车消费者可能存在的影响因素,但也可能存在对预测结果影响程度较小的变量。因此,有必要在决策建模之前进行特征选择,确定影响消费者购买意向的重要因素,提高决策树的准确性。

3.3 决策树分类结果

3.3.1 基于汽车能源类型的分类

在对消费选择汽车品牌类型进行分类时,节点分裂评价准则与特征划分选择信息增益率,内部节点分裂的最小样本数为4、修建严重性设为75。在分类分析过程中,自变量选择汽车消费者的性别、年龄、收入情况、购车时考虑的安全性、内饰/装配、排气量/性能、外观/设计感、维修/售后服务、油耗、品牌偏好/口碑共10个变量,因变量选择汽车消费者购买汽车时选择的能源类型,共有新能源车、油耗车、油电混合车3种。

3.3.2 基于汽车品牌类型的分类

在对消费选择汽车品牌类型进行分类时,节点分裂评价准则与特征划分选择信息增益率,内部节点分裂的最小样本数为2、修建严重性设为75。在分类分析过程中,自变量选择汽车消费者的性别、年龄、购车时考虑的安全性、内饰/装配、排气量/性能、外观/设计感、维修/售后服务、油耗、品牌偏好/口碑、价格预算共10个变量,因变量选择汽车消费者购买汽车时选择的能源类型,共有国内品牌、国外品牌两种。

3.4 决策模型精度检验

建立的基于汽车能源类型的分类结果,训练集的分类准确率为92.3%,测试集数据验证的准确率为86.85%,该分类模式在汽车消费者选择何种类型的汽车的判断时的可信度达八成。

建立的基于汽车品牌类型的分类结果,训练集的分类准确率为91.82%,测试集数据验证的准确率为88.28%,该分类模式在汽车消费者选择何种品牌的汽车的判断时的可信度达八成。

4 结语

本研究以决策树的方法构建了汽车消费者购买意愿的分类模型,并通过实际汽车消费者购买意愿数据的实证分析,得出了有关结论,为汽车消费者购买意愿分析提供了一定的参考价值。

但分类的效果还有待优化。分类数据的调查上可采用揭示行为调查方式,对已购车的用户数据进行分析,减少意愿调查中的不确定因素。本研究并未囊括所有的影响消费者购买意愿的因素,后续可以增加具体的汽车品牌与型号、消费者的工作、家庭信息等因素调查,以及汽车消费者在购买过程中可能受到的其他因素的影响,如政府补贴政策、汽车宣传等因素,这些因素也可能影响消费者的购买意愿。

[参考文献]

[1]施浩,于倩,于沐含.关于影响汽车消费者购买行为因素的探究[J].营销界,2022,(22):17-19.

[2]杨一翁,孙国辉,纪雪洪.消费者视角下的综合品牌效应研究——基于汽车品牌的数据[J].中央财经大学学报,2015,(11):105-112.

[3]郑先锋.面向消费者兴趣点的汽车外观评价方法研究[D].江苏:江南大学,2017.

[4]方明婷.基于汽车论坛数据的用户购车意愿分析[D].大连:大连理工大学,2021.

[5]童威,黄启萍.基于改进决策树算法的汽车销售数据挖掘[J].安徽电气工程职业技术学院学报,2017,22(03):104-109.

[6]肖欣莹.汽车销量影响因素分析——基于综合熵权法和VAR[J].重庆工商大学学报(自然科学版),2018,35(05):72-78.

[7]刘道东.基于半参数模型汽车销量影响因素研究[D].山东:山东财经大学,2021.

[8]李全森.基于消费者感知价值私家车选购影响因素研究[D].重庆:重庆交通大学,2021.

[9]陈美云,洪春男,李建邦.使用决策树建立汽车消费者购买意愿分类模式[J].台湾(数据分析),2016,11(2):47-76.

作者简介:

张利青(1995--),女,汉族,安徽黄山人,硕士,广东东软学院,助教,研究方向:信息资源管理。

张毅(2001--),男,汉族,江西吉安人,本科,广东东软学院,助教,研究方向:统计分析与机器学习。