

doi: 10.19562/j.chinasae.qcgc.2025.01.011

基于三层加权堆叠模型的电动汽车剩余里程预测*

石琴^{1,2,3}, 侯伟路^{1,2,3}, 张晓楠^{1,2,3}, 吴为教^{1,2,3}, 贺泽佳^{1,2,3}

- 合肥工业大学汽车与交通工程学院, 合肥 230009;
- 合肥工业大学自动驾驶车辆安全技术安徽省重点实验室, 合肥 230009;
- 安徽省智慧交通车路协同工程研究中心, 合肥 230009)

[摘要] 为实现电动汽车剩余里程准确预测, 本文提出一种基于三层加权堆叠模型的电动汽车剩余里程预测方法。结合最大信息系数和斯皮尔曼系数作为变量评价准则, 使用最小冗余最大相关算法从候选特征集中优化得到输入特征集。构建考虑原始训练特征的三层堆叠模型, 并利用贝叶斯优化算法得到堆叠模型中基模型权重。最后, 使用输入特征集对三层加权堆叠模型训练并实现电动汽车剩余里程预测, 结果表明所述三层加权堆叠模型的预测精度较高。此外, 与其他模型相比, 所述堆叠三层加权模型的泛化能力更强。

关键词: 电动汽车; mRMR算法; Stacking模型; 剩余里程

Electric Vehicle Remaining Range Prediction with a Three-Layer Weighted Stacking Model

Shi Qin^{1,2,3}, Hou Weilu^{1,2,3}, Zhang Xiaonan^{1,2,3}, Wu Weijiao^{1,2,3} & He Zejia^{1,2,3}

- School of Automotive and Transportation Engineering, Hefei University of Technology, Hefei 230009;
- Key Laboratory for Automated Vehicle Safety Technology of Anhui Province, Hefei University of Technology, Hefei 230009;
- Engineering Research Center for Intelligent Transportation and Cooperative Vehicle-Infrastructure of Anhui Province, Hefei 230009

[Abstract] To achieve accurate prediction of electric vehicle remaining range, a method based on a three-layer weighted stacking model for predicting remaining range of electric vehicles is proposed in this paper. By combining the maximal information coefficient and Spearman correlation coefficient as criteria for variable evaluation, the minimum redundancy maximum relevance algorithm is employed to optimize and obtain the input feature set from the candidate features. A three-layer stacking model that incorporates the original training features is then constructed, and Bayesian optimization algorithm is used to determine the weights of the base models within the stacking model. Finally, the input feature set is used to train the three-layer weighted stacking model and realize electric vehicle remaining range prediction. The results show that the proposed three-layer weighted stacking model has high prediction accuracy and, compared to other models, with stronger generalization capabilities.

Keywords: electric vehicle; mRMR algorithm; Stacking model; remaining range

前言

随着国家经济的快速发展和居民生活水平的不断提升, 汽车逐渐成为居民出行的主要交通工具, 公安部数据显示 2023 年全国汽车保有量达到 3.36 亿

辆^[1]。汽车为居民出行带来便利的同时也带来化石燃料短缺和环境污染等问题, 国际能源署的数据显示 2023 年全国交通运输行业碳排放量增长 1.36 亿 t。针对这一现状, 国务院颁布《2030 年前碳达峰行动方案》, 指出交通运输绿色低碳行动是实现“碳达峰”的重要组成部分; 交通运输部颁布《绿色交通“十四

* 安徽省重点研发项目(202304A05020087)资助。

原稿收到日期为 2024 年 06 月 23 日, 修改稿收到日期为 2024 年 07 月 24 日。

通信作者: 贺泽佳, 博士, E-mail: hezejia@hfut.edu.cn。

五”发展规划》,指出交通运输行业须加快新能源和清洁能源运输装备的推广和应用,以快速形成绿色低碳的运输方式。在此背景下,发展电动汽车成为交通运输行业促进绿色低碳发展的重要战略动向,也是汽车行业转型升级的必由之路。

与传统燃油汽车相比,电动汽车在其生命周期内可以节省35%以上的能源消耗和减少20%以上的碳排放^[2]。然而,电动汽车在其发展过程中仍面临挑战,与传统燃油汽车相比,电动汽车储能容量相对较小,补充能量时间较长,导致其续航里程受到限制^[3]。此外,车辆运行环境和车辆行驶条件的实时变化导致电动汽车剩余里程难以准确预测,造成驾驶员在行驶中极易出现“里程焦虑”问题,极大程度上影响驾驶员对电动汽车的接受度^[4-5]。

针对上述问题,近年来随着人工智能和数据存储技术的快速发展,国内外专家致力于使用机器学习研究电动汽车剩余里程预测。陈德海等^[6]采用粒子群优化算法选出径向神经网络中的相关参数组合,并根据工况对剩余里程预测模型进行了实时更新和误差修正以提高模型预测精度。田慧欣等^[7]根据地图信息数据预测未来行驶工况,并将工况参数作为循环支持向量机预测模型的输入特征,实现对应行驶工况下单位里程预测。Yavasoglu等^[8]采用决策树估计道路类型并对其进行了分类,并采用相关机器学习算法实现剩余里程的在线估计。Ayevide等^[9]在模型训练过程中采用列文伯格-马夸尔特算法优化模型参数,获得较好预测结果。Bi等^[10]建立不同温度条件下的剩余里程非线性估计模型,使用非线性回归方法优化模型结构,进而实现剩余里程的准确估计。然而,车辆内在特性、驾驶员行为习惯和交通环境状况等都会影响电动汽车剩余里程,但目前研究主要使用单一模型对剩余里程预测,忽略多模型融合的潜在优势^[11-12]。此外,目前研究仅采用单一优化方式调整模型结构,容易陷入局部最优。

在机器学习领域,特征提取可以解决原始数据存在数据维度过高导致预测时间过长和预测结果不准确的问题,有效挖掘对模型预测结果有积极影响的特征。Huang等^[13]通过网格搜索方式在充电电压曲线中寻找合适的电压段,利用主成分分析提取特定电压段的主成分,提高预测精度。Wu等^[14]采用随机森林算法选择相关性较高的特性参数作为模型输入特征,有效提高模型鲁棒性。Zhong等^[15]使用多元正则化实现特征聚类并选择特征以提高模型预测精

度。传统特征选择方法往往依赖相关领域专家的经验,在处理变化性较大的数据集时表现不佳^[16-17]。此外,目前研究未同时考虑特征与目标的相关性及特征之间的冗余性,导致模型无法处理冗余特征对模型性能的影响。

因此,本文提出一种基于三层加权堆叠模型的电动汽车剩余里程预测方法。首先,本文使用最小冗余最大相关算法对电动汽车实际运行数据集进行优化改进,从候选特征集中得到12个特征作为输入特征集;然后,构建考虑原始训练特征三层堆叠模型,并利用贝叶斯优化算法得到三层堆叠模型中基模型权重;最后,使用输入特征集对三层加权堆叠模型训练并实现电动汽车剩余里程预测。实验结果表明,与传统电动汽车剩余里程预测算法相比,本文提出的基于三层加权堆叠模型的电动汽车剩余里程预测方法预测精度更高,泛化能力更强。

1 特征提取

1.1 数据来源

本文数据集来自新能源汽车监管平台10辆电动汽车1年的实际运行数据,通过车载终端进行采集,并以15 s的采样时间间隔发送到数据中心。数据集涵盖电动汽车车辆编号、时间标签、车辆状态、充电状态、运行模式、速度、累计里程、电池总电压、电池总电流、电池单体电压阈值等。

本文研究旨在实现电动汽车剩余里程预测,具体涉及电动汽车放电过程。因此,本文首先依据车辆状态和充电状态筛选电动汽车放电过程运行数据,并删除无用数据列,以确保后续建模分析准确性。其次,依据拉依达准则进行异常值判断,将识别出的异常值置为空值,保留数据完整性和信息价值,然后采用拉格朗日插值法填充原始数据中的空值或缺失值。最后,考虑到电池寿命与其充放电深度密切相关,保持30%的剩余电量可以避免深度放电,延长电池寿命^[18],本文以30%的剩余电量作为最低安全放电电量^[19-20],筛选出剩余电量范围在30%~100%构造剩余里程数据集,由此得到电动汽车剩余里程预测候选特征集,如表1所示。此外,本文涉及大量算法、模型以及误差函数,为了方便读者阅读,本文统一将缩写词以表格形式表示,如表2所示。

1.2 最小冗余最大相关算法

考虑到特征与目标存在相关性及特征与特征之

表1 电动汽车剩余里程预测候选特征集

特征类型	特征名称
车辆内在特性	车速、电压、电流、电池荷电状态、电池单体电压最高值、电池单体电压最低值、温度最高值、温度最低值、驱动电机转速、驱动电机转矩、驱动电机温度、驱动电机控制器温度、驱动电机控制器电压、驱动电机控制器电流
驾驶员行为习惯	加速踏板行程、制动踏板状态
外部运行环境	气温、湿度、降水量、风速

表2 缩写词与定义对照表

缩写词	定义
mRMR	最小冗余最大相关算法
BOA	贝叶斯优化算法
XGBoost	极限梯度提升树模型
RF	随机森林模型
GBDT	梯度提升决策树模型
SVM	支持向量机模型
BP	BP神经网络模型
PCC	皮尔逊相关系数
VTM	方差选择
EMS	手动选择
MSE	平均均方误差
MAE	平均绝对误差
MAPE	平均绝对百分比误差
R ²	决定系数

间存在冗余性,传统模型无法处理冗余特征对模型性能影响。因此,本文提出最小冗余最大相关(mRMR)算法,不仅可以分析特征与目标之间是否具有最强相关性,而且探究特征之间是否冗余以提高模型预测精度^[21],具体算法框架如图1所示。

本文引入最大信息系数和斯皮尔曼系数作为评价候选特征集S中的m个特征与目标之间的相关性和特征之间的冗余性,并通过权重因子分别赋予二者不同权重,其融合系数表示为

$$M(u, v) = \omega_m MIC(u, v) + \omega_s SC(u, v) \quad (1)$$

$$MIC(u, v) = \max_{a < b < G} \frac{I(u, v)}{\log_2 \min(a, b)} \quad (2)$$

$$SC(u, v) = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (3)$$

式中: ω_m 和 ω_s 分别是最大信息系数 $MIC(\cdot)$ 和斯皮尔曼系数 $SC(\cdot)$ 的权重,取值范围是0~1,且满足 $\omega_m + \omega_s = 1$;a和b分别是变量u和v方向的网格数目;G是网格总数目; d_i 是观测值的秩次差;n是样本数量。

候选特征集S中的m个特征与目标需要满足以下条件:

$$\begin{cases} \max R(S, y), R = \frac{1}{|S|} \sum_{x_i \in S} M(x_i, y) \\ \min R'(S), R' = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} M(x_i, x_j) \end{cases} \quad (4)$$

式中:|S|是候选特征集S的特征数; x_i 和 x_j 是候选特征集S中的互异特征;y是目标。候选特征集S中m个特征与目标之间的最小冗余最大相关表示为

$$\max \phi(R, R'), \phi = \frac{R}{R'} \quad (5)$$

式中采用增量搜索法最大化 ϕ 近似得到最优特征,在获得包含n个特征的特征子集 S_n 后,从剩余 $S - S_n$ 特征集查找:

$$\max_{x_j \in S - S_n} [I(x_i, y) - \frac{1}{n} \sum_{x_i \in S_n} I(x_i, x_j)] \quad (6)$$

以平均均方误差衡量模型预测效果,选出预测

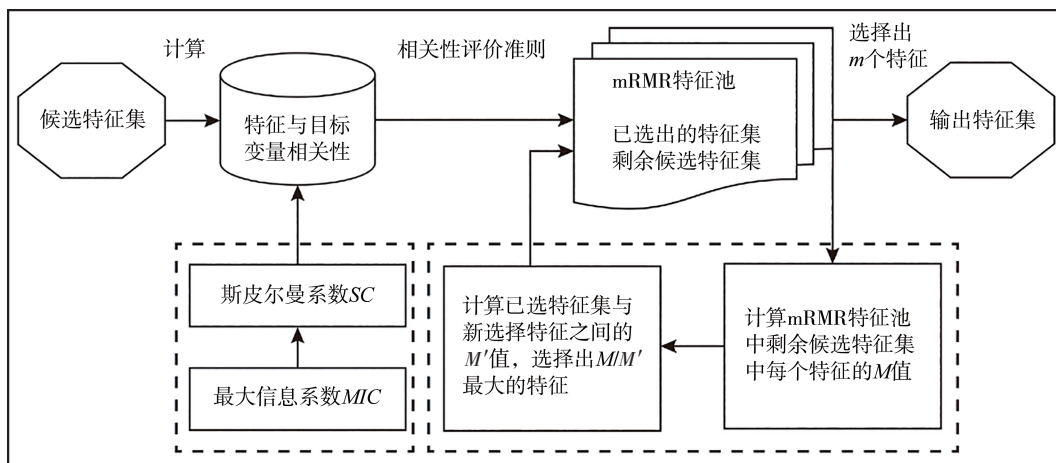


图1 mRMR算法框架

效果最好的特征子集作为输出特征集。

2 三层加权堆叠模型

2.1 贝叶斯优化算法

贝叶斯优化算法(BOA)在已有历史数据和已知先验信息的基础上,通过有针对性地参数空间中进行搜索,并不断自适应地调整采样点的位置,以有效地搜索到全局的最优解^[22]。与其它优化算法相比,贝叶斯优化算法有针对性地选择可以保证在较少函数评估次数内找到全局最优解,很大程度上提高计算效率。贝叶斯优化算法在寻找最优解的过程中实现对目标函数的后验分布估计,并在此基础上寻找下一个需要采样的参数组合。该方法优化过程主要包含概率代理模型和采集函数两部分。

概率代理模型是指贝叶斯优化算法对目标函数的构建方式,可以通过几个观测点实现对目标函数分布估计。贝叶斯优化算法中有多种概率代理模型可供选择,本文选取高斯过程作为贝叶斯优化算法的概率代理模型。高斯过程是一种非参数的概率分布模型,它对函数的每一个取值都定义一个概率分布,因而可以实现对未知目标函数的灵活估计。高斯过程的具体表达形式表示为

$$f(x) \sim GP(\mu(x), k(x, x')) \quad (7)$$

式中: $f(x)$ 是目标函数; $\mu(x)$ 是高斯过程中 $f(x)$ 的均值函数; $k(x, x')$ 是协方差函数。

采集函数用于在当前概率代理模型下选择出一个要评估的点。采集函数根据后验分布在未探索区域与可能出现最优解的区域进行采样,以实现搜索过程的引导。本文选取期望提升函数作为采集函数,期望提升函数的具体表达式表示为

$$EI(x) = E[\max(f(x) - f(best), 0)] \quad (8)$$

式中 $f(best)$ 是当前最佳目标函数值。

确定概率代理模型和采集函数后,需要选择初始采样点,在参数空间中随机选择或者基于先验知识选取初始采样点用于初始化。然后,依据高斯过程和期望提升函数进行不断迭代以实现概率代理模型中参数更新,并不断引导下一个采样点的选择,最终找到使得目标函数最小化的参数值以及对应的目标函数值。

2.2 堆叠模型

堆叠模型(Stacking)是一种有效集成学习方法,通过堆叠多个基础模型的预测结果来提升整体性能^[23]。Stacking模型采用三折交叉验证进行训练,假

设Stacking模型第1层结构有 n 个基模型,第2层有一个学习器,则训练过程如图2所示,具体步骤如下。

步骤1:将所有数据集划分为训练集和测试集并定义 $Train = (X_{tr}, Y_{tr})$ 和 $Test = (X_{te}, Y_{te})$,并将训练集数据划分为3等份 $Train_1, Train_2, Train_3$ 。

步骤2:将 $Train_1$ 和 $Train_2$ 作为训练集, $Train_3$ 作为测试集,选取基模型进行模型训练并得到训练集预测结果 $Train_1^1$ 和测试集预测结果 $Test_1^1$,然后在训练集中依次训练得到训练集预测结果 $Train_2^1$ 和 $Train_3^1$,以及测试集预测结果 $Test_2^1$ 和 $Test_3^1$ 。

步骤3:将训练集预测结果拼接,得到与原始数据样本量相同数据,作为下一层的输入特征 $Train^1 = [Train_1^1, Train_2^1, Train_3^1]^T$ 。

步骤4:将测试集预测结果 $Test_1^1, Test_2^1$ 和 $Test_3^1$ 取均值,作为下一层的模型输入 $Test^1$ 。

步骤5:对剩余 $n-1$ 个基模型重复步骤2到4,并将所有基模型训练集与测试集的预测结果合并得到:

$$Train^1 = \begin{bmatrix} Train_1^1 & Train_2^1 & \cdots & Train_n^1 \\ Train_1^2 & Train_2^2 & \cdots & Train_n^2 \\ Train_1^3 & Train_2^3 & \cdots & Train_n^3 \end{bmatrix} \quad (9)$$

$$Test^1 = [Test^1, Test^2, \cdots, Test^n] \quad (10)$$

步骤6:以基模型预测结果作为元模型输入特征,以数据集真实值作为输出标签,用元模型进行训练和预测,得到最终预测结果。

针对不同基础学习器差异性,本文提出差异化权重融合策略,根据每个基础学习器在不同条件下的性能表现,利用贝叶斯优化算法赋予权重,将模型权重和模型输出相结合代入到下一层得到最终的预测效果:

$$\min_{\vartheta} \frac{1}{N} \sum_{i=1}^N \left[\sum_{j=1}^c \vartheta_j o_{ij} - y_i \right]^2 \quad (11)$$

约束条件为

$$\begin{cases} \sum_{j=1}^c \vartheta_j = 1 \\ \vartheta_j \geq 0, j = 1, 2, \cdots, c \end{cases} \quad (12)$$

式中: ϑ_j 是基模型权重系数; o_{ij} 是基模型的输出; c 是每层结构基模型的个数; y_i 是剩余里程真实值。

所述三层加权Stacking模型包含基础层、泛化层和元模型层,如图3所示。基础层采用极限梯度提升树^[24](XGBoost)、随机森林^[25](RF)、梯度提升决策树^[26](GBDT)、支持向量机^[27](SVM)和BP神经网络^[28](BP)作为基础学习器,用以捕捉数据特征;泛

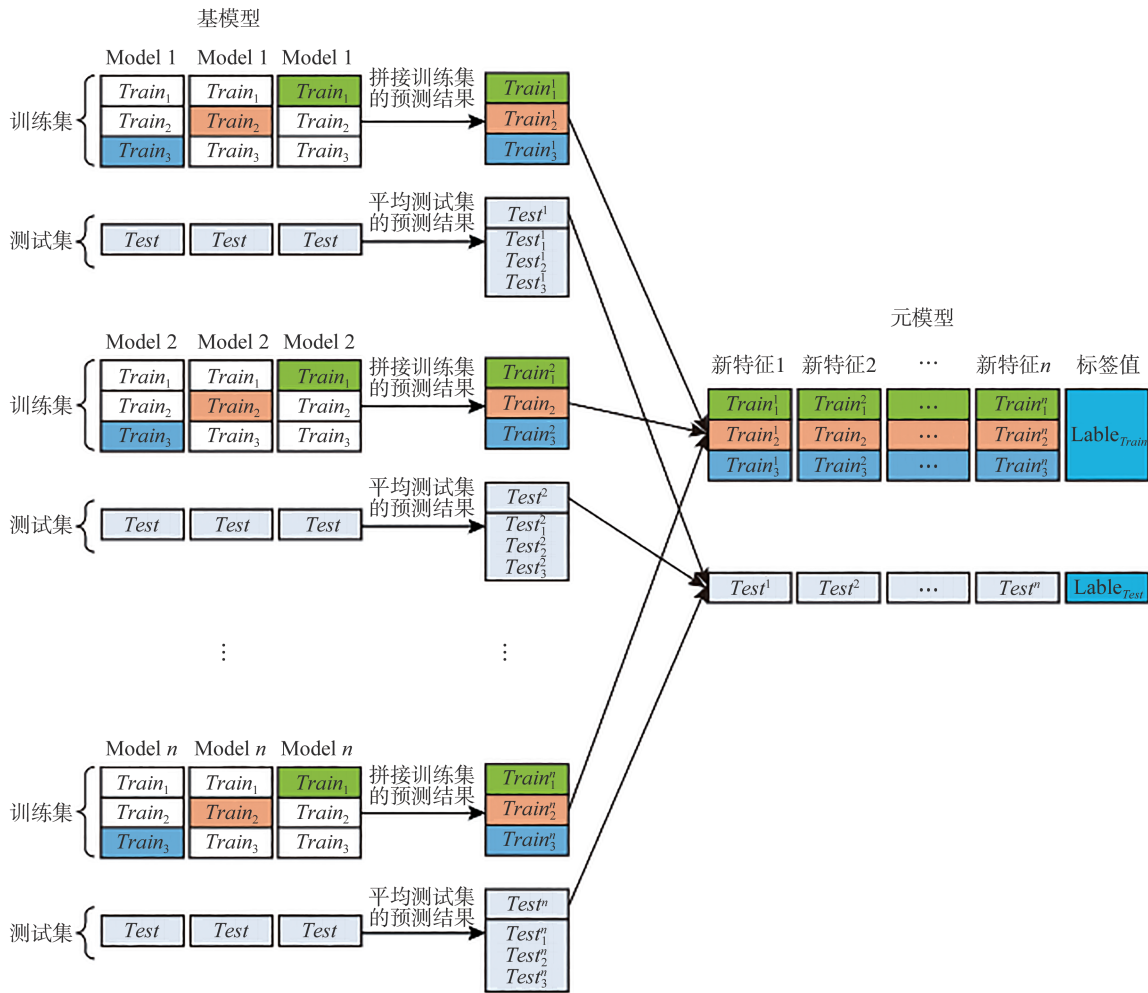


图2 三折交叉验证架构

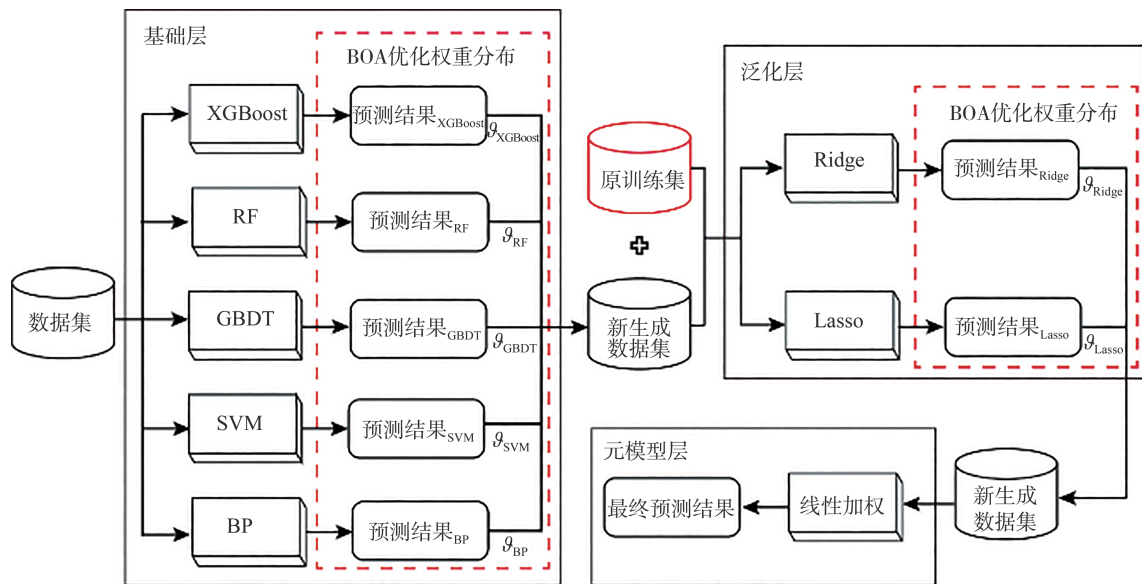


图3 三层加权Stacking模型结构

化层采用 Ridge 回归^[29]和 Lasso 回归^[30]作为基础学习器,用以融合基础层输出和原始特征;元模型层将不

同基础学习器预测结果进行线性加权,从而得到最终的预测结果。基于不同基模型权重系数,得到模

型预测结果,具体表示为

$$f(x) = \vartheta_{\text{Ridge}} f_{\text{Ridge}}(x) + \vartheta_{\text{Lasso}} f_{\text{Lasso}}(x) \quad (13)$$

式中: ϑ_{Ridge} 和 ϑ_{Lasso} 是Ridge回归和Lasso回归的权重系数; $f_{\text{Ridge}}(x)$ 和 $f_{\text{Lasso}}(x)$ 是Ridge回归和Lasso回归泛化层预测结果。

3 结果分析

3.1 mRMR算法结果分析

本文采用平均均方误差(MSE)评估模型预测效果,特征数量依次增加2个,直至输入特征数达上限,将输入特征逐一输入到预测模型中得到不同 ω_m 和 ω_s 权重取值下MSE变化趋势,如图4所示。

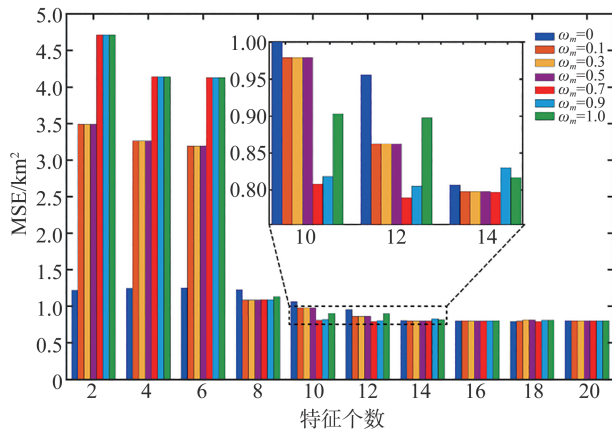


图4 不同权重mRMR算法预测误差

由图4可知,输入特征数量小于10时,不同 ω_m 和 ω_s 权重取值下MSE有显著差异。然而,输入特征数量达到10后,MSE逐渐趋于平稳。 ω_m 取值为0.7时,MSE最小。本文选取 ω_m 为0.7和 ω_s 为0.3作为最大信息系数和斯皮尔曼系数权重进一步优化,结果如图5所示。

由图5可知,随着输入特征数量增加,MSE整体呈下降趋势,输入特征数量达到10后,MSE逐渐趋于平稳。输入特征数量达到12时,MSE达到最小值,说明并非输入特征数量越多,模型预测效果越好。造成这一现象的原因可能是过多特征引入大量噪声或不相关信息,导致模型训练数据时出现过拟合的现象。

本文选取皮尔逊相关系数(PCC)、方差选择(VTM)和手动选择(EMS)3种常用特征选择方法进行对比验证。在采用PCC和VTM特征选择时,从候选特征集中选择相同特征数量,EMS剔除驾驶员行

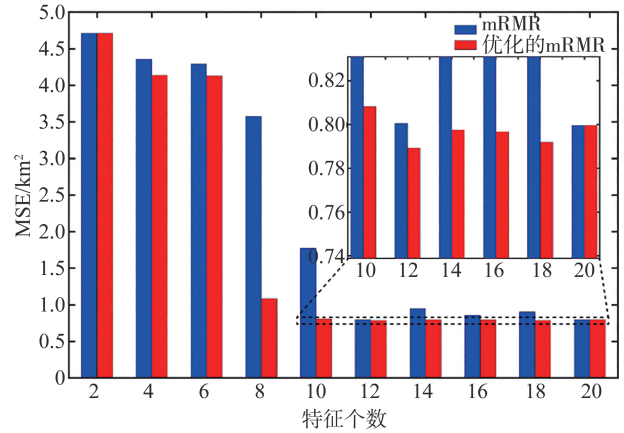


图5 优化前后的mRMR预测误差

为习惯和外部运行环境,代入模型验证得到验证结果,如图6所示。

由图6可知,mRMR特征选择算法的平均均方误差(MSE)、平均绝对误差(MAE)、平均绝对百分比误差(MAPE)和决定系数(R^2)均显著优于其他算法,分别达到0.7983 km、0.6470 km、5.52%和0.9831 km。其中,PCC对数据之间的线性相关性比较敏感,VTM仅选择出方差的特征,部分特征与目标之间毫无关系,EMS仅考虑常见的车辆运行参数。研究结果表明,电动汽车运行过程中影响因素复杂多样,车辆内部参数、驾驶员行为习惯和车辆外部环境等因素对剩余里程有着重要的影响。

因此,通过mRMR特征选择得到输入特征集包含12个特征:电池荷电状态,最低温度值,电池单体电压最高值,电压,电池单体电压最低值,加速踏板行程,气温,车速,制动踏板状态,湿度,电流和最高温度值,如表3所示。

3.2 三层加权Stacking模型结果分析

本文利用BOA优化不同基模型预测效果确定权重分布,如表4所示。由表4可知,基础层中XGBoost、RF和GBDT的权重分别为0.4503、0.3697和0.1720,表明XGBoost、RF和GBDT可以更好地捕捉数据模式,在Stacking模型中担任着重要的预测任务;泛化层中Ridge回归和Lasso回归的权重分别为0.6310和0.3690,原因可能是Lasso回归的L1正则化项并不总能提供明显特征选择优势,很大程度上覆盖数据信息,但仍具有独特价值,可以提供不同拟合特性和正则化效果,进而对整体模型性能的提升起到补充作用。

本文基于三层加权Stacking模型,使用实际运行数据验证模型预测性能、稳定性及泛化能力。本

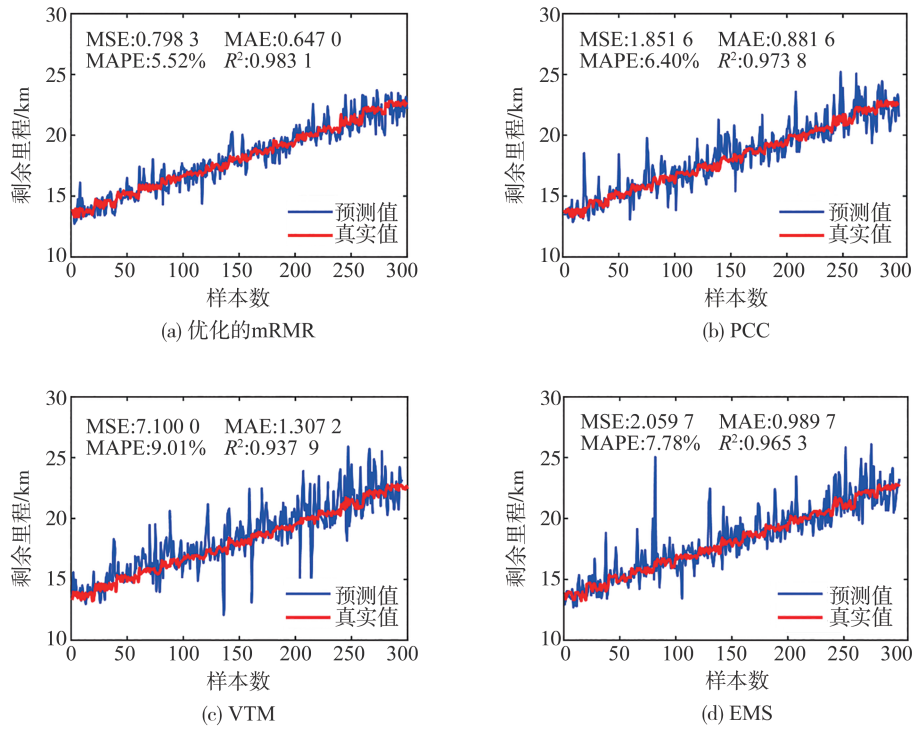


图6 不同特征选择方法预测误差

表3 电动汽车剩余里程预测候选特征集

特征类型	特征名称
车辆内在特性	电池荷电状态、最低温度值、电池单体电压最高值、电压、电池单体电压最低值、车速、电流、最高温度值
驾驶员行为习惯	加速踏板行程、制动踏板状态
外部运行环境	气温、湿度

表4 基础模型在Stacking模型的权重分布

基础层	权重	泛化层	权重
XGBoost	0.450 3	Ridge	0.631 0
RF	0.369 7		
GBDT	0.172 0		
SVM	0.003 6	Lasso	0.369 0
BP	0.004 4		

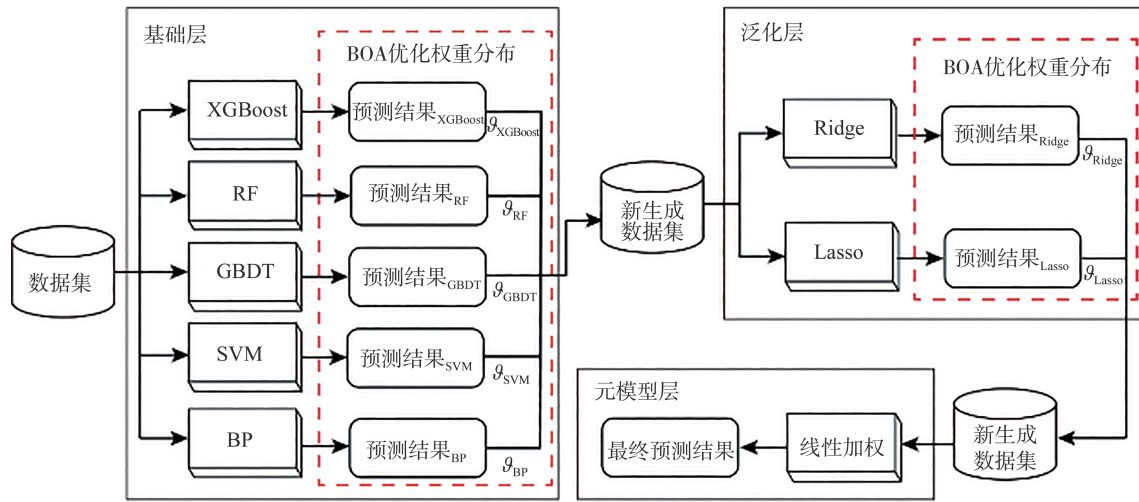
文将三层加权Stacking模型与引入考虑权重分布但不加入原始训练特征的三层Stacking模型、不考虑权重分布且不加入原始训练特征的三层Stacking模型以及两层Stacking模型进行对比,并记为Stacking-A、Stacking-B、Stacking-C和Stacking-D,结构如图7所示,预测结果如表5和图8所示。

由表5和图8可知,Stacking-A模型的正MSE、MAE、MAPE和R²均显著优于其他算法,表明该模型

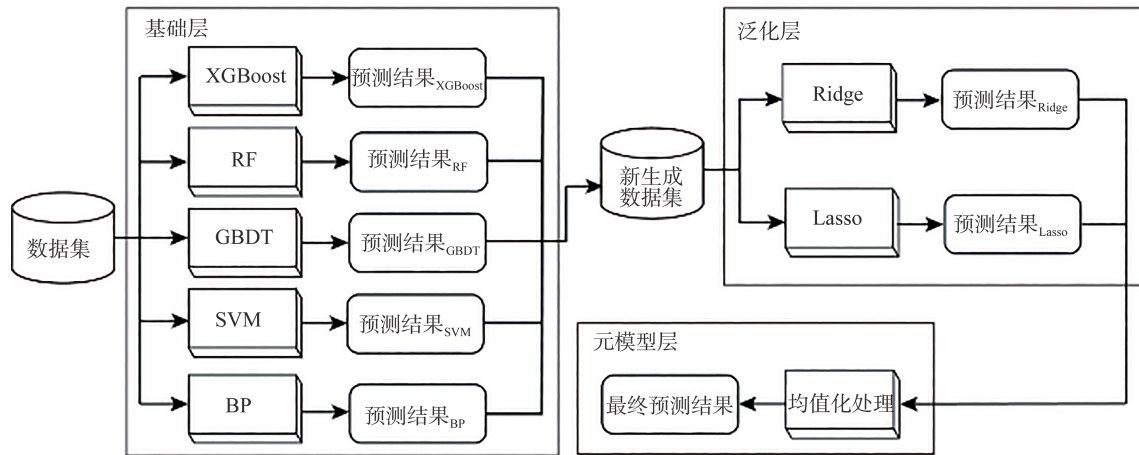
表5 不同Stacking模型预测结果

模型	评价指标			
	MSE	MAE	MAPE	R ²
Stacking-A	0.456 8	0.493 2	4.34%	0.985 2
Stacking-B	0.470 2	0.501 0	4.42%	0.985 0
Stacking-C	0.524 7	0.532 6	4.68%	0.984 4
Stacking-D	0.625 6	0.577 1	5.28%	0.983 8

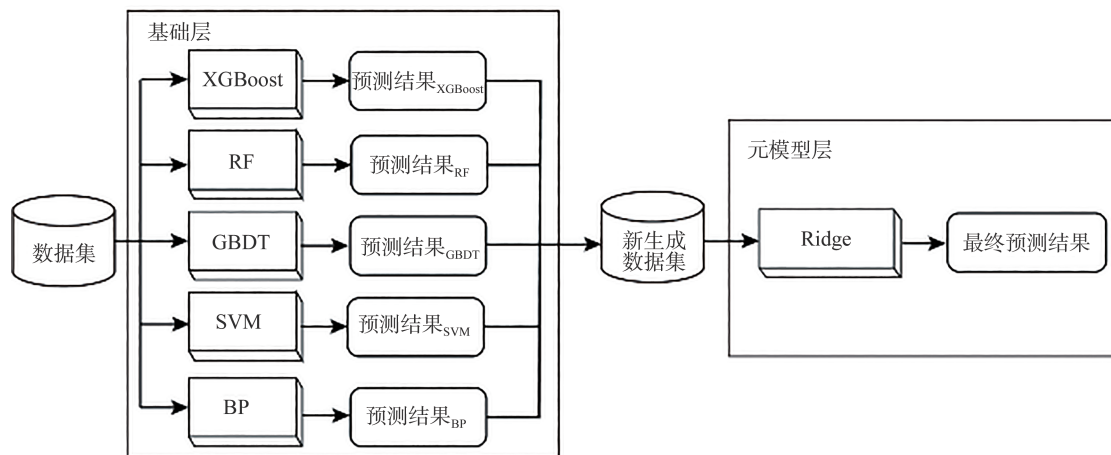
对电动汽车剩余里程预测具有良好性能,原始训练数据引入可以为Stacking模型提供更为全面的特征信息,有助于提高模型的泛化能力。此外,三层加权Stacking模型能够更好地捕捉数据关系,获得更好的预测结果。Stacking-B模型的正MSE和MAPE略高于Stacking-A模型,表明去除原始训练数据对Stacking模型产生负面影响,失去一些关键特征信息,降低模型预测性能。Stacking-C模型的正MSE和MAPE略低于Stacking-B模型,但高于Stacking-D模型,表明不考虑基础模型贡献大小无法充分适应数据特征,但额外层次引入能够使模型具有更强的学习能力,可以更好地适应不同数据分布,减轻过拟合。此外,与mRMR特征选择时只考虑XGBoost模型和其他Stacking模型相比,模型采用层次化结构、原始训练集信息和权重策略可以有效提高整体性能。



(a) Stacking-B



(b) Stacking-C



(c) Stacking-D

图7 Stacking-B、Stacking-C和Stacking-D模型结构

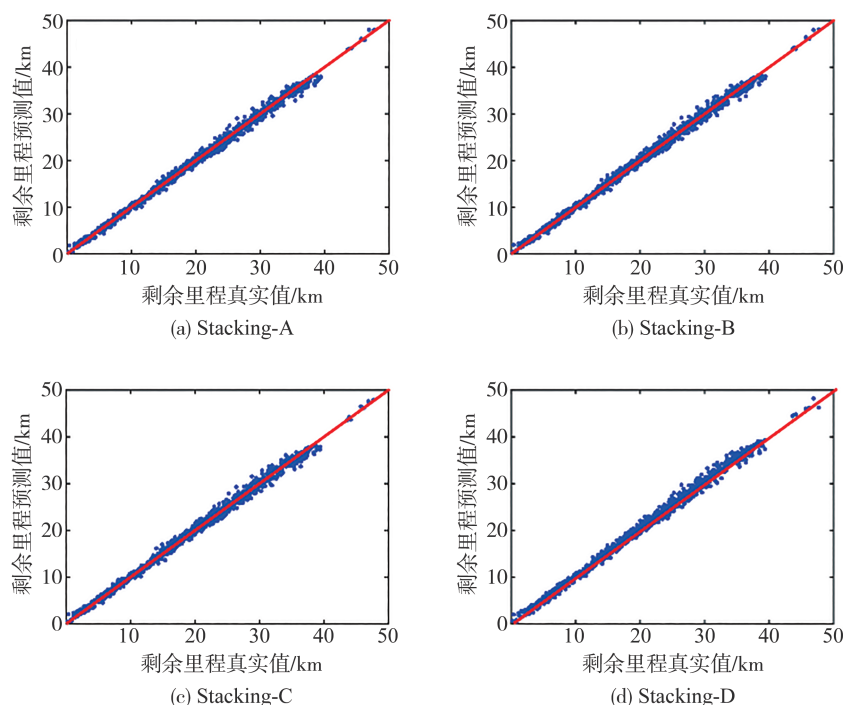


图8 不同 Stacking 模型预测结果对比

4 结论

本文基于新能源汽车监管平台 10 辆电动汽车一年的实际运行数据,使用最小冗余最大相关算法从候选特征集中得到包含 12 个特征的输入特征集,并利用三层加权堆叠模型实现电动汽车剩余里程预测。结果表明:与不考虑层次化结构、原始训练集和权重策略相比,所述三层加权堆叠模型的预测精度较高,泛化能力更强,均方根误差、平均绝对误差、平均绝对百分比误差和决定系数分别达到 0.456 8、0.493 2、4.34% 和 0.985 2。

参考文献

- [1] 卢春房,张航,陈明玉. 新时代背景下的交通运输高质量发展[J]. 中国公路学报, 2021, 34(6): 1-9.
LU C F, ZHANG H, CHENG M Y. Realization of high-quality development of transportation in the new era [J]. China Journal of Highway and Transport, 2021, 34(6): 1-9.
- [2] ZHOU N, PRICE L, YANDE D, et al. A roadmap for China to peak carbon dioxide emissions and achieve a 20% share of non-fossil fuels in primary energy by 2030 [J]. Applied Energy, 2019, 239: 793-819.
- [3] SHU X, YANG W, GUO Y, et al. A reliability study of electric vehicle battery from the perspective of power supply system [J]. Journal of Power Sources, 2020, 451: 227805.
- [4] WANG C Y, ZHANG G, GE S, et al. Lithium-ion battery structure that self-heats at low temperatures [J]. Nature, 2016, 529 (7587): 515-518.
- [5] WANG R, ZHAO X, WANG W, et al. What factors affect the public acceptance of new energy vehicles in underdeveloped regions? a case study of Gansu province, China [J]. Journal of Cleaner Production, 2021, 318: 128432.
- [6] 陈德海,任永昌,黄艳国,等. 基于改进 PSO-RBF 算法的纯电动汽车剩余里程实时预测[J]. 汽车工程, 2018, 40(7): 764-769.
CHENG D H, REN Y C, HUANG Y G, et al. Real time prediction for remaining mileage of battery electric vehicle based on modified PSO-RBF algorithm [J]. Automotive Engineering, 2018, 40(7): 764-769.
- [7] 田慧欣,李晓宇,刘芳. 基于地图信息和循环 SVR 模型的纯电动汽车续驶里程预测[J]. 汽车工程, 2020, 42(9): 1174-1182.
TIAN H X, LI X Y, LIU F. Prediction of continued driving range of battery electric vehicle based on map information and cyclic SVR model [J]. Automotive Engineering, 2020, 42(9): 1174-1182.
- [8] YAVASOGLU H A, TETIK Y E, GOKCE K. Implementation of machine learning based real time range estimation method without destination knowledge for BEVs [J]. Energy, 2019, 172: 1179-1186.

- [9] AYEVIDE F K, KELOUWANI S, AMAMOU A, et al. Estimation of a battery electric vehicle output power and remaining driving range under subfreezing conditions[J]. *Journal of Energy Storage*, 2022, 55: 105554.
- [10] BI J, WANG Y, SAI Q, et al. Estimating remaining driving range of battery electric vehicles based on real-world data: a case study of Beijing, China[J]. *Energy*, 2019, 169: 833-843.
- [11] WEI H, HE C, LI J, et al. Online estimation of driving range for battery electric vehicles based on SOC-segmented actual driving cycle[J]. *Journal of Energy Storage*, 2022, 49: 104091.
- [12] LIPU M S H, HANNAN M A, HUSSAIN A, et al. Real-time state of charge estimation of lithium-ion batteries using optimized random forest regression algorithm[J]. *IEEE Transactions on Intelligent Vehicles*, 2022, 8(1): 639-648.
- [13] HUANG H, MENG J, WANG Y, et al. An enhanced data-driven model for lithium-ion battery state-of-health estimation with optimized features and prior knowledge[J]. *Automotive Innovation*, 2022, 5(2): 134-145.
- [14] WU X, LI M, DU J, et al. SOC prediction method based on battery pack aging and consistency deviation of thermoelectric characteristics[J]. *Energy Reports*, 2022, 8: 2262-2272.
- [15] ZHONG W, CHEN X, WU Q, et al. Selection of diverse features with a diverse regularization[J]. *Pattern Recognition*, 2021, 120: 108154.
- [16] 胡杰, 翁灵隆, 覃雄臻, 等. 基于多模型融合的电动汽车行驶里程预测[J]. *交通运输系统工程与信息*, 2020, 20(5): 100-106, 141.
HU J, WENG L L, QIN X Z, et al. Mileage prediction of electric vehicle based on multi model fusion[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2020, 20(5): 100-106, 141.
- [17] LI X, HUA W, WU C, et al. State estimation of a lithium-ion battery based on multi-feature indicators of ultrasonic guided waves[J]. *Journal of Energy Storage*, 2022, 56: 106113.
- [18] ZOU D, MENG D, DAI Y, et al. Optimal charging strategy of electric vehicles with consideration of battery storage[C]. *E3S Web of Conferences*. EDP Sciences, 2021, 236: 02015.
- [19] ANDRE D, MEILER M, STEINER K, et al. Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. I. Experimental investigation[J]. *Journal of Power Sources*, 2011, 196(12): 5334-5341.
- [20] HE H, XIONG R, GUO H. Online estimation of model parameters and state-of-charge of LiFePO₄ batteries in electric vehicles[J]. *Applied Energy*, 2012, 89(1): 413-420.
- [21] PENG H, LONG F, DING C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(8): 1226-1238.
- [22] WANG X, JIN Y, SCHMITT S, et al. Recent advances in Bayesian optimization[J]. *ACM Computing Surveys*, 2023, 55(13s): 1-36.
- [23] NAIMI A I, BALZER L B. Stacked generalization: an introduction to super learning[J]. *European Journal of Epidemiology*, 2018, 33: 459-464.
- [24] CHEN T, GUESTRIN C. Xgboost: a scalable tree boosting system[C]. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 2016: 785-794.
- [25] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, 45: 5-32.
- [26] FU F, JIANG J, SHAO Y, et al. An experimental evaluation of large scale GBDT systems[J]. *arXiv preprint arXiv:1907.01882*, 2019.
- [27] CRISTIANINI N, SHAWE-TAYLOR J. An introduction to support vector machines and other kernel-based learning methods[M]. Cambridge University Press, 2000.
- [28] DING S, SU C, YU J. An optimizing BP neural network algorithm based on genetic algorithm[J]. *Artificial Intelligence Review*, 2011, 36: 153-162.
- [29] SALEH A K M E, ARASHI M, KIBRIA B M G. Theory of ridge regression estimation with applications[M]. John Wiley & Sons, 2019.
- [30] LEE J H, SHI Z, GAO Z. On LASSO for predictive regression[J]. *Journal of Econometrics*, 2022, 229(2): 322-349.