

基于智能算法的“互联网+”时代二手车 评估定价模型研究

杨 波

(昆明理工大学 质量发展研究院,云南 昆明 650093)

摘 要:“互联网+”战略快速推行,二手车交易市场进入新的发展阶段。本文针对二手车交易中评估定价问题,提出了基于智能算法的二手车评估定价模型:首先利用因子分析法探索影响二手车定价的重要因素,然后利用 MIV 变量筛选方法确定神经网络输入变量,最后构建基于 BP 神经网络智能算法的二手车评估定价模型。实例分析表明,所建预测模型比现有预测模型预测的更为准确、稳定。

关键词:BP 神经网络;MIV;变量筛选;二手车评估定价

中图分类号:F224.9 **文献标识码:**A

DOI:10.13395/j.cnki.issn.1009-0061.2017.05.005

一、引言

二手车是指在办理完登记手续到达国家制定的报废标准之前进行交易并转移所有权的汽车^①。近年来,随着“互联网+”战略的快速推行,消费者通过二手车交易网站可以选择的范围和空间不断扩大,二手车市场进入新的发展阶段^②。阻碍二手车交易市场进一步发展最核心的问题就是二手车价格和质量的匹配问题^③,这也是行业内学者研究的重点课题。

在欧美、瑞士、韩日等国家,均具有专门资格的相关部门,根据汽车市场行情,通过质量综合检测,结合科学的鉴定评估标准,最终确定二手车实际销价^④。与国外发达国家成熟市场相比,我国二手车市场主要存在交易信息不透明、诚信机制不健全、评估价值不可信等问题^{⑤⑥}。合理评估二手车价值,提高客户对二手车的信任度,使二手车估价与交易工作步入科学化、规范化的管理轨道,有利于二手车流通市场的发展。目前常见的二手车评估价值的方法主要有重置成本法、现行市价法、收益现值法和清算价格法^⑦。重置成本法基于平均年限折旧法,并结合眼观、手摸、试驾等过程,完全依据经验来进行价值评估^⑧,具有一定的片面性。现行市价法需要基于公开活跃的市场,并且不易寻找参考对象。收益现值法较易受到人为主观和未来不可测因素的影响。

清算价格法当前在我国缺乏实践应用。2014年6月1日,国家质检总局和国家标准委联合发布《二手车鉴定评估技术规范》^①。此文件较为清晰地改进了基于重置成本法的估值方法,对规范二手车鉴定评估行为具有参考性价值,但它也只是一种推荐性标准,不适合作为当前唯一的定价方式。很多学者将回归分析、Bootstrap 法、GM、AHP、模糊理论等方法应用于评估定价,但由于误差较大,评价结果不尽如人意^⑨。

鉴于上述分析,本文为探索“互联网+时代”二手车科学合理、贴近实际的评估定价方法,提出了一种基于 MIV-BP 的二手车评估定价预测模型。

二、相关原理介绍

BP 神经网络。由于影响二手车价值的因素众多,参数复杂,预测难度较大,普通的评估定价方法难以综合考虑所有指标变化对二手车价值的影响,所以一直难以建立较为有效地二手车评估定价模型。线性回归方法用于处理多因素、非线性的价格评估问题预测精度过低,误差较大;误差反向传播(Back Propagation, BP)神经网络采用梯度搜索技术反传学习,通过不断调整连接权值和阈值,可以得到较好的预测精度。BP 神经网络拓扑结构如图 1 所示;其优点在于:(1)具有良好的鲁棒性和容错性;(2)并行处理信息,计算速度快;(3)具有自学习、自

收稿日期:2017-02-20

基金项目:云南省科技支撑项目“基于 Web 的发动机质量测试与需求调研管理信息系统技术开发”(项目编号:KKSTJ201358015)

作者简介:杨波(1969—),男,博士研究生,高级工程师,主要研究方向:数据挖掘和技术经济研究。

组织和自适应性,能够处理不确定问题;(4)能以无限精度逼近任意平滑函数;(5)具有较强的信息综合能力,能够同时处理定性和定量信息。

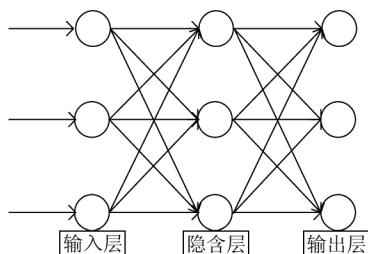


图1 神经网络结构

鉴于BP神经网络的诸多优点,能方便地针对二手车价值进行建模和分析,得出有效地预测和估计。首先,对所有变量进行归一化处理,消除量纲和数量级的影响。其次,对实例样本进行模型训练,当最终误差满足精度要求或者迭代次数的要求时,得到关于影响因素的权值矩阵。最后,进行模型检验,只有通过检验的模型才是科学合理的模型。对于有效的模型,通过比较不同影响因素的权值大小,就可以得到此影响因素对二手车价值的影响程度,从而建立起一个相对比较合理的二手车价值评估模型。此模型能够根据影响因素的变化,直接对二手车价值进行调整和预测。在MIV变量筛选中,平均影响值(Mean Impact Value, MIV)是指用于反映输入层神经元对输出层神经元影响大小、评价变量之间相关性最好的指标,其符号代表相关方向,绝对值大小代表影响的相对重要性^①。

三、指标体系的分析与设计

(一)前提假设

对二手车的评估定价必须以某些假设条件作为前提基础,否则评估定价活动无法进行,或者评估结果毫无意义。二手车评估定价的具体前提假设如下:(1)待估二手车将按现行用途继续使用,或者将转换用途继续使用;(2)待估二手车可以在自由竞争的公开市场上按市场原则进行交易;(3)只针对待估二手车评估基准日的价值。

(二)原始指标的统计与选取

其指标有: X_1 :累计使用年限; X_2 :累计行驶里程; X_3 :品牌认可度; X_4 :车辆市场保有率; X_5 :同款新车价格; X_6 :销售市场销售状况; X_7 :同款车故障率; X_8 :维修保养记录; X_9 :配件供应情况; X_{10} :保险险种及有效期; X_{11} :当地税收及环保政策; X_{12} :车辆行驶路况; X_{13} :同款新车保值率; X_{14} :售后服务情况; X_{15} :整车技术状况。

(三)二手车价值影响因素数字化

综合考虑各个因素,整理相应数据,在同一变量范围的基础上进行各个因素的数字化,为数学模型的建立和分析打下基础。值得注意的是,不同数据需不同的处理方法,具体如下:

$$\tilde{x} = \frac{x - \min}{\max - \min} \quad (1)$$

式中: \max 、 \min 分别为同一影响因素不同车辆的数值中的最大值和最小值。对于定序尺度数据,拟进行如下式(2)的数字化处理:

$$\tilde{x} = \frac{\max - x}{\max - \min} \quad (2)$$

式中: \max 、 \min 分别为同一影响因素不同车辆的数值中的最大值和最小值。对于等级尺度数据,拟采用人工赋值方法;将最高等级赋值为1,依次降低0.2。对于同一影响因素不同条件情形,拟采用算术平均值法进行综合数字化。

(四)影响二手车价值的因子分析

利用SPSS统计软件,对258个二手车交易实例样本15个初始指标的原始数据进行预处理,进行KMO和Bartlett球度检验,P值小于0.05,结果为0.796,其总方差解释表如表1所示。

表1 总方差解释表

因子	特征值	方差贡献率	累计贡献率
1	11.092	71.761	71.761
2	1.95	12.62	84.381
3	1.289	8.34	92.721
4	0.46	2.977	95.698

由表1易知前3个因子的累计贡献率为92.721%>90%,能够比较全面地反映总体信息;碎石图显示,由于因子1、2、3之间的坡度较其他陡峭,并且特征值均不小于1,其旋转后的因子载荷矩阵如表2所示。

根据表2可知,公因子F1包含信息量加大的指标主要有 X_1 、 X_2 、 X_8 、 X_{12} 、 X_{15} ;公因子F2包含信息量加大的指标主要有 X_5 、 X_3 、 X_4 、 X_{13} 、 X_7 、 X_{14} 公因子F3包含信息量加大的指标主要有 X_6 、 X_9 、 X_{10} 和 X_{11} 。综合来看,F1反映了车辆在存放和使用的过程中,由理化原因或自然力作用造成的车辆实体发生的价值损耗,命名为实体性因子;F2反映了由于技术陈旧、性能落后而导致的车辆无形损耗,命名为功能性因子;F3反映了宏观经济政策、市场需求与供应等外部经济环境变化而引发的车辆贬值,命名为经济性因子。

(五)MIV实现变量的筛选

利用MATLAB编程得到15个MIV如下所示:

表2 旋转后的因子载荷矩阵

初始指标	F ₁	F ₂	F ₃
X ₁ : 累计使用年限	0.918	0.342	0.207
X ₂ : 累计行驶里程	0.901	0.139	0.23
X ₁₂ : 车辆行驶路况	0.823	0.036	0.348
X ₈ : 维修保养记录	0.727	0.29	0.04
X ₁₅ : 整车技术状况	0.699	0.205	0.27
X ₅ : 同款新车价格	0.35	0.92	0.331
X ₃ : 品牌认可度	0.303	0.834	0.059
X ₄ : 车辆市场保有率	0.198	0.795	0.433
X ₁₃ : 同款新车保值率	0.112	0.794	0.3
X ₇ : 同款车型故障率	0.07	0.8	0.335
X ₁₄ : 售后服务情况	0.05	0.811	0.322
X ₆ : 销售市场销售状况	0.03	0.502	0.932
X ₉ : 配件供应情况	0.396	0.435	0.717
X ₁₀ : 保险险种及有效期	0.123	0.321	0.819
X ₁₁ : 当地税收及环保政策	0.02	0.22	0.742

MIV_01=-8.103; MIV_02=-0.788;
MIV_03=+2.675; MIV_04=+3.223;
MIV_05=+6.743; MIV_06=+1.667;
MIV_07=-4.328; MIV_08=+3.897;
MIV_09=-2.175; MIV_10=+6.119;
MIV_11=-0.974; MIV_12=-0.693;
MIV_13=+0.583; MIV_14=+3.834;
MIV_15=+7.123;

因此, MIV 变量筛选结果为: 累计使用年限, 累计行驶里程, 整车技术状况, 同款新车价格, 保险险种及有效期。基于 BP 神经网络的 MIV 变量筛选方法的具体流程参见下图 2:

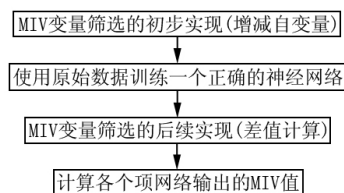


图2 基于 BP 神经网络的 MIV 变量筛选

四、BP 神经网络模型的构建与训练

(一) 样本集合的分配

不同于以往将样本集合依据相关理论与经验随机分为两部分的作法, 本文拟将所采用的样本数据分成三部分以避免训练过程中出现的“过拟合”情形。在三部分样本数据中, 52 个实例数据(约占实例数据总数的 20%) 为变量样本, 180 个实例数据(约占实例数据总数的 70%) 为测试样本, 10 个实例数据(约占实例数据总数的 10%) 为训练样本。

(二) 网络模型的参数设置

鉴于二手车评估定价的模型特定, 可知网络的输入层的节点数为 15, 输出层节点数为 1, 网络拓扑结构的层数为 3。最佳隐含层节点数的选取可参考如下公式^[12]:

$$q < p - 1, q < \sqrt{p + r} + \theta, q = \log_2 p$$

式中: p 为固定不变的输入层节点数, q 为可调整变化的隐含层节点数, r 为固定不变输出层节点数, $\theta \in [0, 10]$ 。将 $p=15, r=1$ 带入上式, 得到: $q < 14, q < 4 + \theta, q = 3.9$ 。故隐含层节点数初始值为 $q=5$ 。运行数学统计软件, 根据出图效果, 不断调整数值; 最终确定最佳的隐层节点数为 8。据此, 最佳的网络拓扑结构为: $15 \times 8 \times 1$ 。

(三) 网络模型的实现

为使实例数据具备优良的典型性和泛化能力, 本文首先采用 MATLAB 软件 dividevec 函数重新随机抽取上述三种分类的数据, 打乱原来的顺序。然后结合马克威统计软件, 完成二手车评估定价 BP 神经网络模型的训练。

(四) 其他方法重复实现与比较分析

为对比研究本文所提方法的可行性和有效性, 同时以基于 15 个影响因素多元线性回归分析方法(称为模型 1)、基于初始影响因素的多元线性逐步回归分析方法(称为模型 2)、基于 3 个公共因子的多元线性回归分析方法(称为模型 3)、基于 3 个公共因子的 BP 神经网络分析方法(称为模型 5)、基于 MIV 变量筛选的 BP 神经网络分析方法(模型 6)重复上述操作步骤, 得到的误差数据与本文上述方法(模型 4)的误差数据比较分析如表 4 所示。由表 3 模型 1 与模型 4、模型 3 与模型 5 的误差比较发现: BP 神经网络方法对二手车评估定价问题的建模分析具有可行性。由上表模型 1 与模型 2、模型 4 与模型 5 的误差比较发现: 变量选取的个数对二手车的评估价值具有一定的影响。由上表模型 2 与模型 3、模型 5 与模型 6 的误差比较发现: 影响因素的选择对二手车评估定价具有一定的影响。然后进一步对六种模型作统计量显著性检验, 检验结果如表 4 所示。由表 4 可知, 模型六应用于二手车评估定价问题, 值小于 0.05, 其模型拟合度较其他方法更理想。

五、结论与意义

二手车交易不仅能够促进新车销售、承担比新车更多的销量和利润, 而且也能够带动汽车服务业的发展、推动汽车售后服务的信息化建设。二手车评估定价是互联网+二手车交易的焦点问题。本文采用理论与实例相结合的分析方法, 对二手车评估

表3 六种模型的误差比较结果分析表

模型	变量样本				测试样本				训练样本			
	最大误差	最小误差	绝对误差	平均误差	最大误差	最小误差	绝对误差	平均误差	最大误差	最小误差	绝对误差	平均误差
1	0.945	0.267	0.678	0.671	0.896	0.3	0.596	0.667	0.988	0.299	0.689	0.626
2	0.705	0.212	0.593	0.554	0.879	0.298	0.551	0.59	0.746	0.223	0.523	0.524
3	0.809	0.382	0.527	0.473	0.616	0.212	0.404	0.497	0.742	0.323	0.419	0.497
4	0.715	0.376	0.339	0.533	0.655	0.356	0.299	0.576	0.5	0.294	0.306	0.553
5	0.51	0.307	0.203	0.445	0.567	0.322	0.245	0.426	0.403	0.192	0.211	0.462
6	0.485	0.293	0.192	0.316	0.287	0.109	0.178	0.384	0.294	0.196	0.098	0.338

表4 六种模型的F检验结果比较分析表

模型	1	2	3	4	5	6
判定系数	0.785	0.814	0.886	0.866	0.907	0.938
调整后的判定系数	0.767	0.795	0.869	0.828	0.892	0.915
F值	268.2	271.45	312.76	398.16	432.88	596.04
P值	0.27	0.04	0.07	0.15	0.019	0.00

定价问题进行了研究。首先,通过实例分析,总结出了影响二手车评估定价的15个初始指标。然后,基于多元逐步回归、PCA、MIV变量筛选等方法分别对初始指标体系进行了分析和设计。最后,基于多元线性回归和BP神经网络方法分别对精简设计的筛选指标进行了预测分析。通过平均误差对比分析,结果表明:基于BP神经网络MIV变量筛选的二手车评估定价模型的预测性能较之其他模型的预测性能更为理想。

此研究的主要意义还在于:(1)通过矩阵计算,定量明确了各个影响因素对二手车价值影响程度的大小;(2)通过模型分析,能够给出二手车价值的合理区间,进一步规范二手车交易;(3)良好的鲁棒性使得此模型能够有效应对新指标的引入、外界环境

变化等突发状况;(4)基于智能算法的评估定价模型能够应用未来大数据发展的要求,模型在未来具有广阔的应用前景。

注释:

①中华人民共和国质量监督检验检疫总局,中国国家标准化管理委员会.二手车鉴定评估技

术规范[A].2014-06-01.

②⑥隋博.信息化条件下现行市价法在二手车评估中的应用[J].轻工科技,2014,(11):115-116,135.

③冯秀荣,王斌.影响二手车价值的因子分析[J].商业研究,2008,(02):102-105.

④王斌,葛新权.二手车价值评估中有关问题的探讨[J].价格理论与实践,2008,(11):53-54.

⑤赵雷明.二手车评估现状分析[J].企业导报,2015,(03):88-89.

⑦王钰斌.现行市价法在二手车评估中的应用[J].武汉商业服务学院学报,2014,(02):33-35.

⑧丁海波,牛丽.基于改进AHP-Fuzzy的二手车性能综合评估研究[J].机电工程技术,2013,(12):79-82.

⑨曹静娴.基于数据挖掘技术的二手车性能评估模型研究[J].商场现代化,2014,(24):17-18.

⑩Stephen Grossberg. Nonlinear neural networks: principles, mechanisms, and architectures[J]. Neural Networks, 1988, 1(1): 17-61.

⑪卢纹岱.SPSS统计分析(第4版)[M].北京:电子工业出版社,2012.

⑫Thomas Andrews, Cynthia Benzing. The determinants of price in internet auctions of used cars[J]. Atlantic Economic Journal, 2007, 35(1):43-57.

Model of Evaluation and Price Setting of Second-hand Car in the Internet Plus Era based on Intelligent Algorithm

Yang Bo

(Quality Development Institute, Kunming University of Science and Technology, Kunming Yunnan 650093, China; Yunnan Institute of Metrology and Testing Technology, Kunming 650093, China)

Abstract: With the promotion of "Internet Plus" strategy, second-hand car market is undergoing a new development. To deal with the evaluation and price setting in second-car transactions, we propose a new model based on intelligent algorithm. Firstly, factor analysis is used to investigate the major factors affecting the price setting of second-hand car. Then, MIV variable screening method is adopted to determine the input variables of neural network. Finally, a new second-hand car evaluation and price setting model based on BP neural network intelligent algorithm is built. Case analysis has shown that the built model is more precise and stable than other prediction models.

Key words: BP neural network; MIV; variable screening; second-hand car evaluation and price setting