

BP神经网络在我国二手车销量预测中的应用

陈丹纯, 梁美丽

广东外语外贸大学数学与统计学院, 广东 广州

收稿日期: 2022年8月6日; 录用日期: 2022年8月16日; 发布日期: 2022年8月30日

摘要

随着我国经济的发展, 汽车行业得到蓬勃发展。本文基于2001年至2021年我国二手车销量, 分析了影响二手车销量的8个因素的相关性, 通过灰色关联法表明这些因素之间的相关性高, 选择其中关联度高于0.6的7个输入指标, 确定二手车每年销售量为输出指标, 建立BP神经网络模型。根据模型结果可知, 三个数据测试集的确定系数均超过90%, 接近1。训练过程的预测值与真实值描点图几乎一致。这表明模型的预测误差小, 精度高。测试数据集与真实数据的相对误差也很小, 证明基于BP神经网络模型的二手车年销售量预测是合理的。也证实了与传统方法回归分析法相比, 机器学习的预测精度更高。

关键词

二手车, BP神经网络, 预测, 灰色关联度分析

Application of BP Neural Network in the Prediction of Used-Car Sales in China

Danchun Chen, Meili Liang

School of Mathematics and Statistics, Guangdong University of Foreign Studies, Guangzhou Guangdong

Received: Aug. 6th, 2022; accepted: Aug. 16th, 2022; published: Aug. 30th, 2022

Abstract

With the development of my country's economy, the automobile industry has developed vigorously. Based on the sales of used-cars in my country from 2001 to 2021, we analyze the correlation of 8 factors that affect the sales of used-cars in this paper. It is shown that the correlation between these factors is high by the gray correlation method. We select 7 input indicators with a correlation degree higher than 0.6 and estimate the annual sales volume of used-cars as the output indicators to establish a BP neural network model. According to the model results, the determination coefficients of the three data test sets are all over 90%, close to 1. The predicted values

of the training process are almost identical to the ground truth plots, which shows that the prediction error of the model is small and the accuracy is high. The relative error between the test data set and the real data is also small, which proves that the prediction of the annual sales volume of used cars based on the BP neural network model is reasonable. It is also confirmed that the prediction accuracy of machine learning is higher than that of the traditional method regression analysis.

Keywords

Used-Car, BP Neural Network, Prediction, Grey Correlation Analysis

Copyright © 2022 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. 绪论

1.1. 研究背景及意义

近些年,我国工业飞速发展,人们生活水平日益提高,使得他们对于日常出行的交通工具越来越重视。由于汽车的便捷性、价格逐渐能为消费者所承担,越来越多的人会购买汽车。这一现象可从我国快速增长的新车销量可窥。由新车市场所带来的二手车在汽车产业的重要性也逐渐提升。我国政府致力于规范二手车市场以及促进市场的合理交易,每年政府工作报告中都会提出关于促进二手车市场发展的相关政策。2018年,我国全面取消二手车限迁政策,促进二手车消费市场潜力释放[1]。2021年两会期间,我国政府进一步提出取消对二手车交易的不合理限制。《关于推进二手车交易等级跨省通办便利二手车异地交易的通知》落地实施克服了二手车跨区域流通的困难,促进二手车交易秩序的规范化,极大地推动了我国二手车市场的发展,营造良好的交易环境[2]。近五年来,我国二手车转移与新车注册登记数量比例由67%不断上升至111%,且二手车转移登记量远远超过新车注册的,这表明我国二手车市场的活跃程度日益增长。

在政府带头的多方共同努力下,我国二手车市场正处于发展阶段。根据中国汽车流通协会调查报告指出:2021年我国二手车累计年交易量为1758.51万辆,较去年同期增长22.62%,远超专家预计的二手车交易规模。2021年中国新车销量达2627.5万辆,二手车销量约为新车销量的1.5倍,达到发达国家二手车与新车流通量比例,这体现了我国二手车市场具有极大的发展空间和市场潜力[3]。此外,中国汽车协会表示:“由于新车产能受限导致其销量处于萎缩状态,让二手车市场迎来量价齐升。估计2022年我国二手车交易量能超过2000万辆。”[4]这些可观的数据都体现了我国二手车市场极好的发展前景。

但是,由于我国二手车市场起步较晚,仍存在一系列问题,主要如下:一,我国缺乏完整的评估体系和监督管理方法,消费者容易被不良商家欺骗购买车况坏且无保障的二手车。二,由于二手车本身的特殊性,可能出现高价烂车、交易流程不规范且售后服务不完善的问题。因此,本文基于我国二手车市场的发展前景和规模的预测,根据结论提出合理建议,能够促进二手车市场的发展以及带来经济效益,为我国汽车行业和经济发展提供新思路。

汽车是我国的支柱产业之一,二手车销量是衡量中国二手车市场的一个重要指标,二手车市场的蓬勃发展对于我国的经济发展具有重要作用。但由于我国对二手车市场尚处于发展探索阶段,难以避免地是存在评价体系不透明、汽车质量参差不齐等问题,不利于二手车市场发展。如何定量分析影响二手车销量的因素与二手车销量的相关性,找出主要影响二手车销量的因素,并据此建立预测模型,准确预测

二手车销量,了解未来二手车销量趋势,探索二手车销量隐含的信息,为二手车市场发展提供合理建议,防止盲目扩张出现产能过剩,促进二手车市场的稳步发展,与此同时也能够推动我国汽车行业成长与壮大。因此,本文的研究有一定程度上的理论价值和现实意义。

1.2. 国内外研究现状

近年来,二手车市场的发展前景广阔。本文基于国内外学者对二手车销量的研究文献,分别从指标选取和预测模型两方面进行综述。

至今,欧美国家二手车市场已达到成熟阶段,二手车市场的法律法规健全完善、新车市场与二手车市场发展协调和二手车的售后服务完善等,促进了国外二手车市场繁荣发展,国外学者对于二手车销量评估体系和影响因素已经展开了较为系统的研究。

在 20 世纪初期,美国二手车市场是世界上最早出现的二手车交易市场,凭借优秀的历史、庞大的汽车产量、完善的法规和统一的市场定价,目前成为全球最成熟、最大的二手车交易市场。国外学者对于二手车销量的研究多采用机器学习的方法。Fu-Kwun Wang 等人(2011) [5]对中国台湾汽车销量与人均收入、汽车销量、CPI 等影响因素进行建模分析。M Hulsmann 等人(2011) [6]运用时间序列分析和决策树两种方法,对德国和美国月度汽车销量数据进行预测,决策树的预测效果更好。Nitis Monburion 等人(2018) [7]运用随机森林回归、多元线性回归两种方法对德国电子商务网收集的二手车数据预测,随机森林回归误差更小。CK Puteri 等人(2020) [8]基于线性回归方法分析了印度尼西亚车龄、汽车行驶里程、颜色、变速器和汽车类型等多种影响因素,建立了二手车销量预测模型,准确率超过 75%。

1) 预测方法

我国学者常用于预测的方法有回归分析法、BP 神经网络模型等机器学习方法,时间序列分析法等。陈欢(2008) [9]通过灰色理论并建立微分方程的预测模型,汽车销量预测拟合度高,能够提供相关决策保证。李任龙(2015) [10]通过相关性分析确定对二手车市场影响最大的因素,运用灰色预测方法和回归分析方法预测二手车销量,提供战略帮助。宋福军(2016) [11]对 2000 年至 2015 年中国汽车月度销量分别采用指数平滑法预测、ARIMA 模型预测、BP 神经网络模型,后者模型的预测结果最为理想。王书鹏等人(2019) [12]基于 2011 年至 2018 年月度数据,运用 BP 神经网络机器学习方法,由于数据量偏少,导致预测精度一般。桂思思等人(2021) [13]基于某一汽车品牌销量预测,采用整合移动平均自回归模型与线性回归模型相结合预测,为二手车研究提供了理论基础和建模参考。

2) 汽车销量影响因素

危高潮(2009) [14]基于回归分析法得出与汽车销售量相关性较高的 5 个重要因素,分别是汽车产量、公路总里程、国民生产总值、石油消费量、居民收入。赵颖(2014) [15]建立以钢材产量、人均国民生产总值、汽车平均市场价格、人均生活能源、私人汽车拥有量、燃料价格指数、营运载客汽车拥有量、公路里程共 8 个指标,通过主成分分析法预测我国汽车销量。杨庆斗(2020) [16]分析 9 个可能影响汽车销量因素与汽车销量的相关性,均高度关联,分别建立了线性回归模型和 BP 神经网络模型,结果表明后者的预测精度更准确。

综上所述,国内外学者研究汽车市场及预测其销量大多采用回归分析、BP 神经网络、随机森林等机器学习模型以及传统的时间序列分析。通过学者们的研究,易发现机器学习的方法用于预测销量的精度更加准确,但对数据量的要求更高;而选取因素主要有:人均国民生产总值、公路里程、钢材产量等与汽车销量息息相关的指标。

1.3. 研究框架

本文围绕我国二手车年度销量预测该问题展开研究,选取以灰色关联分析法为支撑,分析可能因素

与销量的相关性, 选择相关性高的数据进行预测, 分别选用 BP 神经网络法和回归分析法预测, 建立关于我国二手车销量的预测模型, 根据结论提出建议, 总体研究进程如下图所示。

第一部分为绪论, 了解并分析二手车市场在我国的现状及发展前景, 查阅二手车市场的相关文献, 确定本文的研究思路和采用的方法。

第二部分为我国二手车市场现状, 对二手车市场有大体了解。

第三部分为选用方法的理论描述, 说明理论分析中的变量含义, 方法的理论基础以及优缺点, 便于后续的建模。

第四部分基于 BP 神经网络模型和回归分析法进行实证研究, 预测我国二手车年度销量并分析结果。

第五部分根据前文建模得到结论, 提出建议。

第六部分为总结与展望, 反思与总结本文在研究过程的长处与不足, 并查阅资料提出改进的方向。



2. 我国二手车市场现状

由于同样的预算可以买到性价比更高的二手车, 因此大众对于二手车的接受程度也越来越高。由于自驾出行带来的便捷, 越来越多的人够买汽车的意愿更加强烈。但是, 在购买过程会发现同样的预算可以在二手车市场有更多的选择, 能买到高配置的车型, 性能和安全也有所保障。2018 年全国两会期间, 我国政府报告提出, “将新能源汽车车辆购置税优惠政策再延长三年, 全面取消二手车限迁政策。”^[15] 随着政府相关政策的落地带来更多的红利, 二手车消费市场潜力得到释放, 我国二手车交易量已连续 18 年持续增长, 二手车市场发展前景广阔。

2019 年全国交易量 1492.3 万辆, 但由于疫情的影响, 2020 年二手车交易量为 1434 万辆, 首次出现下滑趋势。根据中国汽车流通协会的调查报告, 上一年我国二手车累计年交易量为 1758.51 万辆, 这表明我国二手车市场保持良好发展趋势。二手车销量在我国汽车销量比重越来越高, 对于发展二手车市场具有重要意义, 能在一定程度上促进新车市场的发展, 又能带动汽车相关行业的发展(图 1)。

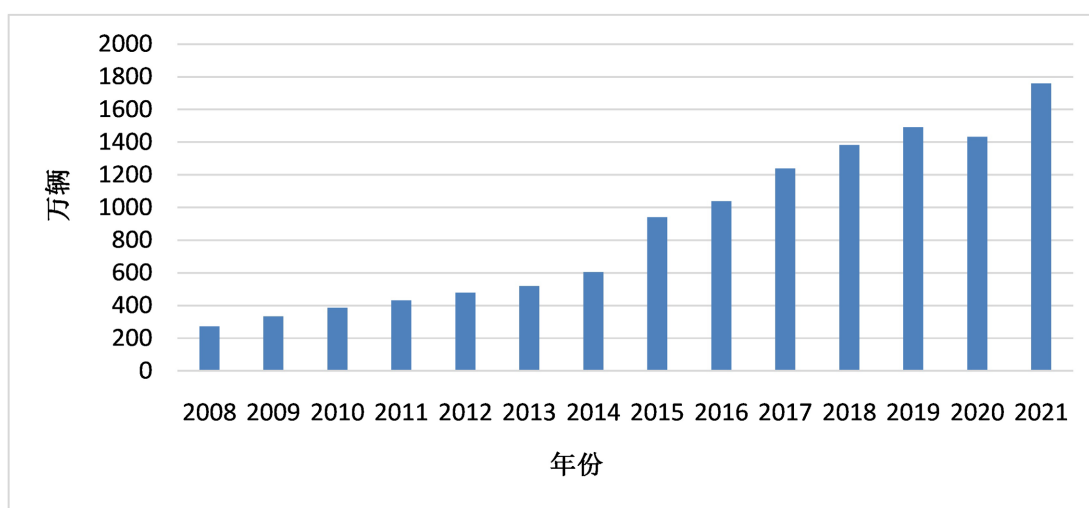


Figure 1. The sales volume of used-cars in China

图 1. 我国二手车销量

虽然二手车市场越来越规范,但由于它在我国发展时间较短,市场机制不够成熟,仍存在以下问题:一是由于二手车的车况经过多方接手,容易致使二手车市场鱼目混珠,可能存在商家为了获利期满消费者的现象。在交易过程,商家无法掌握部分车辆的完整信息,二手车甚至还存在违章事故未处理等。由于消费者对于二手车的行情了解浅薄,就会容易上当被“坑”,买到低质量的二手车;二是评估体系不健全,由于各地评估标准不一、评估过程不够严谨,没有专业的评估设备和体系,商家为了获取更多的利润,会故意造成低价收高价售的状况。三是二手车售后服务问题,相比于新车,二手车的保修服务等可能会有所欠缺,若是商家不负责,买家为了更好的体验,可能会放弃二手车这个选择。

3. 理论基础

3.1. 变量解释

为了简化模型理论基础,做出如下符号规定:

符号	含义
D	原始矩阵
d_{ij}	第 i 个对象的第 j 个指标值
G	评价矩阵
$\xi(d_i)$	比较数列与参考数列在各个时刻(即曲线中的各点)的关联系数
ρ	分辨系数, $\rho > 0$, 通常取 0.5
$\Delta \min$	两级最小差
$\Delta \max$	两级最大差
$\Delta o_i(k)$	各比较数列曲线上的每一个点与参考数列曲线上的每一个点的绝对差值
r_i	灰色关联度
X_0, X_1, \dots, X_n	神经网络输入层
ω_j	X_j 对应权值
F	映射函数
y_i	第 i 个输出层
γ_h	隐藏层第 h 个神经元的阈值
θ_j	输出层第 j 个神经元的阈值

3.2. 灰色关联分析

灰色关联分析法是衡量两因素关联程度的方法之一,它能够体现一个系统因素间发展态势中的相关和不同程度。原理是分析相对序列和比较序列的几何形状相似程度,通过它们的关联程度判断序列的联系是否密切。若曲线间越贴近,则表明序列间的灰色关联度越高,即因素间变化趋势的一致性更高。该方法对于分析的样本数量和规律没有严格要求,计算过程简单易懂,是分析因素间关联程度的常用方法之一。

第一步,在确定指标后,收集数据,从而确定评价体系,记参考序列的评价矩阵为 D ;若评价问题有 m 个对象,每一个对象对应 n 个指标构成,评价矩阵如下所示:

$$D = \{d_{ij}\} \quad \text{公式 1}$$

根据评价目的及指标情况, 假设参考数列为 $D_0 = \{d_{01}, d_{02}, \dots, d_{0n}\}$ 。

第二步, 由于指标的单位含义不同, 为了消除量纲, 先对数据进行标准处理, 得到标准化评价矩阵 G , 其中 $G_0 = \{g_{01}, g_{02}, \dots, g_{0n}\}$ 为处理后的参考数列。

第三步, 分别计算每个被评价对象的指标序列与原有参考数列对应元素间的绝对差值, 如下:

$$|G_i - G_0| = |g_{ij} - g_{0j}| \tag{公式 2}$$

第四步, 确定最小的差值

第五步, 根据公式计算出比较数列与参考数列在各个时刻(即曲线中的各点)的关联系数 $\xi(X_i)$:

$$\xi(X_i) = \frac{\Delta \min + \rho \Delta \max}{\Delta o_i(k) + \rho \Delta \max} \tag{公式 3}$$

第六步, 计算灰色关联度

求比较数列与参考数列间关联程度的平均值, 得到关联程度 r :

$$r_i = \frac{1}{N} \sum_{k=1}^N \xi_i(k) \tag{公式 4}$$

3.3. 神经元模型

假设有 X_0, X_1, \dots, X_n 共 n 个输入层, 其中 $X_0 = -1, \omega_0 = \theta, X_j$ 对应的权值为 ω_j , 则将输入层分别与对应权值相乘求和: $X_1 \times \omega_1 + X_2 \times \omega_2 + \dots + X_n \times \omega_n - \theta$ 记为 net_i , 则通过映射函数 f 映射, 得到输出层 $y_i = f(net_i)$ (图 2)。

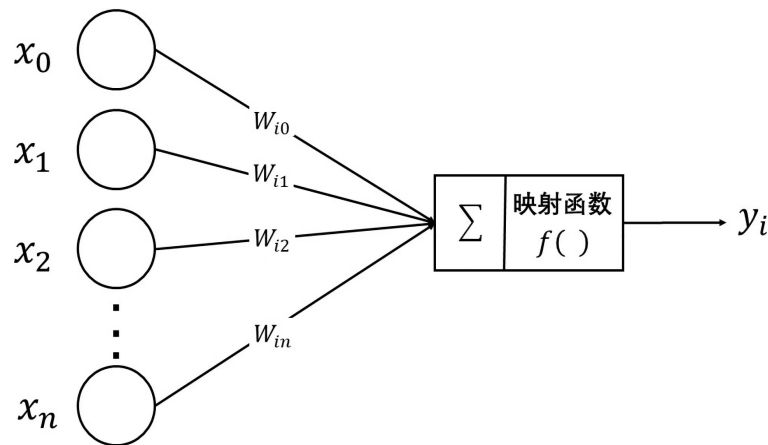


Figure 2. Illustration of the neuron model
图 2. 神经元模型图示

神经元是 BP 神经网络的处理单元。BP 神经网络由多个神经元构成, 就其中单独的一个神经元而言, 它接受来自其他 n 个神经元的输入信号, 并根据两两神经元的权重传递信号, 并对神经元接受的总输入值与阈值进行对比, 通过激活函数的映射, 输出结果。

激活函数是 BP 神经网络的重要组成部分之一, 它能将神经元的输入过程映射到输出层。对于训练 BP 神经网络模型以及处理非线性函数, 激活函数有无可替代的地位。

激活函数有阶跃函数和 Sigmoid 函数。阶跃函数能使得输出层结果离散, 将输入值分别映射为数字 0 或 1, 分别代表神经元抑制和兴奋状态。

$$\text{sgn}(x) = \begin{cases} 1, & x \geq 0; \\ 0, & x < 0. \end{cases} \quad \text{公式 5}$$

因为 Sigmoid 函数处处连续且可导, 并且导数均比 0 大, 它可将输入层的连续实值压缩在 0 和 1 之间(图 3)。

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad \text{公式 6}$$

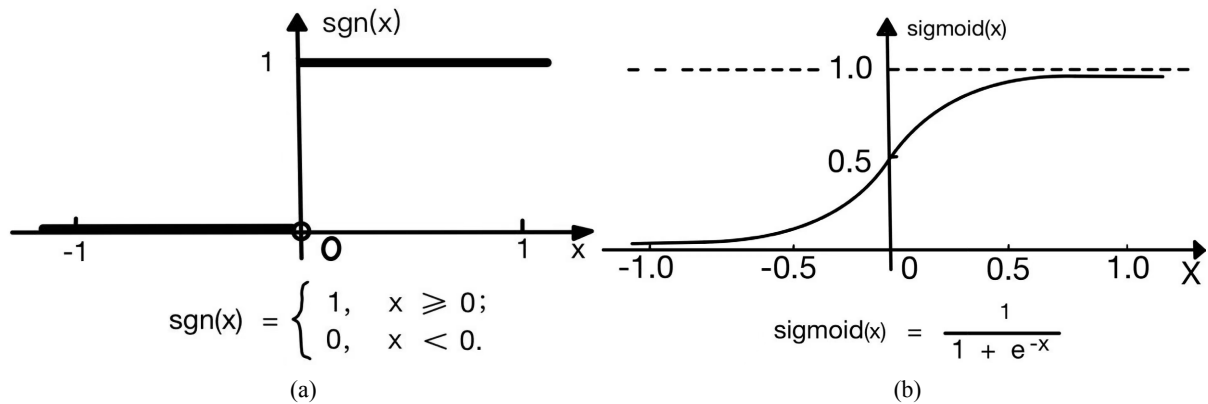


Figure 3. Sgn function & Sigmoid function
图 3. Sgn 函数及 Sigmoid 函数

3.4. BP 神经网络理论

BP 神经网络是 Rumelhart 和 McClelland 等学者研究提出的, 由输入层、隐含层和输出层构成, 分别有 n 、 q 、 m 个节点。通过激活函数处理以及反向传递不断调整对应的权值和阈值, 当网络的输出误差平方达到设定的要求停止。

BP 神经网络包括三层, 分别是输入层、隐含层和输出层。公式推导如下。从输入层到隐藏层:

$$\alpha_h = \sum_{i=1}^d V_{ih} \cdot X_i \quad \text{公式 7}$$

$$b_h = f(\alpha_h - \gamma_h) \quad \text{公式 8}$$

从隐藏层到输出层:

$$\beta_j = \sum_{h=1}^q w_{hj} \cdot b_h \quad \text{公式 9}$$

$$y_j = f(\beta_j - \theta_j) \quad \text{公式 10}$$

则输出 $y_k = (y_1^k, y_2^k, \dots, y_j^k)$ 为 1 维向量。

BP 神经网络模型能在原有的信息量少、数据量有限、不确定性的问题下进行学习训练, 不被非线性模型限制。但在模型的使用过程, 也需要注意如下问题。

1) 节点数

神经网络模型的节点数选择没有统一的标准, 只能通过经验得出的公式来确定计算公式如下, 通常采用第一种:

$$m = \sqrt{n+1} + a \quad \text{公式 11}$$

$$m = \log_2 n \quad \text{公式 12}$$

$$m = \sqrt{nl} \quad \text{公式 13}$$

其中 n 和 l 分别为输入层、输出层节点数, a 为 0 到 10 之间的调节常数。

2) 学习速率

为了确保模型在训练过程的稳定, 模型的学习速率一般设置在 0.01 至 0.8 之间。

因此, BP 神经网络模型能够解决复杂的非线性问题, 在数据量有限的条件下, 全局逼近网络, 在预测方面得到广泛应用。

4. 基于 BP 神经网络的二手车销量预测

4.1. 指标选取

二手车是市场的重要商品之一, 其销量会受到多方面影响。根据王旭文研究, 遵循实际要求, 主要将影响因素分为以下三类: 一为经济发展水平, 例如: 人均 GDP、居民人均可支配收入; 二为汽车市场的影响, 例如: 私人汽车拥有量、新车销量、公路里程、钢材产量、橡胶轮胎外胎产量、汽车驾驶员人数。

数据选取国家统计局官网(<http://www.stats.gov.cn/>)以及中国汽车流通协会官网(<http://www.cada.cn/>), 但由于篇幅过大, 原始数据在附录展示。

4.2. 灰色关联分析

运用 MATLAB R2013 对收集的原始数据进行灰色关联度分析, 计算对二手车销量影响作用的大小, 结果如下表 1 所示。

Table 1. Grey relational grade

表 1. 灰色关联度

相关因素	私人汽车拥有量	新车销量	人均 GDP	居民人均可支配收入	公路里程	钢材产量	橡胶轮胎外胎产量	汽车驾驶员人数
关联性	0.8878	0.735	0.7434	0.7484	0.5901	0.6474	0.6143	0.7696

该表显示, 本文选取的相关因素与二手车市场的关联度均大于 0.5, 总体两辆之间相关性高。

私人汽车拥有量与二手车市场的关联度最高, 与实际一致, 因为二手车的车源大部分来源于私人汽车, 因此, 私人汽车拥有量是至关重要的指标。汽车驾驶员人数关联度位次之, 该人数的增加表示更多人要开车, 买车的购买意愿会逐渐增多。紧接着是居民人均可支配收入和人均国民生产总值, 这表明国家经济发展状况也会对二手车销量造成一定影响。二手车的前身是来自使用后的新汽车, 因此新车销量的关联型较高。

本文选取关联度大于 0.6 的指标, 分别是以下 7 个指标: 私人汽车拥有量、新车销量、人均 GDP、居民人均可支配收入、汽车驾驶员人数、钢材产量和橡胶轮胎外胎产量。

4.3. BP 神经网络

本文将我国二手车年度销量为研究对象, 利用 2001 年至 2021 年共 21 年数据。由于样本数据量较少, 训练集仅有 21 组学习样本, 因此两层的 BP 神经网络已满足要求。

4.3.1. BP 神经网络参数设置

1) 归一化处理

本文采用的软件是 MATLAB R2013a, 将 21 个数据按照 9:1 分为训练集和测试集, 分别将 7 个指标构成的训练输入层和二手车销量训练输出层存放在 P_train 和 T_train 中, 而测试输入层和测试输出层存放在 P_test 和 T_test 中。

为了消除量纲影响, 保证训练效果更加准确, 对所有的数据集归一化处理, 在此处, 运用 MATLAB R2013a 中自 mapminmax 函数将原始数据归一化处理, 映射为(0, 1)之间的数, 取消各维数据间的数量差别, 凸显指标的本质含义, 也能防止出现大数吃小数的现象。

2) 建立 BP 神经网络模型

结合上述分析和多次参数的调试, 建立模型。本模型包含输入层、隐含层、输出层。由于本文选取了 7 个指标, 故输入层节点数为 7 个, 输出层为我国二手车年度销量, 输出层节点数为 1。隐含层节点数根据经验公式(公式 11)确定为 4。经过调试, 训练次数和学习速率分别为 100 次和 0.01 为最佳, 具体参数如下。

训练次数	100 次
误差	0.00001
学习速率	0.01
输入层节点数	7
输出层节点数	1
隐含层节点数	4
激活函数	Sigmoid 函数

4.3.2. 模型结果

BP 神经网络模型对训练集进行学习, 并将测试集带入测试, 得到结果如下: 由图 4 可知, 本网络允许的迭代次数最大为 100, 实际仅迭代了 5 次, 泛化能力检查在神经网络训练过程中始终为 0, 表明没有出现误差不降反升的情况。表明输出误差已在初始设定的范围内, 停止训练。BP 神经网络中最大误差为 0.0643, 设定的目标误差为 0.00001, 实际误差为 2.19×10^{-7} , 表明该网络构建效果好, 具有参考性。

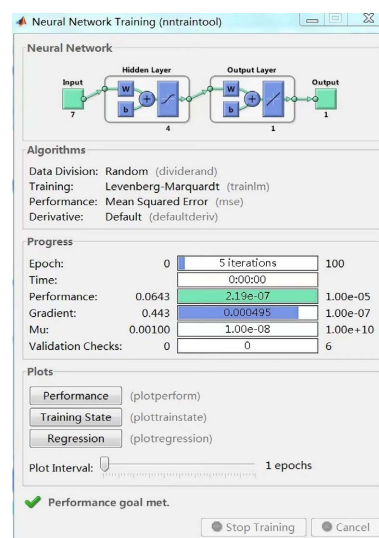


Figure 4. The result of BP neural network

图 4. BP 神经网络训练结果

1) 真实值与预测值的描点图

如图 5, 真实值与预测值几乎接近一致, 仍有个别数据误差较大, 原因是测试数据较少导致, 模型整体拟合预测效果较好。

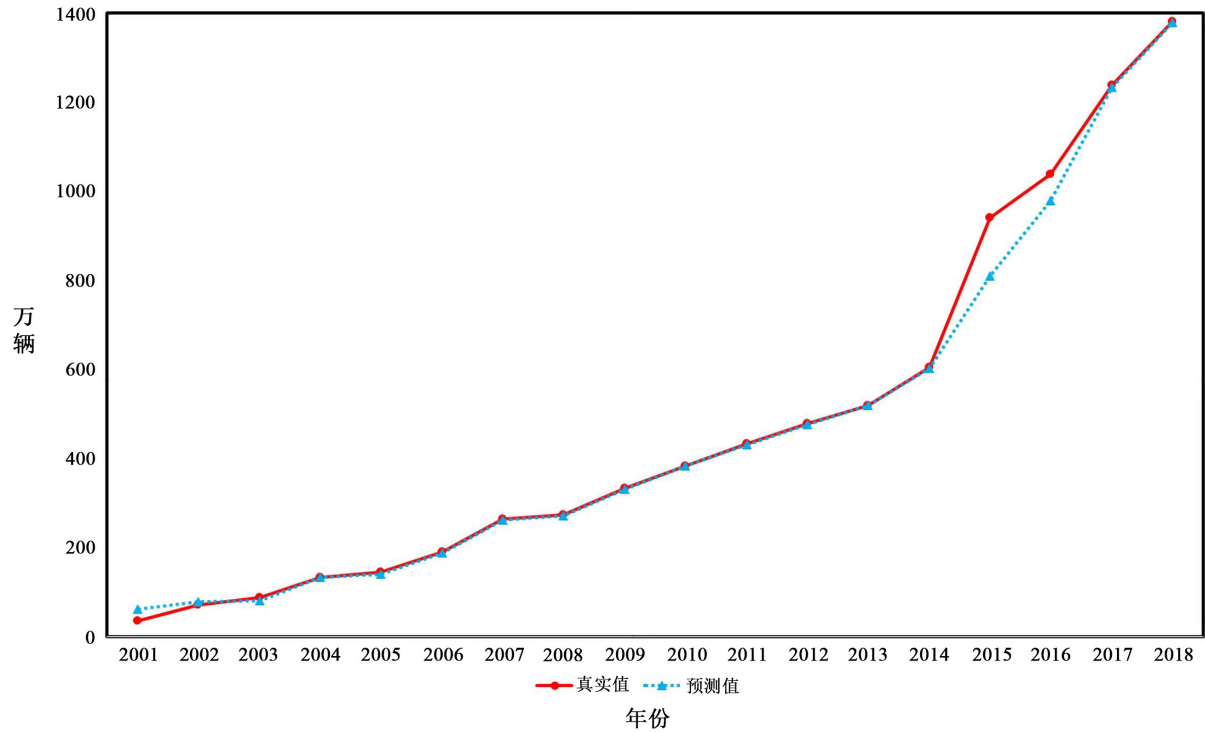


Figure 5. Plot of true value and predicted value
图 5. 真实值与预测值描点图

如图 6 所示, 该网络训练过程中梯度为 0.00049546, 达到设定值, 停止训练。训练迭代到 5 次的时候, 动量因子为 10^{-8} 。

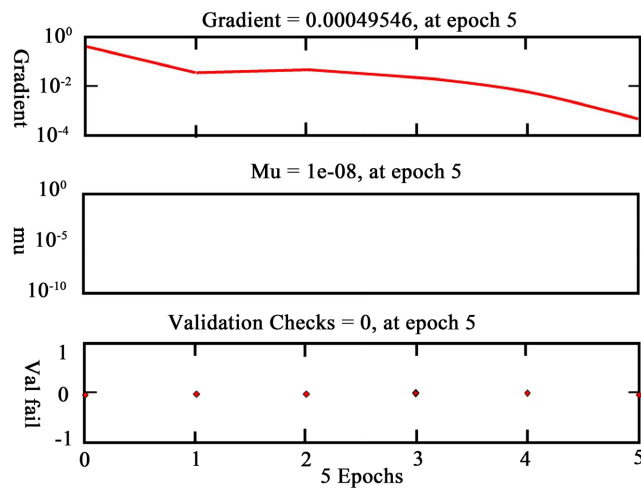


Figure 6. The figure of gradient
图 6. 梯度图

2) 确定系数

R^2 是统计和数据分析中能够体现模型拟合优度的量, 最大值为 1, 计算得出越接近 1, 拟合效果越好。本模型的确定系数如图 7 所示, 训练集、验证集和测试集、训练集的相关系数 R 分别为 1、0.99451、0.9969、0.99592, 可知所有训练数据 R^2 几乎接近接近 1。误差均匀分布在 zero 线附近, 这表明模型的拟合程度高, 证明该模型能够很好地反应选取指标和二手车销量评估的关系。

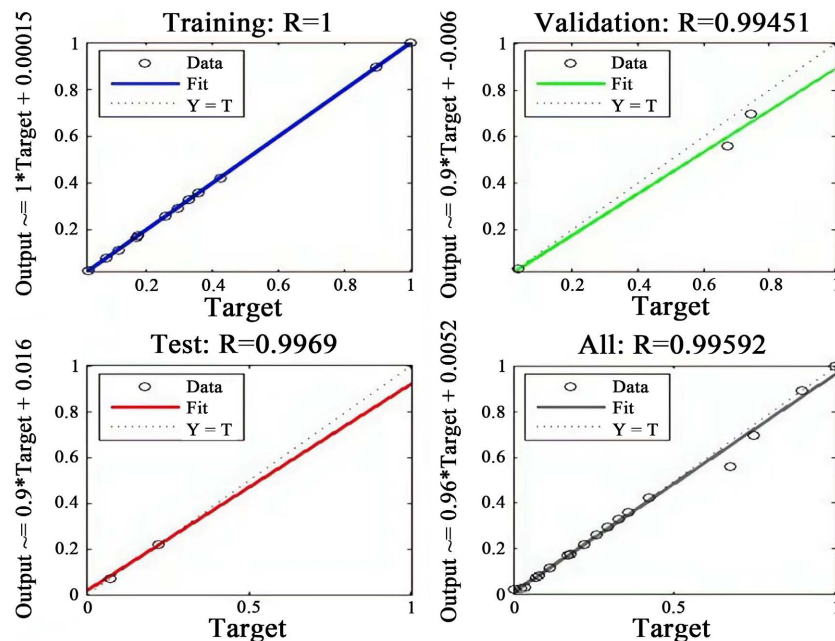


Figure 7. Certain coefficient

图 7. 确定系数

通过前面的网络训练, 预测 2019 年至 2021 年我国二手车的年度销量, 计算相对误差, 如下表 2 所示。

Table 2. The prediction of used-car's sales volume based on BP neural network
表 2. 基于 BP 神经网络的二手车销量预测表

年份	真实值	预测值	相对误差
2019	1492.28	1457.07	2.3595%
2020	1434.14	1496.56	4.3524%
2021	1758.51	1745.85	0.7199%

根据上表的相对误差数据, 可得出 BP 神经网络模型的平均相对误差为 2.4473%, 这表明模型预测精度较高。在实证过程中发现, BP 神经网络每次训练的结果不一样, 查阅相关资料可知原因分别是输入样本是随机排列, 导致输出值不一样; 隐含层权值是随机赋予的。

4.4. 多元线性回归法

利用软件 SPSS26, 将 2001 年至 2018 年我国二手车销量数据进行多元线性拟合, 结果如下: 表格中显著性为 0, 远小于 0.05。表明支持原假设, 即线性回归方程显著(表 3)。

Table 3. The significance**表 3.** 显著性

ANOVA ^a						
模型		平方和	自由度	均方	F	显著性
	回归	2871614.072	7	410230.582	249.464	0.000 ^b
1	残差	16444.466	10	1644.447		
	总计	2888058.538	17			

a. 因变量: 二手车成交量; b. 预测变量: (常量), 橡胶轮胎外胎产量, 私人汽车拥有量, 新车销量, 钢材产量, 汽车驾驶员人数, 人均 GDP, 居民人均可支配收入。

根据回归方程系数表(见附录)可得方程:

$$y = -0.075 \cdot \text{私人汽车拥有量} + 0.144 \cdot \text{新车销量} + 0.048 \cdot \text{人均GDP} - 0.039 \cdot \text{居民人均可支配收入} + 0.069 \cdot \text{汽车驾驶员人数} - 0.022 \cdot \text{钢材产量} + 0.006 \cdot \text{橡胶} - 188.652$$

用方程预测 2019 年至 2021 年的我国二手车销量, 预测结果如下表所示。相对误差十分大, BP 神经网络法的误差小很多(表 4)。

Table 4. The prediction of used-car's sales volume based on multiple linear regression**表 4.** 基于多元线性回归的二手车销量预测表

年份	真实值	预测值	相对误差
2019	1492.28	1258.97	15.6345%
2020	1434.14	999.9305	30.2766%
2021	1758.51	1936.141	10.1012%

5. 建议

针对我国二手车市场的现状, 需要国家政府、相关部门和二手车车商等多方面共同合理, 促进我国二手车市场的完善和壮大, 保证消费者的合法权益, 具体可以从以下措施去推进。

1) 建立二手车车况信息库

由于二手车自身的特点, 车况存在不同程度的影响。因此, 防止车商为了促进销量而篡改二手车车况, 如: 过户次数、变速箱、表显里程等指标。通过汽车监管部门、车商等多渠道合力, 增加汽车信息的透明度, 构建二手车数据库, 使得二手数据真实可靠, 促进二手车交易公平合理, 保证消费者的合法权益, 推动二手车市场发展。

2) 培训二手车评估师

相比于其他行业的从业工作者, 对于二手车评估师的培训以及考核存在欠缺。为了保证现有评估过程更加公正, 应该加强对二手车评估师专业知识的培训, 使得二手车价格评估更加科学, 能一定程度改善二手车市场鱼目混珠的情况。

3) 完善的售后服务

消费者之所以最终放弃购买二手车, 是因为二手车的售后服务, 相比于新车可能存在诸多麻烦。因此, 可以借鉴西方国家的二手车售后服务经验, 使消费者能够安心买二手车, 权益也能够得到保障, 敢于买二手车。

4) 充分利用电商平台

近些年来,我国电商平台发展快速且成熟。二手车市场可以利用电子商务平台自身存在的优势,减少一些可省的费用,如铺租、汽车保养等。此外,这也能够让消费者足不出户,货比三家,对二手车车况等有了解,推动二手车市场的发展。

6. 总结及展望

6.1. 总结

近几年,我国二手车销量飞跃增长,这也体现了越来越多的消费者购买二手车的意愿且二手车市场在我国的极好的发展前景,因此促进二手车市场发展以及协调好与新车市场的关系,进而推动我国重要支柱产业之一汽车蓬勃发展。伴随着人们消费观念的改变以及国家的政策法规引导,二手车市场在我国快速发展,发展势头猛。但至今为止,二手车在我国仍处于起步阶段,需要各方面配合努力。

本文总结阐述了我国二手车市场的现状,2021年,我国二手车年交易量为1758.51万辆,较去年同期增长35.34%,远超预计交易规模。在我国二手车年度销量预测问题中,采用了私人汽车拥有量等10个指标。根据灰色关联分析显示选用的指标与二手车销量有很高的关联度,选取关联度高于0.7的6个指标,作为BP神经网络的输入层,得到每年我国二手车销量的预测结果。神经网络模型三个数据集确定系数均超过99%,接近1,表明模型拟合效果很好。测试数据集与真实数据的相对误差为2.3595%、4.3524%、0.7199%,证明BP神经网络模型预测是合理的。

6.2. 不足和展望

本文基于BP神经网络模型对我国二手车年销售量预测效果较好,但是,二手车价值评估蕴含的信息量、影响因素都比较多,研究肯定还有不足之处,需要在以下几个方面继续加强:

1) 本文采用的数据是2001年至2021年我国二手车年销售量共21个数据样本,数据量较少,若获得更多的数据量,这才能使效果更加精准,结果更加可靠。

2) 本文仅采取了BP神经网络方法进行预测,应该多引入其他机器学习的方法进行预测,再比较效果,使得评估体系更加精准。

3) 本文根据文献采取7个影响二手车年销售量的指标,指标之间存在的相关性较高的问题,可以尝试加入一些其他因素改善模型的预测质量。

4) 该BP神经网络模型预测我国二手车年销售量精度高,但若要预测未来几年的二手车年销售量,需要预测7个指标的数据,并代入模型预测。

参考文献

- [1] 我国全面取消二手车限迁政策[EB/OL]. http://www.gov.cn/xinwen/2018-03/05/content_5271005.htm, 2018-03-05.
- [2] 商务部、公安部、税务总局联合出台便利二手车异地交易登记新举措[EB/OL]. http://www.gov.cn/xinwen/2021-04/20/content_5600768.htm, 2021-04-20.
- [3] 再现政策利好,二手车行业迎来发展新机遇[EB/OL]. <https://baijiahao.baidu.com/s?id=1698236720903458416&wfr=spider&for=pc>, 2021-04-28.
- [4] 2022年中国二手车行业发展前景预测分析[EB/OL]. <https://www.askci.com/news/chanye/20220110/1804491721152.shtml>, 2022-01-10.
- [5] Wang, F.-K. and Chang, K.-K. (2011) Using Adaptive Network-Based Fuzzy Inference System to Forecast Automobile Sales. *Expert Systems with Applications*, **38**, 10587-10593. <https://doi.org/10.1016/j.eswa.2011.02.100>
- [6] Hulsmann, M., Borscheid, D., Friedrich, C.M. and Reith, D. (2011) General Sales Forecast Models for Automobile Markets Based on Time Series Analysis and Data Mining Techniques. *Industrial Conference on Data Mining*, New York, 30 August-3 September 2011, 255-269. https://doi.org/10.1007/978-3-642-23184-1_20

- [7] Monburinon, N., Chertchom, P., Kaewkiriya, T., Rungpheung, S., Buya, S. and Boonpou, P. (2018) Prediction of Prices for Used Car by Using Regression Models. *5th International Conference on Business and Industrial Research*, Bangkok, May 2018, 17-18.
- [8] Puteri, C.K. and Safitri, L.N. (2020) Analysis of Linear Regression on Used Car Sales in Indonesia. *Journal of Physics: Conference Series*, **1469**, Article ID: 012143. <https://doi.org/10.1088/1742-6596/1469/1/012143>
- [9] 陈欢. 灰色理论在汽车销售预测和投资中的应用研究[D]: [硕士学位论文]. 合肥: 合肥大学, 2008.
- [10] 李任龙. 基于灰色理论的二手车市场发展研究[D]: [硕士学位论文]. 西安: 长安大学, 2015.
- [11] 宋福君. 我国汽车产量的分析与预测研究[D]: [硕士学位论文]. 沈阳: 东北大学, 2016.
- [12] 王书鹏, 连恒鹏, 王涛, 黄素珍, 刘桂兰. 机器学习在汽车销量预测中的应用[J]. *中阿科技论坛*, 2019(2): 22-36.
- [13] 桂思思, 孙伟, 徐晓锋. 基于 ARIMA 与线性回归组合模型的汽车销量预测分析[J]. *计算机与数字工程*, 2021, 49(8): 1719-1722.
- [14] 危高潮. 我国汽车销量主要影响因素的分析[J]. *西安财经学院学报*, 2009, 22(6): 88-92.
- [15] 赵颖. 基于回归分析的我国汽车销量预测模型[D]: [硕士学位论文]. 武汉: 华中师范大学, 2014.
- [16] 杨庆斗. 基于 BP 神经网络的长城汽车销量预测研究[D]: [硕士学位论文]. 天津: 天津财经大学, 2020.

附录

原始数据

年份	二手车成交量	私人汽车拥有量	新车销量	人均GDP	居民人均可支配收入	汽车驾驶员人数	钢材产量	橡胶轮胎外胎产量
2001	37.01	770.78	236.4	8622	4070	4462.68	16067.61	13573
2002	71.33	968.98	324.8	9398	4532	4827.08	19251.59	16306.59
2003	87.89	1219.23	439.1	10542	5007	5368.07	24108.01	19311.96
2004	134.08	1481.66	507.1	12336	5661	7107.64	31975.72	32709.37
2005	145.02	1848.07	575.1	14185	6385	8017.76	37771.14	34390.06
2006	190.59	2333.32	720	16500	7229	9317.24	46893.36	43547.08
2007	265.76	2876.22	879.15	20169	8584	10567.1	56560.87	55832.97
2008	273.73	3501.39	930	23708	9957	12276.8	60460.29	51956.94
2009	333.86	4574.91	1300	25608	10977	13740.73	69405.4	65601.56
2010	385.19	5938.71	1806	30015	12520	15129.89	80276.58	77611.83
2011	433	7326.79	1850.51	35181	14551	17416.76	88619.57	83566.22
2012	479.14	8838.6	1900	38420	16510	20028.52	95577.83	89370.49
2013	520.24	10501.68	2189.4	41909	18311	21742.7	108200.54	110440.12
2014	605.29	12339.36	2349.19	46531	20167	24812.07	112513.12	111913.11
2015	941.71	14099.1	2460	49922	21966	28012.99	103468.41	92831.39
2016	1039.07	16300.22	2802.8	53783	23821	30328.77	104813.45	94974.05
2017	1240.09	18515.11	2887.89	59592	25974	31685.2	104642.05	92789.58
2018	1382.19	20574.93	2808.1	65534	28228	36923.42	113287.33	88608.72
2019	1492.28	22508.99	2576.9	70078	30733	39752.86	120456.94	84445.29
2020	1434.14	24291.19	2531.1	71828	32189	41794.89	132489.19	80747.48
2021	1758.51	27650	2627.5	80976	35128	44000	103279	89910.8

回归方程系数表

模型	未标准化系数		标准化系数	t	显著性	共线性统计
	B	标准错误	Beta			容差
(常量)	-188.652	124.452		-1.516	0.161	
私人汽车拥有量	-0.075	0.048	-1.177	-1.552	0.152	0.001
新车销量	0.144	0.091	0.329	1.585	0.144	0.013
人均 GDP	0.048	0.024	2.143	2.002	0.073	0.000
居民人均可支配收入	-0.039	0.071	-0.742	-0.546	0.597	0.000
汽车驾驶员人数	0.069	0.022	1.696	3.060	0.012	0.002
钢材产量	-0.022	0.009	-1.819	-2.310	0.044	0.001
橡胶轮胎外胎产量	0.006	0.006	0.486	1.073	0.309	0.003