

汽车燃油消耗预测模型的研究进展

关鹏^{1,2}, 任烁今², 沈义涛^{1,3*}, 赵健福²

1. 哈尔滨工业大学(威海)汽车工程学院, 山东 威海 264209;

2. 中汽研汽车检验中心(天津)有限公司, 天津 300300;

3. 上海交通大学动力机械与工程教育部重点实验室, 上海 200240

摘要:为准确预测车辆在不同工况下的油耗,帮助研发人员深入了解发动机的油耗变化规律并进一步优化发动机性能,全面总结和分析现有的油耗预测模型,将其分为传统油耗预测模型和基于数据驱动的机器学习油耗预测模型两大类,并将基于数据驱动的机器学习油耗预测模型进一步细分为多元回归、浅层机器学习、深度学习和混合式油耗预测模型四类,分析各种模型的预测方法及其变体的应用现状、优势与局限性。明确各类模型的最佳应用场景,指出当前研究中存在的主要问题和挑战。数据驱动的机器学习类预测模型中,多元回归方法对线性相关性强的数据具有很好的表现,且模型透明度高,易于理解;机器学习能够有效应对复杂的非线性关系,特别是深度学习,可以充分挖掘数据中的特征,实现燃油消耗的精准预测,但对数据质量要求较高,模型相对复杂。根据不同模型的特点与应用,对油耗预测未来的发展进行展望。

关键词:油耗预测;数据驱动;机器学习;深度学习;混合油耗模型

中图分类号:TP18

文献标志码:A

文章编号:1673-6397(2025)02-0001-15

引用格式:关鹏,任烁今,沈义涛,等.汽车燃油消耗预测模型的研究进展[J].内燃机与动力装置,2025,42(2):1-15.

GUAN Peng, REN Shuojin, SHEN Yitao, et al. Research progress of vehicle fuel consumption prediction models[J]. Internal Combustion Engine & Powerplant, 2025, 42(2): 1-15.

0 引言

全球变暖已经成为日益严重的社会问题^[1]。为有效解决温室气体的过量排放,中国宣布到 2030 年实现碳达峰目标,到 2060 年完成碳中和这一重要任务。中国提交的《强化应对气候行动——中国国家自主贡献》明确提出:到 2030 年,单位国内生产总值 CO₂ 排放比 2005 年下降 60%~65%的目标,非化石能源占一次能源消费的 20%左右^[2]。我国约有 10%的能源消耗来自运输行业,随着汽车工业和物流行业的蓬勃发展,中国汽车保有量仍然保持增长趋势。2023 年全国机动车保有量高达 4.35 亿辆,汽车保有量达到 3.36 亿辆^[3];相比 2022 年,机动车保有量增加了 1 800 万辆,增长率约为 4.32%,汽车保有量增加了 1 700 万辆,增长率约为 5.33%。截止到 2024 年上半年,我国机动车保有量已经上涨至 4.40 亿辆,其中汽车保有量为 3.45 亿辆^[4]。2019 年至 2023 年全国汽车相关统计数据^[5-8]如表 1 所示。由表 1 可知:汽车保有量增长趋势明显。目前,汽车污染物排放与能源消耗对我国的环保和可持续发展构成严峻挑战,对人类的身心健康构成巨大威胁。减少能源消耗与环境污染成为当下亟待解决的问题^[9]。重型商用车

收稿日期:2025-01-02

基金项目:动力机械与工程教育部重点实验室开放课题(202301)

第一作者简介:关鹏(2000—),男,黑龙江佳木斯人,硕士研究生,主要研究方向为机器学习、深度学习、油耗预测等, E-mail:23S130316@stu.hit.edu.cn。

*通信作者简介:沈义涛(1982—),男,湖北潜江人,工学博士,副教授,硕士生导师,主要研究方向为车辆智能化、电动化与低碳化, E-mail:shenyitao@hitwh.edu.cn。

第四阶段油耗限值将于 2025 年 7 月 1 日实施,越来越多的学者将目光聚焦于燃油消耗预测模型的研究。准确的油耗预测不仅能够反映车辆在不同工况下的实际油耗,还能够帮助研发人员深入探索发动机的油耗规律,并进一步优化发动机性能,还可以为协助国家制定相关政策法规提供参考。此外,油耗预测还能协助驾驶员调整驾驶策略,提高燃油经济性,节约燃料成本并减少环境污染。

目前,油耗预测模型可以大致分为传统油耗预测模型和基于数据驱动的机器学习油耗预测模型两类。对

传统油耗预测模型的研究起步较早,主要基于物理机理和统计回归方法。其核心思想是通过理论推导或经验公式计算油耗,分析车辆行驶过程中的能量消耗路径或拟合历史数据中的油耗规律构建模型,涉及详细的物理与机械原理,如车辆动力学公式、热力学原理等,模型透明度高,较为经典。随着传感器功能与通信技术的提高,汽车行驶数据被记录并保存在平台中,形成大量的车辆行驶数据,通过机器学习方法在历史数据中寻找油耗规律,构建油耗预测模型,成为近年来迅速发展的模型构建新方法。机器学习油耗预测模型的预测精度较高,但该类方法多为黑盒模型,透明度低且不易解释。本文中分别对这两类模型进行详细阐述,特别是基于数据驱动的机器学习油耗预测方法,旨在使研究人员能够清晰了解当前油耗预测模型所采用的方法类型、每种方法的优缺点以及未来的发展方向。

表 1 2019 年至 2023 年全国机动车统计数据^[3,5-8]

年份	机动车保有量/ 亿辆	汽车保有量/ 亿辆	汽车占比/ %
2019 ^①	3.48	2.60	74.7
2020	3.72	2.81	75.5
2021	3.95	3.02	76.5
2022	4.17	3.19	76.5
2023	4.35	3.36	77.2

①截至 2019 年上半年。

1 传统油耗预测模型

根据模型的尺度和粒度,传统油耗预测模型分为宏观、中观和微观三类。由于油耗与排放密切相关,因此经典的传统模型通常同时考虑燃油消耗和污染物排放两方面,又称为能耗排放模型。

1.1 宏观模型

宏观能耗排放模型通常用于描述大范围、长时间尺度的交通系统,如城市或高速公路系统的整体油耗预测。1978 年,美国环保署(environmental protection agency, EPA)开发了一个专门用于评估公路机动车排放因子的模型——移动源排放因子模型(MOBILE 模型)^[10],旨在协助国家、州和地方政府准确估算机动车的排放,并制定有效的空气质量管理计划。MOBILE 模型以大量联邦测试程序(Federal test procedure, FTP)的测试数据和在用车测试数据为基础,根据不同的车型配置,将整体车辆分为 8 类。首先计算各类车辆单独的排放因子,再根据每种类别车辆的占比权重计算总排放;在计算出每一类的基本排放后,还需考虑不同影响因素对其进行修正,最终预估出实际的排放因子。经过 20 余年发展和 10 余次修订,随着模型计算方法的不断改进,MOBILE 模型能够更加全面地考虑影响排放的各种因素。这些因素涵盖了车型、车质量、发动机类型、车辆的维护保养情况等车辆的具体特征,同时,还纳入了行驶里程、环境温度、湿度和所使用的燃油类型等车辆运行条件,确保了估算的准确性和可靠性,能够广泛应用在不同地区。

计算道路运输排放的计算机程序(computer program to calculate emissions from road transport, COPERT)^[11]由欧洲环保局(European environment agency, EEA)开发,可以对不同污染物和温室气体排放进行估计,目前最新版本为 COPERT IV。相比 MOBILE, COPERT 的车型分类更详细,模型的排放是热排放、冷排放和蒸发排放三种排放的总和。该模型主要输入参数为车辆平均速度、燃油相关参数、天气条件等,结合程序中内嵌的各类排放系数、冷起动行驶里程比例等,最终计算出机动车排放^[12]。Lang 等^[13]根据 COPERT 模型计算不同排放标准车辆的排放因子,估算了 1999—2011 年中国道路车辆排放量,分析 CO、非甲烷挥发性有机物(non-methane volatile organic compound, NMVOC)、NO_x 等排放污染物在这些年中不同省份的变化趋势,研究结果表明:CO、NO_x 排放总体呈上升趋势, NMVOC 排在 2002 年之前呈上

升趋势,2003年后开始逐渐下降;通过与先前对当地排放测量结果对比发现,COPERT IV模型可用于估算中国的车辆排放。

1988年,美国加州空气资源局(California air resources board, CARB)发布了车辆排放模型 EMFAC (emissions factor)^[14],并多次对其进行改善和更新,广泛用于预测区域污染物排放和燃料消耗,其工作原理与 MOBILE 模型基本相同,均是对基本排放因子进行影响因素的修正。从2000年5月 CARB 首次推出 EMFAC2000 模型以来,该模型经历了多次迭代,最新的 EMFAC2017 模型的功能和精确度显著提高,该版本提供缺省模块、定制模块和项目模块3个污染物排放计算模块。缺省模块最简单,简单输入机动车保有量即可计算宏观排放水平,算法内的修正因子均为默认;定制模块和项目模块需要输入更加复杂的参数,根据不同要求计算排放。除 CO、NO_x、颗粒物(particulate matter, PM)等传统的机动车污染物计算,EMFAC2017 还引入了对温室气体排放的评估,包括 CO₂、N₂O 和 CH₄。

宏观能耗排放模型侧重于排放估算,但通过各种方法以及模型版本迭代,MOBILE、COPERT 和 EMFAC 模型也广泛应用于燃油消耗研究中^[15]。

1.2 中观模型

中观油耗预测模型通常用于描述中等范围及时间尺度的交通系统,如交叉路口、某个街区等的燃油消耗。机动车排放模拟器(motor vehicle emissions simulator, MOVES)^[16]模型是 EPA 开发的新一代机动车能耗排放模型,利用瞬时速度和加速度计算车辆比功率(vehicle specific power, VSP)。以 kW/t 为单位的车辆比功率 P_{vs} 的数值

$$\{P_{vs}\} = (A\{v\} + B\{v^2\} + C\{v^3\} + \{m\}\{v\}\{a\})/\{m\}, \quad (1)$$

式中: $\{v\}$ 为以 m/s 为单位的速度 v 的数值, $\{a\}$ 为以 m/s² 为单位的加速度 a 的数值, $\{m\}$ 为以 t 为单位的的质量 m 的数值, A 为滚动阻力系数, B 为旋转阻力系数, C 为空气阻力系数。

根据 VSP, MOVES 模型将车辆的运行状态(包括起动、空转和运行模式)划分为 23 个运行单元(Bin),与不同运行单元对应的是污染物排放速率,然后根据当前车速自动匹配排放单元,计算基础排放因子,并通过燃油、里程、气象等影响因素对其进行修正,得到综合排放因子。相比 MOBILE 模型,MOVES 在中观层面具有更好的表现。2003 年,机动车排放模型 IVE 被推出,该模型是由美国加州大学河滨分校牵头,与全球可持续系统研究组织和国际可持续发展研究中心共同研发^[17]。该模型同样引入了 VSP 作为影响排放的特征参数,同时引入发动机特征强度(engine stress, ES) 共同表征车辆瞬时工作状态与排放的关系。IVE 模型依据 VSP 和 ES 划分为 60 个 Bin 区间,每个区间内嵌基本排放因子,最后采用宏观排放模型思想,根据各类修正系数对基本排放因子进行修正,最终得到实际排放因子。MOVES 模型的开发参考了 IVE 排放模型,不仅构建了瞬时工作状态与能耗排放之间的紧密联系,还为评估不同行驶工况下的能耗和排放提供了更为精准的基础,提高了预测的准确性和可靠性。

1.3 微观模型

微观层次的油耗预测模型适用于小范围、短时间尺度的交通系统,更关注单个车辆的详细行为,如速度、加速度等对瞬时油耗的影响。综合模态排放模型(comprehensive modal emission model, CMEM)^[18]是由加州大学河滨分校工程学院环境研究与技术中心联合密歇根大学、劳伦斯伯克利国家实验室的研究团队共同开发的一个排放模型,该模型从微观角度对汽车在不同行驶模式下的逐秒燃油消耗进行估算,模型将尾气排放过程和运行工况进行分块处理,通过向对应的模块输入车辆行驶参数,即可得到车辆燃油消耗和排放,但所需输入的参数较复杂,包括车辆质量、发动机燃烧率、发动机排放指数等。Ahn 等^[19]在 2002 年提出了 VT-Micro 模型,用于预测轻型汽车和轻型卡车在稳定热状态下的油耗和排放率,利用分类回归树方法将测试车辆分为不同类别,通过瞬时变量和对数变换技术,以速度和加速度不同幂次的乘积作为自变量,建立了燃油消耗率函数关系式。

传统油耗模型通常基于车辆物理学原理和统计经验规律,由于汽车内部构造及核心部件的变化都影响发动机的燃油消耗率,因此传统油耗预测模型需要对汽车理论知识有深刻的理解。构建合理的油耗预测模型必须考虑车辆类型、车辆动力学特性、发动机性能、驾驶行为、道路条件、交通状况等多种因素,因

此,这类模型构建相对复杂,但明确的物理意义可以使模型更加清晰、透明,易于理解。

2 基于数据驱动的机器学习油耗预测模型

随着智能网联技术的普及,汽车行业产生了大量实时运行数据,伴随着大数据分析技术不断发展,基于数据驱动的油耗预测方法受到越来越多学者的青睐。该方法主要依靠车载诊断系统(on board diagnostics, OBD)、便携式排放测量系统(portable emissions measurement systems, PEMS)、全球定位系统(global positioning system, GPS)或车辆传感器上传至平台的实时数据,因此数据量巨大且包含十分丰富的物理信息,再通过机器学习方法提取数据中的特征,构建输入特征与油耗的映射函数。OBD 系统是嵌入在车辆内部的实时监控系統,已经经历了三次迭代,目前又称为远程 OBD 系统或 OBD III。OBD III 将前两代的诊断技术与远程传输技术相结合,具有故障检测、排放控制、故障报警、数据储存等功能,通过车辆内置传感器获取发动机运行参数(如转速、负荷)、排放相关参数(氧传感器电压、催化转换器效率)、其他系统参数(如 PM 过滤器状态)等,最后将车辆运行数据、当前排放状况、故障检测信息实时发送至监管部门的数据中心,从而对在用车进行远程监测^[15]。远程 OBD 系统上传至监测平台的数据类型如表 2 所示,采集频率为 1 Hz。

表 2 远程 OBD 系统数据采集类型

序号	数据项	序号	数据项	序号	数据项	序号	数据项
1	VIN	8	发动机燃料流量	15	DPF 压差	22	发动机转矩模式
2	采集时间	9	SCR 上游 NO _x 传输器输出结果	16	定位状态	23	油门踏板
3	车速	10	SCR 下游 NO _x 传输器输出结果	17	经度	24	累计油耗
4	大气压	11	尿素液位	18	纬度	25	尿素箱温度
5	发动机净输出转矩	12	进气量	19	发动机冷却液温度	26	实际尿素喷射量
6	摩擦转矩	13	SCR 入口温度	20	油箱液位	27	累计尿素消耗
7	发动机转速	14	SCR 出口温度	21	累计里程	28	DPF 排气温度

PEMS 是实际道路行驶条件下实时测量车辆排放的移动设备,具有污染物测量、驾驶工况记录、法规符合性验证等功能。在对车辆进行实际运行数据检测时,需要专门安装复杂的采集设备,如气体分析仪、排气流量计、数据采集单元、供电单元等,因此成本较高,试验过程复杂,但得到的车辆逐秒运行数据更加准确。在实际行驶过程中,行驶环境、驾驶行为、车辆自身特性、交通及道路状况等油耗影响因素均能够在车辆实时行驶数据中体现^[20],如速度的变化可以体现驾驶员的驾驶习惯,发动机的转速和转矩可以大致反映出车辆的行驶环境。汽车在市区行驶时,由于错综复杂的交通条件,车速一般在 30~60 km/h 的低速区间内波动,不断地起停、加速以及换挡操作导致发动机转速及其他参数的频繁突变、单位距离内燃油消耗增加。汽车在市郊行驶时,随着交通复杂程度的降低,车速一般保持在 50~80 km/h,发动机转速也会维持在一个相对经济的区间内,相比市区,燃油消耗与排放也有所改善。汽车在高速行驶时,车速较高且稳定维持在 80~120 km/h,加速、减速次数较少,这时车辆通常处于高档位状态,发动机转速更加平稳,各项参数也趋于稳定,因为稳定的车速与较高的发动机效率,燃油消耗率相对较低。

数据驱动方法的重点在于对历史数据的挖掘,通过数据预处理(如数据清洗、数据降维)和有关算法(如聚类、分类、关联性分析)从大规模复杂的车辆运行数据中提取与油耗相关的特征信息,建立数据特征与油耗之间的非线性关系,实现汽车在不同状态下的油耗映射,从而对油耗进行准确预测^[21]。与传统方法相比,数据驱动方法不需要详细考虑汽车中不同部件对油耗的影响以及其中蕴含的物理知识,更加简单易行。机器学习是人工智能(artificial intelligence, AI)的一个分支,研究计算机算法和统计模型,能够在不进行显式编程的情况下使计算机自动从数据中学习和改进。Janiesch 等^[22]将机器学习划分为浅

层机器学习和深度学习,划分结构^[11]如图1所示。

深度神经网络一般由多个隐藏层构成,并组成深层嵌套的网络架构。与浅层机器学习相比,深度神经网络通常包含更高级的神经元,可以完成数据局部特征提取、时序特征提取等高级操作,如数据随时间的变化趋势或周期性变化等。决策树、支持向量机等机器学习方法没有这种深层次的网络结构,被划分为浅层机器学习。本文中依据该划分标准对使用机器学习方法的油耗预测模型进行阐述。由于构建多元回归油耗预测模型的核心也是利用历史数据来训练模型,因此归入基于数据驱动油耗模型的类别。

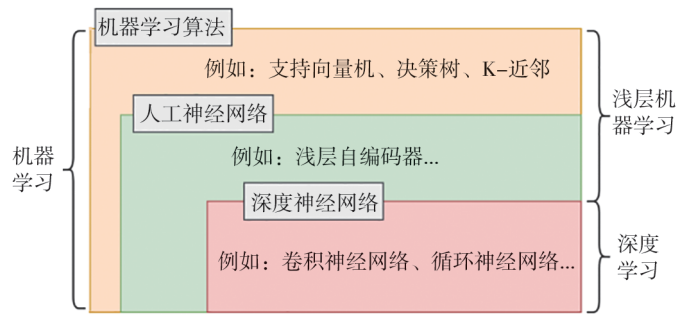


图1 机器学习结构划分图

由于传感器信号的不稳定性,获取的数据质量往往参差不齐,在进行模型训练之前需对数据进行预处理操作,如剔除极端数据和重复数据、插值补充空缺数据等。为提高预测模型的鲁棒性和收敛速度,一般还会对数据进行特征选择及特征降维。特征降维是在保留特征主要信息的同时,通过某种算法降低输入参数维度,实现模型简化。常用的特征选择方法包括皮尔逊相关系数、斯皮尔曼相关系数、灰色关联度等,这三种方法均是常用的相关性分析方法,用于衡量变量之间的相关程度,再根据相关程度排序,选择相关性更大的特征作为模型输入。皮尔逊相关系数适合线性关系检测,斯皮尔曼相关系数适合单调关系的检测,灰色关联度适合多变量系统的相似性评估。还有学者通过各特征参数在神经网络中的贡献度确定模型的最终输入。余舒^[23]通过反馈神经网络(backpropagation neural network, BP)分析拓展参数对预测结果的影响权重,选择权重较大的参数作为最终预测模型的输入特征。主成分分析法(principal component analysis, PCA)是一种广泛采用的特征降维方法,这种方法利用线性变换技术,将原始数据集中的特征映射到一个更低维度的空间,同时尽可能保留数据的主要变异信息,然后通过最大化方差增加样本之间的可分离性,尽可能多的保留与目标变量相关的原始数据信息。除此之外,还有学者利用模态分解方法将原始数据进行滤波处理^[24],去除噪声对试验结果的影响。

为了验证基于数据驱动模型的预测性能,一般通过以下指标进行评价:决定系数 R^2 、平均绝对误差 E_{MA} 、相对误差 E_R 、均方误差 E_{MS} 、均方根误差 E_{RMS} 、平均绝对百分比误差 E_{MAP} 等,具体计算式及其评价标准如表3所示。

表3 评价指标计算式及其评价标准

计算式	评价标准	计算式	评价标准
$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$	越接近1越佳	$E_{MS} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$	越小越佳
$E_{MA} = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $	越小越佳	$E_{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$	越小越佳
$E_{Ri} = y_i - \hat{y}_i / y_i $	越小越佳	$E_{MAP} = \frac{100\%}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right $	越小越佳

注: $\hat{y}_i$ 为第 i 次的预测结果, $i = 1, 2, \dots, n$, n 为样本数量; y_i 为第 i 次的真实结果; E_{Ri} 为第 i 次的相对误差。

不同评价指标适用于不同场景,同时也能够提供模型性能的不同视角。 R^2 易于解释,值越接近1模型效果越好,说明模型越能捕捉到数据中蕴含的趋势; E_{MS} 、 E_{RMS} 对异常值敏感,能够惩罚模型中较大的错误;若研究人员更加关心模型的相对误差,可以选择 E_{MAP} 或 E_R 作为模型评价指标。

2.1 多元回归油耗预测模型

多元回归是常用的预测方法,用于预测一个连续的因变量与多个自变量之间的关系。车辆实际行驶过程中,燃油消耗往往受车速、加速度等多个因素的影响,因此很多学者通过多元回归方法构建合适线性或非线性函数,评估燃油消耗与其影响因素之间的关系,以此预测油耗。

阮连杰^[25]基于行驶车辆的油耗数据集,以车辆行驶速度、加速度和道路坡度等组合得到的特定特征,构建多元回归瞬时油耗预测模型。该研究首先分析了油耗与车辆功率之间存在十分明显的线性关系,通过车辆动力学公式中的功率平衡搭建了 3 种油耗模型框架、6 种基本瞬态油耗模型,模型经过 Elastic Net 回归、遗传算法和贝叶斯准则的简化与特征筛选,有效提高了其预测精度。通过试验分析,该瞬时油耗预测模型的预测平均绝对误差为 0.117 mL/s,相比传统的 VT-Micro 模型,平均误差降低了 12.1%。

郭娟^[26]以多元统计分析为基础,建立了以功率、车质量、传动比、功率和车质量的乘积的交互项为自变量,单位体积的汽油行驶里程为因变量的多元线性回归模型,以此反映车辆的油耗水平。该研究首先分析了各变量之间的相关性,然后通过 R 软件中的 $\text{lm}()$ 函数拟合多元线性回归模型,再根据相关性程度,去掉不显著的特征并增加交互项,构建出最终回归模型。试验结果表明,决定系数 R^2 可达 0.848 8。为简化数据收集和模型识别过程,Madhusudhanan 等^[27]提出了一种使用预测误差识别油耗模型参数的建模方法,以发动机转速和转矩作为模型输入,使用参数化的传递函数构建预测模型以及预测误差模型,采用最小化预测误差平方和的方法估算模型参数。在发动机模型中,训练集和测试集的拟合度分别为 91.0% 和 91.2%,与传统方法相比,在保证拟合度的情况下,识别参数所用数据集缩减了 99%,有效减少了测试时间和成本。

功率可以较准确地反映发动机的燃油消耗状态,高功率输出需要消耗更多的燃油,Ren 等^[1]提出一种基于实际道路数据的分箱重构油耗预测模型,该研究以大量试验数据为基础,依据发动机转速和转矩划分 Bin 区间,同时每个区间中嵌入油耗,将油耗压缩为 2 维的转速-转矩 map。通过分析不同区间跨度对油耗预测的准确性,证明了通过发动机转速与转矩可以实现对发动机油耗的准确映射。

Jiménez^[28]提出了描述车辆瞬时功率需求的指标 VSP,将车辆的动能、势能、滚动阻力、空气阻力等所有负荷归一化为单位质量的功率(kW/t),引入 VSP 可以更加直观方便地描述车辆在某一时刻的功率总需求。许多学者将 VSP 作为油耗模型的重要影响因素:1) 范童谣^[29]利用 VSP 和车辆速度 v 建立了油耗预测模型,精确评估车辆在不同行驶工况下的能源消耗,采用区间划分的概念,首先根据 VSP 和 v 划分 Bin 区间,并依据 v 将所有 Bin 区间分为怠速、中低速和高速三部分,最后拟合出每一部分的油耗、功率、 v 曲线,构建油耗回归模型;在试验过程中,由于车辆在中低速时的行驶工况更复杂,油耗预测效果一般,但在怠速和高速工况下,油耗预测误差均在 5% 以内。2) 陈旗^[30]通过 VSP 描述交通运行分布状态,在不同的速度区间,构建了因变量为 VSP 的分段线性回归油耗预测模型,同时加入机动车排量参数对油耗模型进行修正,提出基于排量区间的修正方法,将排量分为 4 个区间,并根据 VSP 是否大于 0 进行不同方法的油耗计算,若 VSP 大于 0,依据排放和速度区间寻找对应油耗模型 $y=kx+b$ 的拟合参数,若 VSP 小于 0,直接寻找对应的燃油消耗率;对应参数和燃油消耗率由大量已有车辆运行数据总结归纳得到。

采用多元回归方法构建油耗模型时,需要寻找车速、加速度、载重、坡度、VSP 等能够描述燃油消耗的影响因素。通过多个重要影响因素解释油耗变化,该方法具有较强的可解释性,便于理解和应用。然而车辆燃油消耗是极其复杂且多变的过程,具有较强的非线性特性。一般的线性或非线性方程往往难以准确描述油耗变化。此外,当自变量之间存在高度相关性时,会导致回归系数的估计不稳定,容易出现多重共线性现象,从而增加模型的不确定性和解释难度。

2.2 浅层机器学习油耗预测模型

机器学习是一种利用算法和统计模型从观测数据中学习一般规则预测未知变量的方法,具有很强的处理非线性关系的能力,越来越多地被应用于油耗模型的构建中。主要的浅层机器学习方法包括支持向量机(support vector machine, SVM)、随机森林(random forest, RF)、梯度提升决策树(gradient boosting

decision, GBDT)等。

SVM 主要用于解决二分类问题,依靠核函数实现从低维数据到高维数据的映射,其原理如图2所示。

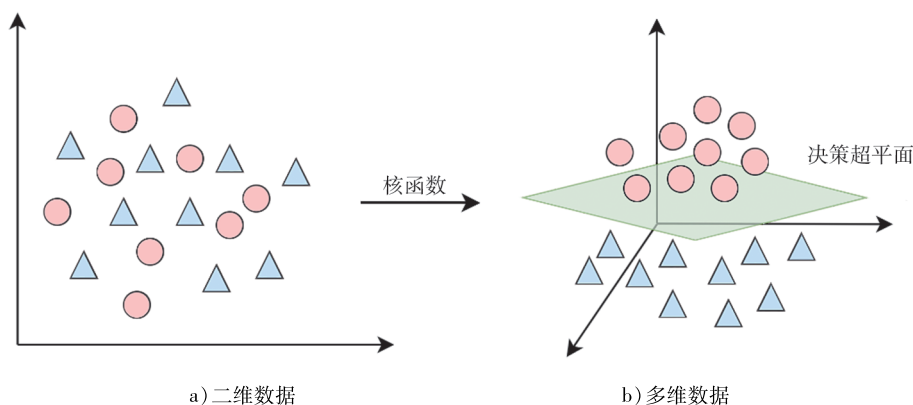


图2 SVM 原理图

支持向量回归(support vector regression, SVR)是 SVM 的拓展,能够用于解决回归问题,其核心思想是通过找到一个最优的超平面,以达到最佳分类或回归预测。由于 SVR 善于处理小样本和复杂非线性问题,因此在油耗预测领域受到较多关注。刘昌海^[31]分别构建了 BP 神经网络、长短期记忆网络(long short term memory, LSTM)和 SVR 三种算法,分别应用于市区、市郊和高速路段的油耗预测中,研究表明:BP 神经网络的泛化能力最差, LSTM 在三种路段的预测差异较小,但整体准确率低于 BP 神经网络;在准确度和泛化能力上,SVR 的表现最佳,优于其他两种算法。此外,该研究在初始化、收敛因子、等级制度和淘汰机制四方面对灰狼算法(grey wolf optimizer, GWO)进行了改进,利用改进后的灰狼算法(improved grey wolf optimizer, IGWO)对 SVR 进行参数优化。改进后的 SVR 算法在市郊和高速路段的油耗预测中, R^2 均超过 0.95。Suykens 等^[32]对 SVM 算法进行改进,提出了最小二乘支持向量机(least squares support vector machine, LSSVM)。在改进传统支持向量机的过程中,损失函数采用了最小二乘线性系统。这一变化使得原本的不等式约束能够被转换为等式约束,不仅简化了优化问题的结构,还显著提高了计算效率。周祎^[33]以高斯核函数为基础构建 LSSVM 油耗预测模型,预测拥堵、市郊、高速三种工况下的油耗,并利用自适应粒子群算法进行参数寻优。在与 SVM、递归神经网络(recurrent neural network, RNN)算法的对比中, LSSVM 在不同工况下的 R^2 均在 0.96 以上,远超其他两种算法,可以有效实现不同行驶工况下的油耗预测。应用 SVM 算法时,合适的核函数可以有效提高模型的预测精度,但当面对新数据集时,需要重新选择核函数,计算量增大。

随机森林是 Breiman^[34]提出的由多个决策树组成的集成模型,其核心思想是通过构建多棵决策树,并通过投票或平均每棵树的预测结果,得到最终预测结果, RF 原理如图3所示。每个决策树都在训练集的一个随机子集上训练,而且在每个节点的分裂过程中,只考虑一部分随机选择的特征。算法根据当前数据集遍历所有候选特征和可能分裂点,选择使分裂后子节点的误差下降最大的特征和分裂点,将数据划分为左右子节点,重复该分裂过程直到满足停止条件,最终叶节点的预测值为该节点所有样本标签的均值,该方法在多维度数据分析和预测中应用广泛,油耗预测中也有很好的表现。

赵明^[35]通过试验证明,车辆的载质量和季节对车辆实际油耗的影响较大,因此以载质量、月份和平均车速

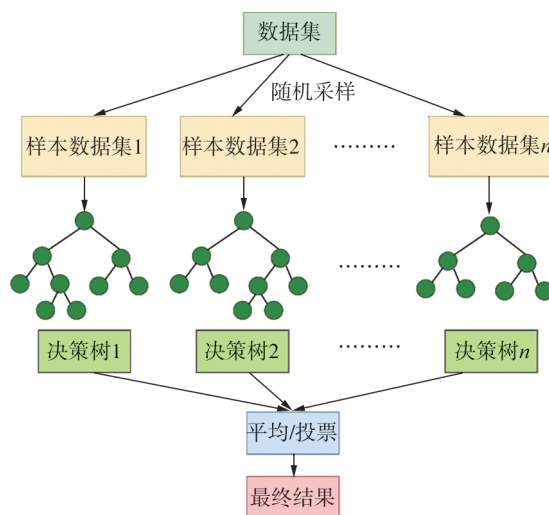


图3 RF 原理图

为输入构建随机森林模型,预测车辆的平均百公里油耗,验证集为三辆不同质量冷藏车在12个月的油耗结果,预测结果的 R^2 为0.831。

Massoud 等^[36]利用 RF 分析了驾驶行为与油耗的关系,以代表车速和发动机转速的特征参数作为输入, R^2 和 E_{MS} 分别为0.896和1.506。阮连杰^[25]构建 RF 油耗模型并经过遗传算法优化,以速度、加速度、道路因素作为输入特征预测油耗,预测平均绝对误差为0.067 mL/s,优于传统油耗模型。

在构建每棵决策树时,RF 方法从所有可用特征中随机选取一个子集进行节点分裂,这种随机选择机制降低了特征间的相关性,使模型在处理高维数据时表现更为出色。通过集成多个决策树,RF 不仅提高了模型的泛化能力和稳定性,还增强了对复杂数据的适应性。然而,决策树数量的增加也意味着更高的计算资源需求。

梯度提升决策树与随机森林相同,均是一种高效的集成学习技术,但梯度提升决策树通过连续串行训练新的决策树,迭代减小模型预测误差,每棵树依赖前序树的残差,不可并行。梯度提升回归器 (gradient boosting regressor, GBR) 是 GBDT 算法的直接实现, Nagar 等^[37]构建并对比了18种适用于功率为5.88~35.30 kW 拖拉机的集成学习燃油预测模型,这18种集成学习算法均使用决策树作为基础学习器,但在构建、组合决策树以及优化策略上有所不同。该研究选择最大取力器(power take-off, PTO)功率、发动机转速以及发动机转速降幅作为模型输入,结果表明,GBR 在测试和训练阶段的性能较好,训练集与测试集的 R^2 分别为0.999和0.914;对功率为31.63 kW 拖拉机应用该模型,田间试验结果表明油耗预测精度 R^2 为0.897, E_{MAE} 为11.49%。Bousonville 等^[38]基于500多辆16种不同型号的卡车远程数据,以卡车类型、平均速度、速度方差、道路类型和总质量为输入特征,分别构建了GBR、人工神经网络(artificial neural network, ANN)和K近邻(K-nearest neighbors, KNN)三种机器学习方法对油耗进行预测,结果表明GBR的预测精度最高。

极端梯度增强(extreme gradient boosting, XGBoost)在GBDT的基础上引入了多种优化技术,包括块结构并行、将数据按特征列分块存储、支持并行分裂点查找。近似分裂算法将连续特征分桶后计算候选分裂点,可以减少计算量,是基于GBDT的改进版本。XGBoost是一种串行集成学习模型,因此计算效率一般低于随机森林,但其运行是不断迭代优化的过程,在模型准确性上普遍优于随机森林。林若奇^[39]分别构建RF和XGBoost模型,对猛踩油门型、高速行驶型以及频繁变速型三种驾驶行为进行百公里油耗预测,试验结果表明XGBoost在三个数据集上的预测结果均优于RF,在猛踩油门数据集中, R^2 达到了0.81;同时该研究使用Shapley value方法分析特征变量对油耗预测的影响机制,提出了节能驾驶策略建议。Su 等^[40]使用XGBoost预测了能够运输各种滚装货物的纯汽车和卡车运输船(pure car and truck carrier, PCTC)的燃油成本,选择PCTC大多使用的航线,以船型、距离、燃油价格、航速、燃油成本和港口日期为特征输入,试验结果表明,XGBoost在RF、SVM和BP算法中表现最优, R^2 为0.968。LightGBM算法由微软开发,是GBDT和XGBoost的升级版,它通过引入直方图算法和梯度单边采样等技术,显著提高了训练效率和模型性能。Zhu 等^[41]将LightGBM算法应用于分段油耗预测,并利用Optuna对模型超参数进行优化,以提高模型的预测精度。该研究提取运动学片段的驾驶行为特征作为模型输入参数,预测片段的百公里油耗,试验结果表明LightGBM算法优于XGBoost、RF、SVR等现有算法, E_{MA} 、 E_{MS} 、 R^2 分别为2.036、13.042、0.873。GBDT方法能够很好地捕捉车辆在实际行驶过程中各影响因素,如速度、加速度变化导致的油耗变化过程,但由于其串行集成学习的特性,往往需要较多的计算资源。

2.3 深度学习油耗预测模型

深度学习的核心在于神经网络,特别是深度神经网络,它通常由多个隐藏层组成,隐藏层介于输入与输出层之间,负责从输入数据中提取特征信息,每层包含若干个神经元。这些神经元接收输入信号,执行加权求和与运算处理,并通过激活函数输出处理结果。激活函数决定神经元的信息是否传递到下一个神经元,常用的激活函数包括Sigmoid函数(常用于二分类)、ReLU函数和Tanh函数。模型利用反向传播算法计算损失函数相对参数的梯度,并依据该梯度调整权重,通过不断迭代训练优化性能。深度学习具有极强的非线性拟合能力,相比浅层神经网络,高级神经元能够自动提取和发现原始数据中的特征^[42],

无需进行核函数选择或其他手动特征工程。由于深度学习方法具有强大的非线性拟合能力和出色的特征提取能力,许多学者将该方法应用于机动车油耗预测领域,并取得了显著成果。

ANN 是一种受生物神经元启发的机器学习技术^[43],可用于创建准确的时间效率预测模型,其基本结构如图 4 所示。Sun 等^[44]开发了一种基于 ANN 的混合动力柴油公交车油耗预测模型,该模型能够根据实际运营数据在微观(采集频率 1 Hz)和中观(时间为 5~60 min)两个层面上进行精确预测。微观模型使用车辆速度、加速度、道路坡度、环境温度和频率为 1 Hz 的 VSP 作为输入变量,为了构建中观模型,将收集到的 1 Hz 数据处理成 11 个自变量和 1 个预测变量,即片段平均速度、时间跨度、累计距离和油耗率(每百公里消耗的燃油体积,以 L 计)等。试验结果表明,ANN 模型在预测混合动力公交车油耗方面优于传统线性回归模型,其中微观模型的累计绝对误差仅为 1%~2%,中观模型的绝对预测误差为 5%~9%。Katreddi 等^[45]基于两辆相同品牌及相同型号柴油机的 PEMS 道路测试数据,构建了一个 ANN 燃油消耗模型。考虑到参数获取的难易程度,选择了发动机负载、发动机转速和车速这三个易于获取的数据作为模型输入;使用一辆车的数据进行模型的训练和测试,并将训练好的模型用于预测另一辆车的油耗,以验证模型的泛化能力;结果表明:训练集车辆的 R^2 为 0.780 6,验证集车辆的 R^2 为 0.776 2。由于该研究选择的输入参数较少,因此预测效果一般,不能很好地捕捉燃油变化,但更多的参数需要车辆硬件设备的加持。因此需着重挖掘参数中的特征,以强化模型的捕捉能力。反向传播神经网络作为人工神经网络的一种,通过前向传播和反向传播算法进行训练,能够学习和表示复杂的非线性关系。郭宇^[46]为了优化 BP 神经网络的权值和阈值,提出了一种遗传退火算法改进的 BP 神经网络油耗预测模型,结合遗传算法全局搜索和退火算法避免局部最优的优势,提高模型训练效果:对负荷率、转速、速度、转矩、换挡次数、刹车次数等 6 个与油耗强相关的因素进行 PCA 降维处理,将重构后的特征作为模型输入;该模型在春季、夏季和秋季三个数据集上进行验证,结果表明,模型在全时域数据集集中的平均准确率为 96.51%,平均 E_{RMS} 和平均 E_{MA} 分别为 13.69 和 8.80。

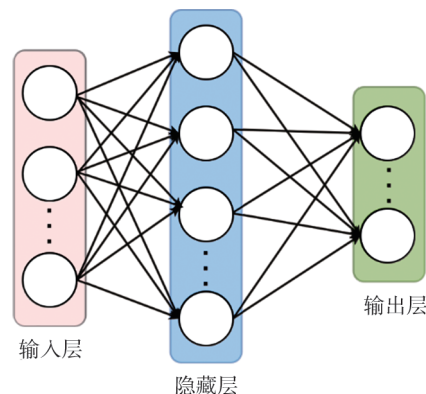


图 4 ANN 基本结构图

RNN 是一种具有时延特性的神经网络,与前馈神经网络不同,其内部包含一个循环结构,使其能够递归地传播输入并保留先前有用的数据处理状态^[47]。Xu 等^[48]利用车联网动态燃油消耗-速度信息,开发了广义回归神经网络(generalized regression neural network, GRNN),以更准确地估计卡车燃油消耗。该研究使用来自某汽车集团的车辆行驶真实数据,涵盖了不同驾驶行为和路况,结果表明,GRNN 模型在新路线上的燃油消耗预测性能优于现有的车辆特定功率模型、VT-Micro 微观模型和综合模式排放模型 CMEM。由于特有的循环结构,RNN 十分擅长处理时序数据,但随着迭代进行,容易面临梯度消失和梯度爆炸的挑战。

LSTM 通过独特的门控单元设计,成功克服了传统 RNN 在训练过程中遇到的梯度消失与梯度爆炸难题^[49],其包含长时记忆和短时记忆,能够确保长时间序列依赖的学习能力。LSTM 的单元结构如图 5 所示。Wang 等^[50]基于 LSTM 构建了燃油消耗预测模型,并通过实地试验验证了其预测性能,该研究对输入组合以及训练集大小进行分析,在保证预测精度的前提下减少训练时间,提高模型效率,选择速度、加速度、GPS 信号作为输入,并在信号灯交叉口、校园环境、城市道路和高速公路上的预测性能进行了验证,相对误差 E_{R} 分别为 0.049、0.030、

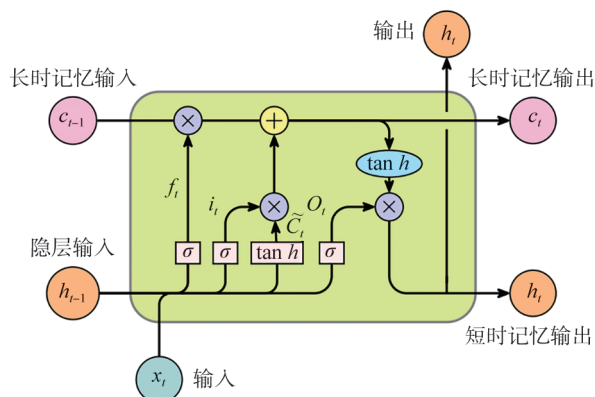


图 5 LSTM 单元结构图

0.077、0.097。张立成^[51]构建了 LSTM 油耗预测模型,并将驾驶行为的加速度一阶导数 Jerk 融入到油耗模型中,加强模型与驾驶行为的关联。选择速度、加速度、Jerk 作为模型输入参数,分别在校园、城市、高速三种工况下对模型进行预测。与 RNN、GRNN 等其他深度学习模型相对,LSTM 的效果最佳,三种工况下的 R^2 分别为 0.998、0.998、0.996。当数据量过大时,LSMT 网络需要花费较长的时间才能收敛。王一婷^[52]在保持标准 LSTM 门结构不变的前提下,提出了 Shared-LSTM 模型。该改进通过共享门结构中的权重,将输入门、输出门和遗忘门整合为一个统一的共享门。在不破坏 LSTM 原有功能的基础上,优化了网络结果,减少模型的收敛时间。相比普通 LSTM、GRU、BP 神经网络,该模型预测效率提高了 3% 以上,在模型预测精度方面, E_{MA} 、 E_{MS} 、 E_{RMS} 均降低了 6% 以上。双向长短期记忆网络(bidirectional long short-term memory, BiLSTM)是 LSTM 架构的改进版本,通过沿时间序列的正向和反向分别训练 LSTM 模型,能够整合来自过去和未来的信息。这种方法使得 BiLSTM 在捕捉长时间依赖关系方面更加有效,因为它不仅考虑了当前时刻之前的序列信息,还结合了之后的信息,为每个时间点提供了更全面的理解。吴丽宁^[53]构建了 BiLSTM 油耗预测模型并分别应用在激进、谨慎、标准三种驾驶行为数据集中,与 RNN、GRU、普通 LSTM 循环神经网络相比,该模型的 E_{MA} 与 E_{RMS} 评价指标均最小。LSTM 在已有的数据集中的表现良好,但当车辆遇到新的环境时或应用到不同类型车辆上时,需要重新训练模型,同时该模型对数据质量要求较高,过少的数据影响模型预测精度。

2.4 混合式油耗预测模型

面对实际道路行驶中的复杂情况,浅层的机器学习方法和单一的深度学习模型仍然存在一定的局限性。为了从数据中有效提取特征并显著提升模型性能,可以融合不同的机器学习模型,集合每个模型的优势,深入挖掘不同的特征信息,提取更准确且有效的特征,增强模型的整体性能与鲁棒性。

师国东等^[54]构建了 XGBoost-MSIWOA-LSTM 模型预测瞬时油耗,根据 XGBoost 算法原理、方法补充提取有效特征,并提出一种多策略改进的鲸鱼优化算法(multi-strategy improved whale optimization algorithm, MSIWOA)对 LSTM 参数进行寻优。通过改进 MSIWOA,可有效提升该算法的收敛速度和搜索能力,将优化后的超参数反馈给 LSTM,试验结果表明,该融合算法可以实现瞬时油耗的准确预测, R^2 达到 0.997 5。

张隅希^[55]提出一种 MIPCA-MLP 油耗预测模型,通过改进后的 PCA 算法对输入参数进行特征提取,消除了原始特征冗余关系,也注意特征参数与油耗之间的相关性,既简化了模型又提高了模型的预测精度,研究表明:MIPCA-MLP 模型在处理这种复杂非线性回归问题时,预测结果较准确, R^2 达到 0.92。

卷积神经网络(convolutional neural network, CNN)是专为处理具有网格状拓扑结构的数据(如图像、视频和时间序列)而设计的一种深度学习架构,通过应用卷积运算,CNN 能够高效地从这些数据中提取出有意义的特征,使得 CNN 在诸如图像识别、物体检测以及自然语言处理等领域表现出色。研究人员将 CNN 与其他模型融合使用,利用 CNN 提取数据中的局部特征,丰富特征表示,强化融合模型的特征提取能力。杨冉^[56]构建了 CNN-LSTM 融合油耗预测模型,同时也将 Jerk 融入到油耗模型中,以速度、加速度、Jerk 作为模型特征输入,研究表明,加入 Jerk 后可以有效提高模型精度,在车辆速度为 30~80 km/h 时, E_{RMS} 下降了 35.39%, R^2 增加了 10.55%。Vaswani 等^[57]在 2017 年首次提出了一种基于自注意力机制的 Transformer 深度学习模型,通过并行处理输入序列和捕捉长距离的依赖关系,有效克服了在处理长序列数据时常见的挑战。Zhou 等^[58]在 2021 年提出 Informer 深度学习模型,通过引入概率稀疏自注意力机制和其他优化技术,有效解决了传统 Transformer 模型在处理长序列时的计算复杂度和内存占用问题。Song 等^[59]将 Informer 应用到油耗预测领域,提出了 VMD-Informer 油耗模型用于预测液压挖掘机油耗,通过变分模态分解(variational mode decomposition, VMD)方法分解原始信号,有效去除原始数据中的高频噪声,得到固有模态函数序列作为 Informer 油耗模型输入,对油耗模型进行迭代训练,预测挖掘机油耗变化趋势,试验结果表明,在 GRU、LSTM、CNN-LSTM、LSTM-Attention、Informer 深度学习模型中,VMD-Informer 表现最好,不同挖掘机工作状态下 R^2 都在 0.97 以上。

虽然融合模型可以提取数据中更丰富、更深层的特征,但融合更多模型代表着模型整体复杂度增加,模型运算成本也随之增加。当多种深度学习方法组合使用时,很容易遇到组合梯度爆炸问题,即出

现梯度在反向传播时急剧增大,导致权重更新幅度过大,引发模型参数不稳定甚至发散的现象。即便是融合模型,也无法解决机器学习在油耗预测领域中的根本问题,即油耗模型的泛化性,不能囊括各种车型以及驾驶场景。虽然有些模型在类似车型或者相近的驾驶环境中能够表现出不错的泛化性,但如果车型差异过大,或核心部件变更,则需要重新训练模型。目前有研究人员尝试采用迁移学习方法解决油耗模型泛化性的问题。厉浩^[60]提出了一种基于模块图卷积与配置迁移的多类型商用车油耗预测方法,通过构建模块图卷积网络(module graphy convolutional netural network, MGCN)和配置迁移模块(configuration transfer, CT),提取商用车模块特征和整体特征的融合特征,以及不同配置对油耗的分配权重,实现多类型商用车的油耗预测,试验结果表明,MGCN-CT模型在多个评价指标上均优于基线模型,具有较高的预测精度和鲁棒性。

综上所述,没有任何单一模型能始终确保燃油消耗预测的高度准确性。相反,为了达到最佳预测效果,需要依据具体的应用场景来挑选最合适的模型。本文中讨论的几种模型之间的对比与总结如表4所示,这些模型代表了当前在数据驱动应用领域中使用最为广泛的几种方法。通过模型特性针对性的选择,可以更有效地提高预测精度并满足实际需求。

表4 油耗模型对比分析

类型	作者	模型	模型输入	模型表现	模型特点
多元回归	阮连杰 ^[25]	多元回归模型	行驶速度、加速度和道路坡度组合特征	$E_{MA} = 0.117$	适用于规模较小、有明显线性关系的数据集,具有较强的可解释性,便于理解和应用,但是自变量之间存在显著相关性的情况下,容易遇到多重共线性的问题
	郭娟 ^[26]	多元线性回归模型	功率、车质量、传动比和功率和车质量乘积的交互项	$R^2 = 0.8488$	
	Madhusudhanan ^[27]	多元回归模型	发动机转速和转矩	拟合度为91.2%	
	范童谣 ^[29]	多元回归模型	车速、VSP	$E_R < 10\%$	
浅层机器学习	刘昌海 ^[31]	SVR	速度、加速度、坡度、转矩、发动机转速	$R^2 = 0.95$	SVM、SVR 适用于规模较小、非线性明显的数据集中,合适的核函数可以有效提升模型的预测精度,因此当面对新数据集时,需要考虑核函数选择,增大计算开销
	周祎 ^[33]	LSSVM	平均速度、平均加速度、平均减速度、匀速比例等	$R^2 > 0.96$	
	赵明 ^[35]	RF	车重、月份、平均车速	$R^2 = 0.831$	RF 在处理高维数据方面表现出色,拥有较强的泛化能力和稳定性;然而,随着决策树数量的增加,计算成本也会相应上升
	阮连杰 ^[25]	RF	行驶速度、加速度和道路坡度组合特征	$E_{MA} = 0.067$	
	Nagar ^[37]	GBR	最大 PTO 功率、发动机转速以及发动机转速降	$R^2 = 0.897$, $E_{MAP} = 11.49\%$	
	Bousonville ^[38]	GBR	卡车类型、平均速度、速度防擦好、道路类型、总车重	$E_{MA} = 3.34$, $E_{RMS} = 4.65$, $R^2 = 0.917$	
	林若奇 ^[39]	XGboost	平均速度、平均油门开度、平均加速度等	$R^2 = 0.81$	GBDT 善于处理高维数据,具有较强的泛化能力和稳定性,预测精度较高,同时可以对输入特征重要性进行评估。由于其串行的计算特点,计算成本高,同时需要调优的超参数过多
	Su ^[40]	XGboost	船型、距离、燃油价格、航速、燃油成本、港口时间	$R^2 = 0.968$	
	Zhu ^[41]	LightGBM	平均车速、平均减速度、平均加速度、加速占比等	$E_{MA} = 2.036$, $E_{MS} = 13.042$, $R^2 = 0.873$	

表 4(续)

类型	作者	模型	模型输入	模型表现	模型特点
	Sun ^[44]	ANN	车辆速度、VSP、加速度、环境温度、道路坡度	绝对误差(微观)为 1%~2%、绝对误差(中观)为 5%~9%	ANN 适用于规模较大,非线性明显的数据集中,具有很强的非线性映射能力和自适应学习能力,但对数据质量要求较高
	Katreddi ^[45]	ANN	发动机负载、发动机转速和车速	$R^2=0.776\ 2$	
	郭宇 ^[46]	BP	负荷率、转速、速度、转矩、换挡次数、刹车次数	$E_{\text{RMS}}=13.69$, $E_{\text{MA}}=8.8$	
	Wang ^[50]	LSTM	速度、加速度、GPS 信号	$E_{\text{R}}<0.097$	
	张立成 ^[51]	LSTM	速度、加速度、Jerk	$R^2>0.996$	
深度学习	王一婷 ^[52]	Shared-LSTM	平均车速、最大车速、车速标准差、加速比例、减速比例等	$E_{\text{RMS}}<0.057$, $E_{\text{MS}}<0.238$, $E_{\text{MA}}<0.018$	LSTM 适用于规模较大的时序数据,能够捕捉时序数据中的依赖关系,有效处理时序任务,且解决了传统 RNN 容易出现梯度消失、爆炸的问题,但对数据质量要求较高
	吴丽宁 ^[53]	BiLSTM	最大平均速度、最大速度、最小速度、平均速度、速度标准差等	$E_{\text{MA}}<0.195$, $E_{\text{RMS}}<0.366$	
	师国东 ^[54]	XGBoost-MSIWOA-LSTM	发动机转速、发动机转矩、油门开度、机油压力等	$R^2=0.997\ 5$	
混合式 油耗预测 模型	张隅希 ^[55]	MIPCA-MLP	速度、发动机转速、转矩百分比、发动机负荷率等	$R^2=0.92$	混合油耗预测模型能够集合每个模型的优势,实现对不同特征信息的深入挖掘,从而提取出更为准确且有效的特征,适用于大规模、高纬度数据集。但模型相对复杂,运算成本增加
	杨冉 ^[56]	CNN-LSTM	速度、加速度、Jerk	$R^2>0.9$	
	Song ^[59]	VMD-Informer	主泵压力、主泵流量、摆动马达压力等	$R^2>0.97$	

3 总结与展望

本文中系统梳理了车辆油耗预测传统模型与数据驱动模型两类模型的研究进展。传统模型基于汽车相关理论、车辆结构和发动机原理,结合驾驶行为、道路条件等多种油耗影响因素构建模型,具有逻辑清晰的优势,但对专业知识依赖度高且建模复杂。随着大数据技术发展,基于机器学习的预测方法凭借其高效性和高精度逐渐成为主流。由于多数机器学习方法属于黑盒模型,无需研究人员掌握深刻的汽车领域知识,更易于领域外的学者开展研究。本文中着重对机器学习方法进行了整理分析,将其细分为多元回归油耗预测模型、浅层机器学习、深度学习以及混合式油耗预测模型。

多元回归模型通过线性或非线性方程拟合变量关系,为有效提高模型性能,多数学者倾向于选择车速、加速度或 VSP 作为模型的自变量。然而,车辆燃油消耗是极其复杂且动态变化的过程,一般的线性或非线性方程往往难以准确描述油耗的变化。SVM、RF 和 GBDT 等浅层学习模型擅长通过特征参数处理中小规模数据,同时 RF 和 GBDT 的集成学习机制可有效处理高维特征,但当面临大规模且结构更为复杂的数据时,浅层机器学习模型的表现可能会受到限制。深度学习模型,如 ANN、LSTM 等通过多层神经网络以及高级神经元,深入挖掘影响油耗因素之间的复杂交互作用,自动特征提取适应大规模复杂数据,提供更加精准的预测结果,但对数据质量和数量敏感;混合模型融合多种算法优势,在特征挖掘深度上表现突出,但存在模型结构复杂且计算成本较高的问题。

当前基于机器学习的油耗预测研究仍存在一定的局限性:1) 环境因素考虑不足,多数模型仅纳入坡度等个别参数,缺乏温度、湿度、交通流量等多元环境影响因素的系统建模;2) 动态变量关注欠缺,目前多

数研究没有考虑车辆载质量变化对油耗的影响,仅针对同一载质量进行建模,或默认载质量不变;3)模型泛化能力受限,特别是深度学习方法严重依赖特定车型数据,迁移应用成本较高。

未来发展方向应聚焦以下三个维度:1)多源数据融合,构建涵盖人-车-路-环境的多维输入体系,整合车载传感、地理信息和环境监测等多模态数据;2)模型结构优化,在混合模型框架下探索轻量化设计,通过特征选择、模型剪枝等技术平衡精度与效率;3)迁移学习应用,开发跨车型自适应算法,利用预训练模型结合少量新数据实现模型迁移,突破数据壁垒。此外,需加强能耗机理与数据特征的耦合研究,构建解释性更强的智能预测模型。

参考文献:

- [1] REN S J, LI T T, LI G, et al. Investigation of heavy-duty vehicle chassis dynamometer fuel consumption and CO₂ emissions based on a binning-reconstruction model using real-road data[J]. *Atmosphere*, 2023, 14(3): 528.
- [2] 帅石金, 王志, 马骁, 等. 碳中和背景下内燃机低碳和零碳技术路径及关键技术[J]. *汽车安全与节能学报*, 2021, 12(4): 417-439.
- [3] 中华人民共和国公安部交通管理局. 全国机动车保有量达 4.35 亿辆 驾驶人达 5.23 亿人 新能源汽车保有量超过 2000 万辆[EB/OL]. (2024-01-11) [2024-12-06]. <https://www.mps.gov.cn/n2254098/n4904352/c9384864/content.html>.
- [4] 中华人民共和国公安部交通管理局. 2024 年上半年全国机动车达 4.4 亿辆 驾驶人达 5.32 亿人 新能源汽车保有量达 2472 万辆[EB/OL]. (2024-07-09) [2024-12-06]. <https://www.mps.gov.cn/n2254098/n4904352/c9650715/content.html>.
- [5] 中华人民共和国公安部交通管理局. 全国机动车保有量达 4.17 亿辆 驾驶人超过 5 亿人 新能源汽车保有量达 1310 万辆 同比增长 67.13% [EB/OL]. (2023-01-11) [2024-12-06]. <https://www.mps.gov.cn/n2254314/n6409334/c8837510/content.html>.
- [6] 中华人民共和国公安部交通管理局. 2021 年全国机动车保有量达 3.95 亿 新能源汽车同比增 59.25% [EB/OL]. (2022-01-11) [2024-12-06]. <https://www.mps.gov.cn/n2254314/n6409334/c8322353/content.html>.
- [7] 中华人民共和国公安部交通管理局. 2020 年全国新注册登记机动车 3328 万辆 新能源汽车达 492 万辆[EB/OL]. (2021-01-07) [2024-12-06]. <https://www.mps.gov.cn/n2254098/n4904352/c7647179/content.html>.
- [8] 中华人民共和国公安部交通管理局. 今年上半年全国机动车保有量达 3.4 亿辆[EB/OL]. (2019-07-04) [2024-12-06]. <https://www.mps.gov.cn/n2254314/n6409334/c6830663/content.html>.
- [9] 张东. 机车设备的能源管理与节能减排策略研究[J]. *中国设备工程*, 2024(20): 48-50.
- [10] 傅立新, 贺克斌, 何东全, 等. MOBILE 汽车源排放因子计算模式研究[J]. *环境科学学报*, 1997, 17(4): 474-479.
- [11] GIANNOULI M, KALOGNOMOU E A, MELLIOS G, et al. Impact of European emission control strategies on urban and local air quality[J]. *Atmospheric Environment*, 2011, 45(27): 4753-4762.
- [12] DEMIR E, BEKTAS T, LAPORTE G. A comparative analysis of several vehicle emission models for road freight transportation[J]. *Transportation Research Part D: Transport and Environment*, 2011, 16(5): 347-357.
- [13] LANG J L, CHENG S Y, ZHOU Y, et al. Air pollutant emissions from on-road vehicles in China, 1999-2011[J]. *Science of the Total Environment*, 2014, 496: 1-10.
- [14] REID S, BAI S, DU Y, et al. Emissions modeling with MOVES and EMFAC to assess the potential for a transportation project to create particulate matter hot spots[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2016, 2570: 12-20.
- [15] 彭飞. 基于比功率的混合动力汽车交通路网能耗量化方法[D]. 北京: 北京交通大学, 2023.
- [16] EPA. Draft motor vehicle emission simulator (MOVES) [R]. Washington, USA: Environmental Protection Agency, 2009.
- [17] DAVIS N, LENTS J, OSSES M, et al. Development and application of an international vehicle emissions model[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2005, 1939(1): 156-165.
- [18] BARTH M, AN F, YOUNGLOVE T, et al. Development of a comprehensive modal emissions model[R]. Washington D C, USA: NCHRP, 2000.
- [19] AHN K, RAKHA H, TRANI A, et al. Estimating vehicle fuel consumption and emissions based on instantaneous speed and

- acceleration levels[J]. *Journal of Transportation Engineering*, 2002, 128(2): 182–190.
- [20] CHEN Z, YANG C, CHEN A. Estimating fuel consumption and emissions based on reconstructed vehicle trajectories[J]. *Journal of Advanced Transportation*, 2014, 48(6): 627–641.
- [21] ZHAO D F, LI H Y, HOU J J, et al. A review of the data-driven prediction method of vehicle fuel consumption[J]. *Energies*, 2023, 16(14): 5258.
- [22] JANIESCH C, ZSCHECH P, HEINRICH K. Machine learning and deep learning[J]. *Electronic Markets*, 2021, 31(3): 685–695.
- [23] 余舒. 重型车辆实际道路 NO_x 排放算法研究[D]. 武汉: 武汉理工大学, 2019.
- [24] CHEN Y M, SUN B Z, XIE X W, et al. Short-term forecasting for ship fuel consumption based on deep learning[J]. *Ocean Engineering*, 2024, 301: 117398.
- [25] 阮连杰. 高速公路小汽车瞬态油耗模型构建及节油措施设计[D]. 广州: 华南理工大学, 2023.
- [26] 郭娟. 基于存在交互项的多元线性回归汽车油耗预测模型[J]. *广西质量监督导报*, 2019(11): 138–140.
- [27] MADHUSUDHANAN A K, NA X X, AINALIS D, et al. Engine fuel consumption modelling using prediction error identification and on-road data[J]. *IEEE Transactions on Intelligent Vehicles*, 2023, 8(2): 1392–1402.
- [28] JIMÉNEZ-PALACIOS J L. Understanding and quantifying motor vehicle emissions with vehicle specific power and TILDAS remote sensing[J]. *Massachusetts Institute of Technology*, 1999: 345–361.
- [29] 范童谣. 基于数据驱动的商用物流车行驶工况构建与能效特征研究[D]. 重庆: 重庆理工大学, 2024.
- [30] 陈旗. 基于 VSP 的城市道路油耗预测模型研究[D]. 北京: 北方工业大学, 2019.
- [31] 刘昌海. 基于 IGWO 优化支持向量回归的重型车油耗预测研究[D]. 重庆: 重庆交通大学, 2024.
- [32] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers[J]. *Neural Processing Letters*, 1999, 9(3): 293–300.
- [33] 周祎. 基于道路行驶工况的重型卡车油耗预测研究[D]. 西安: 长安大学, 2023.
- [34] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5–32.
- [35] 赵明. 基于远程传输技术的重型柴油车运行特征分析研究[D]. 天津: 天津科技大学, 2023.
- [36] MASSOUD R, BELLOTTI F, BERTA R, et al. Exploring fuzzy logic and random forest for car drivers' fuel consumption estimation in IoT-enabled serious games [C]//*Proceedings of IEEE 14th International Symposium on Autonomous Decentralized System*. Utrecht, Netherlands: IEEE, 2019: 1–7.
- [37] NAGAR H, MACHAVARAM R, AMBUJ, et al. Cloud-driven serverless framework for generalised tractor fuel consumption prediction model using machine learning[J]. *Cogent Engineering*, 2024, 11(1): 2311810.
- [38] BOUSONVILLE T, KAMGA D C, KRÜGER T, et al. Data driven analysis and forecasting of medium and heavy truck fuel consumption[J]. *Enterprise Information Systems*, 2022, 16(6): 1856417.
- [39] 林若奇. 基于驾驶风格的重型货车油耗预测模型研究[D]. 杭州: 杭州电子科技大学, 2022.
- [40] SU M, SU Z Q, CAO S L, et al. Fuel consumption prediction and optimization model for pure car/truck transport ships[J]. *Journal of Marine Science and Engineering*, 2023, 11(6): 11061231.
- [41] ZHU X Y, SHEN X Y, CHEN K L, et al. Research on the prediction and influencing factors of heavy duty truck fuel consumption based on LightGBM[J]. *Energy*, 2024, 296: 131221.
- [42] LI Y W, ZENG I Y, NIU Z H, et al. Predicting vehicle fuel consumption based on multi-view deep neural network[J]. *Neurocomputing*, 2022, 502: 140–147.
- [43] KARLAFTIS M G, VLAHOIANNI E I. Statistical methods versus neural networks in transportation research: differences, similarities and some insights[J]. *Transportation Research Part C: Emerging Technologies*, 2011, 19(3): 387–399.
- [44] SUN R X, CHEN Y C, DUBEY A, et al. Hybrid electric buses fuel consumption prediction based on real-world driving data[J]. *Transportation Research Part D: Transport and Environment*, 2021, 91: 102637.
- [45] KATREDDI S, THIRUVENGADAM A. Trip based modeling of fuel consumption in modern heavy-duty vehicles using artificial intelligence[J]. *Energies*, 2021, 14(24): 8592.
- [46] 郭宇. 基于改进 BP 神经网络的重型卡车油耗预测研究[D]. 西安: 长安大学, 2022.
- [47] CARVALHO E, FERREIRA B V, FERREIRA J, et al. Exploiting the use of recurrent neural networks for driver behavior profiling [C]//*Proceedings of International Joint Conference on Neural Networks*. Anchorage, USA: IEEE, 2017:

- 3016–3021.
- [48] XU Z G, WEI T, EASA S, et al. Modeling relationship between truck fuel consumption and driving behavior using data from Internet of vehicles[J]. Computer-Aided Civil and Infrastructure Engineering, 2018, 33(3): 209–219.
- [49] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to forget: continual prediction with LSTM[C]//Proceedings of 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. Edinburgh, UK: IET, 1999.
- [50] WANG G Q, ZHANG L C, XU Z G, et al. Predictability of vehicle fuel consumption using LSTM: findings from field experiments[J]. Journal of Transportation Engineering, Part A: Systems, 2023, 149(5): 7643.
- [51] 张立成. 考虑 Jerk 的汽车油耗预测模型与速度优化方法研究[D]. 西安: 长安大学, 2022.
- [52] 王一婷. 基于车联网大数据的驾驶风格经济性预测模型与方法研究[D]. 西安: 长安大学, 2023.
- [53] 吴丽宁. 基于驾驶风格分类的卡车油耗预测[D]. 西安: 长安大学, 2020.
- [54] 师国东, 胡明茂, 宫爱红, 等. 基于 XGBoost-MSIWOA-LSTM 的车辆油耗优化预测模型[J/OL]. 计算机集成制造系统: 1–24. (2023–07–25) [2024–10–11]. <http://kns.cnki.net/kcms/detail/11.5946.TP.20230724.1508.002.html>.
- [55] 张隅希. 面向节能的重卡驾驶行为分析方法研究[D]. 西安: 长安大学, 2022.
- [56] 杨冉. 融合驾驶行为特征的 CNN-LSTM 油耗预测方法研究[D]. 西安: 长安大学, 2023.
- [57] VASWANI V, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS). California, USA: NIPS, 2017: 6000–6010.
- [58] ZHOU H Y, ZHANG S H, PENG J Q, et al. Informer: beyond efficient transformer for long sequence time-series forecasting[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(12): 11106–11115.
- [59] SONG H J, LI G Q, LI X H, et al. Developing a data-driven hydraulic excavator fuel consumption prediction system based on deep learning[J]. Advanced Engineering Informatics, 2023, 57: 102063.
- [60] 厉浩. 基于模块图卷积与配置迁移的多类型商用车油耗预测问题的研究[D]. 济南: 山东大学, 2023.

Research progress of vehicle fuel consumption prediction models

GUAN Peng^{1,2}, REN Shuojin², SHEN Yitao^{1,3*}, ZHAO Jianfu²

1. School of Automotive Engineering, Harbin Institute of Technology (Weihai), Weihai 264209, China;

2. CATARC Automotive Research and Inspection Center(Tianjin) Co., Ltd., Tianjin 300300, China;

3. Key Laboratory for Power Machinery and Engineering of Ministry of Education, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: To accurately predict the fuel consumption characteristics of vehicles under various operating conditions, assist researchers in gaining a deeper understanding of the patterns of fuel consumption changes in engines and further optimize engine performance, this paper aims to comprehensively summarize and analyze existing fuel consumption prediction models and categorize them into two main types: traditional fuel consumption prediction models and data-driven machine learning-based fuel consumption prediction models. For the latter category, this paper further divides it into four subcategories: multiple regression, shallow machine learning, deep learning, and hybrid fuel consumption models, detailing the application status, advantages, and limitations of each method and its variants. Through comparative analysis of these models, this paper not only clarifies their optimal application scenarios but also points out the main problems and challenges present in current research. When dealing with data that exhibits strong linear correlations, multiple regression methods perform well, offering high model transparency and ease of understanding. Machine learning approaches, especially deep learning, can effectively address more complex nonlinear relationships, fully exploiting features within the data to achieve precise predictions of fuel consumption, albeit with higher requirements for data quality and relatively complex models. Finally, based on the characteristics and applications of different models, this paper provides an outlook on the future development of fuel consumption prediction.

Keywords: fuel consumption prediction; data-driven; machine learning; deep learning; hybrid fuel consumption model

(责任编辑:刘丽君)