

人工智能在二手车特征及价值评估中的应用研究

黄乐

江苏车置宝信息科技股份有限公司, 江苏 南京 210000

摘要 : 随着人工智能的发展, 二手车特征及价值评估通过人工智能涌现出速度快、并行处理量大、准确度高的通用能力, 在二手车交易、金融、保险的商业场景中价值凸显。本文首先研究了个人在二手车价值评估中客观存在的传统计算方法, 有限历史成交经验和记忆容量受限的痛点。随后, 探讨了人工智能的单体原理及应用前景, 提出了基于人工智能的二手车特征及价值评估的科学应用, 包括数据挖掘、特征预处理、数据编码、模型计算等, 同时分析了机器学习方法包含线性、深度学习、决策树、神经网络等回归、分类算法及引用案例; 最后, 展望了人工智能在二手车特征及价值评估中赋能及解决个人痛点的价值。

关键词 : 人工智能; 二手车特征; 二手车价值评估; 机器学习

Research on the Application of Artificial Intelligence in the Feature and Value Evaluation of Used Cars

Huang Le

Jiangsu Chezhibao Information Technology Co., Ltd., Nanjing, Jiangsu 210000

Abstract : With the development of artificial intelligence, the feature and value evaluation of used cars have emerged with fast speed, large parallel processing capacity, and high accuracy through artificial intelligence, highlighting its value in the business scenarios of used car transactions, finance, and insurance. This article first studies the objective existence of traditional calculation methods in the value evaluation of used cars by individuals, as well as the pain points of limited historical transaction experience and limited memory capacity. Subsequently, it explores the individual principles and application prospects of artificial intelligence, and proposes scientific applications of artificial intelligence-based feature and value evaluation of used cars, including data mining, feature preprocessing, data encoding, model calculation, etc. At the same time, it analyzes machine learning methods including linear regression, deep learning, decision trees, neural networks, and other regression and classification algorithms, along with citation cases. Finally, it looks forward to the value of artificial intelligence in empowering and solving personal pain points in the feature and value evaluation of used cars.

Keywords : artificial intelligence; used car features; used car value evaluation; machine learning

引言

伴随中国经济的强劲增长, 过去十年, 我国二手车交易从605万辆/年增长至1841万辆/年, 实现三倍增长。需求量的激增, 准确二手车价值评估是消费者需求也是商家降本增效的关键能力。评估师个人基于历史成交价格和专业知通过现行市价法、重置成本法计算二手车价值(以下简称:传统二手车评估)。相较人工智能通过机器学习挖掘特征和模型计算输出二手车价值(以下简称:智能二手车评估)^[1]。本文将研究多种人工智能机器学习框架, 提出多种智能二手车评估的科学技术路径。

一、传统二手车评估个人方法解析与局限

(一) 传统二手车评估速度慢

在传统的二手车评估过程中, 车况是关键特征, 比如事故车判定项目又是关键特征的子特征, 如没有第三方检测, 评估师首先进行非举升状态下结构部件状态检查, 包括纵梁、水箱框架、

叶子板骨架、ABCD柱、车顶框架、底边梁、减震器座、后围板
的事故损伤, 比如严重锈蚀、褶皱、凹陷、断裂; 再看维修(行业俗称整备)状态, 比如钣金、烧焊、切割、更换等。其次, 举升检查底盘结构部件(结构部件的下侧部分及车身底板框架); 再次, 检查泡水、火烧、结构外观、动力(油动、电动、混动、增程)、配置、安全等多项目, 还需包括漆膜仪、OBD、出险记

录、维保记录的数字化结论，这些项目均是子特征且项目中又包含多个部件状态。这些车况特征都应在二手车评估中参与计算，评估师可在10至60分钟完成检测并评估出价^[2]。

（二）传统二手车评估准确度低

根据车况特征，评估师会采用现行市价法、重置成本法（以下简称，一步乘法）来评估二手车价格。前者方法参考同款车型的在售车辆价格，再考虑车况特征，确定折价率，再一步乘法得到评估价；后者方法参考同款新车在售车辆的成交价，再考虑车况特征，确定折价率，再一步乘法得到评估价。个人评估无风险控制、决策主观单一，消费者对二手车估价或报价不信任，成交效率低^[3]。

（三）传统二手车评估的记忆容量受限，无法并行

传统二手车评估中多维度可达60+，多特征可达100+，车款更是1000+，个人无法记忆全部并比较差异。其次，个人评估师的历史成交经验，未能充分利用历史数据和市场数据，限制了评估结果的准确性，个人记忆容量受限。评估影响买卖双方的信任度，也让车辆交易风险不易管理，规模化受到了限制。面对市场条件和消费者偏好的快速变化，导致评估结果可能与市场实际情况不符，成交率低^[4]。

（四）人工智能在赋能传统二手车评估的能力

传统二手车评估需得到智能二手车评估的赋能，利用机器学习能力来克服速度慢、准确度低、无法并行，提高评估出价的效率、准确性和规模化。人工智能技术在二手车评估领域展现出显著的优势。二手车产品是一车一价但每个商家或者车主的卖车需求是一人一价。一车一价是产品客观事实，一人一价是需求主观意愿。人工智能技术的机器学习能力可精准计算一车一价，同时可根据客户需求精准计算一人一价，兼具客观性、主观性价格；从效率上，人工智能技术可并行计算全量市场的二手车价格；从服务上，人工智能技术可7/24小时履约服务客户；从风险上，人工智能技术可识别真实性，提高效率。最后，从体验上，人工智能技术可满足客户情绪或叠加产品价格需求^[5]。

二、人工智能在二手车价值评估中的应用框架

（一）人工智能技术的单体原理

人工智能技术把特征之前的关系或选择集表达为参数集并通过数学公式单体表达（以下简称表达式），称为单体原理。求解表达式参数集的自动运算称为机器学习；假设特征变量集的提升，导致参数集提升，被称为尺度定律（Scaling laws）。单体原理假设为可复制的分段决策，通过机器学习，达到高标准通用决策准确度，被称为通用人工智能 AGI。机器学习通过建立多样的（比如正向拆解计算、反向推导计算、模拟对抗测试等等）模拟器类计算模型，自动计算关系或选择集的参数集，直至精准匹配，导致分段准确决策。新任务数据输入已知参数集并行计算输出目标结果。所以人工智能是数字化基础设施以后，数智化的新基础设施，人工智能技术通过单体原理、机器学习及尺度定律等实现智能体^[6]。

（二）人工智能技术在二手车特征的应用

人工智能技术应用于二手车估值服务是二手车服务的一个分支，全流程还有检测、客户、车商服务；二手车的人工智能一定会向全流程智能发展。近年来，中国新能源汽车品牌每年发布的车型、车款有上百种，叠加燃油车多达上千种，导致二手车市场的价格变化周期从每月逼近到每日。客观上全球汽车保有量的增加，会导致整体单车利润下滑，最终逼迫二手车估价能力必须达到秒级定价。事实上，传统二手车评估中有竞争力的车商都可在10分钟内定价^[7]。所以，智能二手车评估更需实时迭代。

特征预处理的核心原理是：数据一定多模态；假的数据不可能预测出真数据。特征预处理的假设是：关系和选择表达式无法充分表达；数据量少可以通过工程解决。多维数据需通过机器学习进行特征预处理。解决数据特征预处理，前提是结合二手车行业经验最大化收集特征及数据，并通过常用的分类器比如 boost 类（ada、xg、light、cat 等）做特征重点选择，聚类器（k-means），如果特征维度多，但数据量很小，比如一般交易规模很小的商家或个人则需要做特征过滤来选择特征，比如二分类器（SVM）过滤、包装、树类 GBDT、维度转换 embedding 嵌入和 deep learning^[8] 寻找特征。

特征编码的原则结合行业通用和人类常识，兼具独立、序列、分类、离散、子集等，不同特征采用不同编码，通用及常识即可。这里面需要注意的是，如果数据源有缺失数据或人为篡改的数据，一定会影响智能二手车价格评估。

（三）人工智能技术在二手车价值评估的应用

线性回归 Linear regression，表达式为 $Y = \sum a_n X_n + b$ ，通过输入特征 X_n ，求解 a_n 和 b 。由于是多元一次公式，假设所有特征在一个坐标轴。这种假设可以成立于一些特征确实是相关常数关系，二手车由于非标，特征可多达100个以上，特征很显然非常数关系。另外，bias 这样用于二手车估值预测也过于武断。但线性回归为人工智能技术发现了机会并实现了表达式的改进。人工智能技术通用的检验技术标准指标，例如 R^2 least square 和标准差 standard error，根据实际场景需求，我们还可以加入客户标准指标用来适应客户需求。根据 Noor&Jan 的学术级研究，在多特征被假设的情况下线性回归^[4] 机器学习模拟器对于二手车估价项目的准确率可达98%。

为了追求准确度的提升，Sharma&Mitra 在调研了多家当地商家，根据当地二手车商家的经验，总结了核心4个特征，表达式为 $Y = \sum a_n X_n + \sum c_n X_n + dO + b + e$ 。当地二手车有平均年份高，公里数高，平均换手次数高，价格较低的特征属性，所以在人工筛选掉影响较小的特征，把有限的精准数据运用到线性回归中，得到了可以应用于当地的二手车价格评估。^[9] 他们说明当地二手车很多高于20年车龄，包括未含有政府政策、市场变量，所以整体结果还有比较大的改进空间。

线性回归从表达式即看出基于对特征的核心假设关系及与目标的线性假设关系，二手车行业和很多其他行业一样，特征有未知且多维，所以二手车估价采用线性回归作为唯一计算模型确实不能满足需求，还需要解决多维空间问题。

梯度提升决策树 Gradient Boosting Decision Tree, 表达式为 $f_M(x) = \sum_{m=1}^M T(x; \theta_m)$, x 代表特征, DT 运用了是与否的表达式。通过建树和叶子, 层层决策特征与目标的最佳切点, 直至最佳切点无法切割。数值型项目计算准确度高, 但对于多维度数据量较小的项目比较难达到较高准确性。所以在模型超参数设置中, 可以尝试扩大树的数量以提升准确率^[10]。

神经网络 neural network, 包含多种模型, DNN deep neural

network 的表达式为 $Y = \sum_{n=1}^m a_n x_n + b$, 神经网络分为输入层或 embedding 层、隐藏层和输出层, 其目的是扩张或压缩特征, 把多维神经元通过链接下一层神经元实现线性关系, 由于网络链接密集且可设置多层去挖掘未知特征^[11]。

智能二手车特征及评估弥补了传统二手车评估的局限, 旨在为二手车行业提供一个更快、更准、并行的二手车价值评估工具, 实现交易链条各方的价值, 将促进行业的持续健康发展。

参考文献

- [1] A. Dutulescu, M. Iamandei, L. M. Neagu, S. Ruseti, V. Ghita, M. Dascalu, What is the Price of Your Used Car? Automated Predictions using XGBoost and Neural Networks, IEEE access, 24th International Conference on Control Systems and Computer Science (CSCS), 2023, pp.418-425.
- [2] A. K. MANDAL, M. NADIM, H SAHA, T SULTANA, M. D HOSSAIN, AND E. H, "Feature Subset Selection for High-Dimensional, Low Sampling Size Data Classification Using Ensemble Feature Selection With a Wrapper-Based Search", IEEE access, vol.12, May 2024, pp.62341-62357.
- [3] A. SHRESTHA, A MAHMOOD, Review of Deep Learning Algorithms and Architectures, IEEE access, VOL. 7, 2019, pp.53040-53065.
- [4] K. Noor, S. Jan, Vehicle Price Prediction System using Machine Learning Techniques, International Journal of Computer Applications (0975 - 8887) Volume 167 - No.9, June 2017, pp.27-31.
- [5] J. Sharma, S. K. Mitra, Developing a used car pricing model applying Multivariate Adaptive regression Splines approach, Expert Systems With Applications, 0957-4174/© 2023 Elsevier Ltd. September 2023.
- [6] A. SHRESTHA, A. MAHMOOD, Review of Deep Learning Algorithms and Architectures, IEEE access, VOL.7, May, 2019, PP.53040-53065.
- [7] 陈潇. 人工智能在广播电视节目推荐系统中的应用 [J]. 电视技术, 2023, 47(7):163-165, 182. DOI:10.16280/j.videoe.2023.07.043.
- [8] 赖小馨. 基于人工智能的个性化推荐系统在电子商务中的应用 [J]. 知识经济, 2024, 670(6):37-39.
- [9] 田丽. 人工智能在推荐系统中的应用与分析 [J]. 文渊 (高中版), 2023(9):232-234. DOI:10.12252/j.issn.2096-6288.2023.09.078.
- [10] 曾新士. 基于人工智能的智慧在线服务信息推荐方法及云计算系统: CN202110682159.6 [P]. 2021-09-03.
- [11] 深圳宏途教育网络科技有限公司. 一种基于人工智能的在线教育资源管理系统及其推荐方法: CN202010328707.0 [P]. 2020-07-24.