文章编号:1674-2974(2025)05-0057-09 DOI:10.16339/j.cnki.hdxbzkb.2025047

基于改进Stacking多模型融合的高速公路隧道建设碳排放预测模型

吴佳润[1],林宇亮[1,3†],邢浩[2],宁曦[2]
(1. 中南大学 土木工程学院,湖南 长沙 410075;
2. 中铁七局集团第四工程有限公司,湖北 武汉 430074;
3. 中南大学 高速铁路建造技术国家工程研究中心,湖南 长沙 410075)

摘 要:隧道工程作为交通基础设施的重要组成部分,其建设过程中的碳排放问题日益受到关注.构建合理的隧道建设碳排放预测模型为控制隧道工程碳排放量、实现隧道减碳设计提供重要的科学依据.据此,以勐绿高速公路隧道工程建设为依托,考虑围岩级别、隧道总长度等12个特征参数,建立了120个隧道不同衬砌设计每延米隧道建设碳排放样本.在传统Stacking算法的基础上,提出了一种改进Stacking多模型融合的隧道建设碳排放预测方法.通过残差赋权方式组合基学习器交叉验证得到的各训练模型,降低了对噪声的敏感性.并以原始训练集和组合基学习器的预测结果作为元学习器输入,保留了原始数据集的信息.研究结果表明,改进Stacking算法预测碳排放量与实际碳排放量的均方根误差(E_{RMSE})、平均绝对误差(E_{MAE})以及决定系数(R^2)不仅优于3种单一基学习器,也优于传统Stacking算法.因此,推荐将改进Stacking算法用于隧道建设碳排放预测.

关键词:隧道建设;碳排放;单一基学习器;Stacking模型融合;预测模型
中图分类号:U455 **文献标志码**:A

A Carbon Emission Prediction Model for Highway Tunnel Construction Based on Improved Stacking Multiple Model Integration

WU Jiarun[1], LIN Yuliang[1,3†], XING Hao[2], NING Xi[2]
(1. School of Civil Engineering, Central South University, Changsha 410075, China;
2. The Fourth Engineering Co., Ltd., the Seventh Bureau Group of China Railway, Wuhan 430074, China;
3. National Engineering Research Center of High-speed Railway Construction Technology, Central South University, Changsha 410075, China)

Abstract: Tunnel engineering, as a significant component of transportation infrastructure, has increasingly received attention for its carbon emissions during the construction process. It is a significant scientific basis for controlling carbon emissions and achieving carbon reduction in tunnel design by establishing a rational model for predicting carbon emissions in tunnel construction. Accordingly, the dataset concludes a total of 120 samples of

* **收稿日期**:2024-05-21
基金项目:国家自然科学基金资助项目(51878667), National Natural Science Foundation of China(51878667);湖南省自然科学基金资助项目(2021JJ30830), Natural Science Foundation of Hunan Province(2021JJ30830)
作者简介:吴佳润(2000—),男,湖南常德人,中南大学博士研究生
† 通信联系人,E-mail:linyuliang11@csu.edu.cn

tunnel construction carbon emissions (per meter) with different lining designs based on the construction of the Menglü highway tunnel project, considering 12 characteristic parameters, including the surrounding rock level, the total length of the tunnel, and so on. Based on the traditional Stacking algorithm, this study proposes an improved Stacking algorithm of multi-model fusion to predict carbon emissions in tunnel construction. The improved Stacking algorithm combines the various base learner training models with the residual weighting approach, which is obtained through cross-validation. Besides, the improved Stacking algorithm utilizes the original training set and the prediction results of the combined base learner as the meta-learner inputs. Therefore, the improved Stacking algorithm is not only less sensitive to noise, but also retains the original dataset information. The results demonstrate that the improved Stacking algorithm is superior to three single base learners as well as the traditional Stacking algorithm in terms of root mean square error (E_{RMSE}), mean absolute error (E_{MAE}), and determination coefficient (R^2). Consequently, it recommends the improved Stacking algorithm for predicting carbon emissions in tunnel construction.

Key words: tunnel construction; carbon emissions; single base learner; stacking model integration; prediction model

随着近些年来大规模基础设施建设的开展,我国已跃居成为世界上隧道数量第一、总长度第一的国家[1]. 根据交通运输部发布的《2022年交通运输行业发展统计公报》,截至2022年底,我国公路隧道共计24 850座、2 678.43万延米,增加了1 582座、208.54万延米. 有研究表明,隧道是能量和物质最集中的基础设施[2],在建设过程中会向环境释放大量的温室气体. 在我国"双碳"目标背景下,开展隧道建设温室气体排放研究显得十分重要.

生命周期评价(life cycle assessment,LCA)方法是一种评估产品或设施整个生命周期内对环境影响程度的手段,已经广泛应用于隧道碳排放研究领域[3-4]. Guo等[5]基于LCA方法评估了中国某公路隧道施工阶段、运营阶段和维护阶段的温室气体排放量,分别占总碳排放量的30.89%、56.12%和12.99%. 苏征宇等[6]利用数字化技术建立了公路隧道碳排放计算模型,并对隧道各阶段的碳排放量进行分析,认为建材生产的碳排放是总碳排放量的主要来源. Xu等[7-8]考虑地质环境和施工参数对隧道碳排放的影响,采用LCA方法计算了8条隧道的49种衬砌设计的碳排放量,建立了两种隧道碳排放的线性回归模型.

机器学习是目前盛行的一种人工智能技术[9],能够通过对历史数据的学习实现较高精度的预测. 刘凯华等[10]基于机器学习算法构建了再生混凝土耐久性的预测模型,研究结果表明,梯度提升树能够较好地预测再生混凝土的抗氯离子侵蚀性能,高斯过程回归模型可以较好地预测抗碳化性能. 齐婉婉等[11]采用XGBoost模型对地震峰值加速度进行预测,并与实测记录和一维数值计算结果对比,结果表明XGBoost模型能够较好地预测地震峰值加速度,且预测结果稳定. Zhang等[12]利用梯度增强回归树算法构建软黏土的压缩模量的预测模型,并采用遗传算法对梯度增强回归树模型的超参数进行调整,预测结果与实际结果的均方误差和相关系数分别为0.13和0.91. 陈嘉等[13]提出一种基于VGG的浅层深度卷积模型的路面破损分类识别方法,对于小规模图像的分类识别具有显著优势,且比其他方法结果更精确.

机器学习的各种算法已在建筑行业得到了广泛的应用,但在隧道建设碳排放分析方面的报道并不多见. 本文以云南省勐绿高速公路为工程依托,建立隧道建设碳排放样本,提出一种改进Stacking算法用于隧道建设碳排放的预测,以XGBoost、SVR和Ridge为基学习器,对隧道建设碳排放样本进行预测,并与单一基学习器和传统Stacking算法的预测结果进行对比分析.

1 隧道建设碳排放样本构建

勐绿高速公路工程位于中国云南省勐腊县以及江城县境内,主要包含曼滩隧道、那萨箐隧道、石楼

梯山隧道、莫作山隧道、景嚷山隧道以及阿蔑寨隧道6座隧道，如表1所示. 每座隧道包含多种衬砌设计，如：那萨箐隧道为分离式隧道，包含F51、F52、F55和F47四种衬砌断面设计，各衬砌设计种类的围岩级别、开挖面积、材料用量均有所不同，导致隧道建设的碳排放量不同. 石楼梯山隧道、莫作山隧道、景嚷山隧道和阿蔑寨隧道也属于分离式隧道，涉及的衬砌断面设计形式有F53、F54、F41、F42、F43、F44、F45、F46、F31、F32、F33. 曼滩隧道为连拱式隧道，包含5种连拱式衬砌设计，分别为L51、L52、L53、L41、L42. 因此勐绿高速公路隧道工程总共包含20种不同的衬砌设计.

表1　勐绿高速公路6座隧道的基本参数

Tab.1　The basic parameters of the six tunnels of the Menglü Expressway

隧道名称	隧道长度(左线/右线)/m	净宽/m	净高/m	结构形式	衬砌设计种类
曼滩隧道	345	10.25	5	连拱式	L51、L52、L53、L41、L42
那萨箐隧道	479/384	10.25	5	分离式	F51、F52、F55、F47
石楼梯山隧道	177/216	10.25	5	分离式	F51、F52、F53、F42
莫作山隧道	4 305/4 308	10.25	5	分离式	F51、F52、F42、F43、F44、F45、F31、F32、F33
景嚷山隧道	651/657	10.25	5	分离式	F51、F52、F54、F41、F31
阿蔑寨隧道	641/736	10.25	5	分离式	F51、F52、F43、F46

注：衬砌设计种类的第1个字母表示隧道断面形式，“L”指连拱式隧道，“F”指分离式隧道；第2个数字表示隧道围岩级别；第3个数字代表隧道衬砌设计的类别.

隧道总长度是影响隧道建设碳排放的重要因素之一. 长隧道建设往往难度更大，每延米隧道建设的材料消耗量和机械耗能会随之增加，导致每延米隧道建设的碳排放量增加. 我国的《公路工程预算定额》(JTG/T 3832—2018)对不同隧道总长度中的材料成本、机械耗能等进行定额. 设隧道总长度为L，将隧道总长度分为$L\leqslant$1 km、1 km$<L\leqslant$2 km、2 km$<L\leqslant$3 km、3 km$<L\leqslant$4 km、4 km$<L\leqslant$5 km以及5 km$<L\leqslant$6 km共6种情况，可据此对不同总长度隧道建设的材料消耗量、机械耗能等进行估算. 结合工程中涵盖的20种衬砌设计类型，考虑前述6种不同隧道长度的情况，可以建立得到隧道工程建设120个不同衬砌设计的工况.

隧道建设碳排放计算依据《建筑碳排放计算标准》(GB/T 51366—2019)，分为材料生产阶段、材料运输阶段及建造阶段3个阶段. 材料生产阶段的碳排放需要预估材料的消耗量，材料消耗量根据设计资料和《公路工程预算定额》(JTG/T 3832—2018)进行确定；材料运输阶段的碳排放主要为材料运输时运输机械所消耗的能源，且考虑运输距离的影响；建造阶段的碳排放主要依据为建造过程中施工机械的能源消耗，施工机械耗能根据《公路工程预算定额》(JTG/T 3832—2018)和《全国统一施工机械台班费用定额》确定. 隧道建设碳排放的计算流程如图1所示. 因此，结合上述120个不同衬砌设计的工况，可确定每延米隧道建设的碳排放量. 在隧道建设碳排放样本分析时，主要考虑了围岩级别、隧道总长度、结构形式和开挖面积等对材料消耗量与机械耗能影响显著的因素[14-15]，同时也考虑了隧道建设过程中消耗的主要材料与能源，包括锚杆、混凝土、钢架、防水材料、水泥、燃油以及电能. 限于篇幅，这里仅列出20个隧道建设碳排放样本，如表2所示.

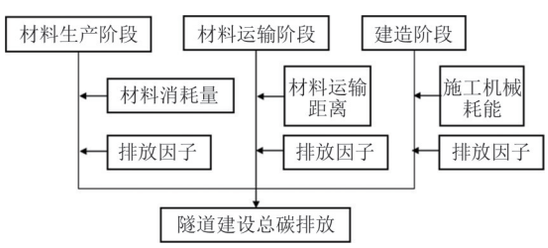


图1　隧道建设碳排放计算流程图

Fig.1　Flowchart of carbon emission calculation for tunnel construction

2　改进Stacking隧道建设碳排放模型

2.1　基学习器选取

2.1.1　XGBoost算法

极值梯度提升(extreme gradient boosting，XGBoost)是基于梯度提升树的一种优化算法，其目标函数由损失函数和正则化项两部分组成，在优化目标函数时，将损失函数二阶泰勒展开，去除常数项，再将正则化项展开，去除常数项，合并一次项、二次项系数，得到最终目标函数[16]，其表达式如式(1)所示.

表2 每延米隧道建设碳排放样本

Tab.2 Carbon emission samples per meter of tunnel construction

编号	结构形式	围岩级别	隧道总长度/km	开挖面积/m^2	锚杆/m	混凝土/m^3	钢架/t	钢筋/t	防水材料/m^2	水泥/t	燃油/kg	电能/(MW·h)	总碳排放量/(tCO_{2eq})
1	连拱式	Ⅴ级	0~1	219.76	84.84	72.07	1.96	3.94	117.61	44.12	148.82	5.41	75.58
2	分离式	Ⅳ级	3<~4	127.68	112.11	40.36	0.77	2.33	92.6	25.64	115.95	3.91	44.42
3	分离式	Ⅴ级	4<~5	207.83	87.12	60.7	1.73	3.05	115.85	35.81	219.34	5.93	64.89
4	分离式	Ⅲ级	2<~3	167.12	0	27.78	0.89	2.38	113.35	17.16	158.99	3.34	34.3
5	连拱式	Ⅴ级	1<~2	211.82	70.7	65.69	1.67	4.15	116.21	40.14	155.6	5.15	69.88
6	连拱式	Ⅳ级	2<~3	201.16	0	54.48	0.9	3.14	115.51	34.43	150.44	4.23	56.11
7	分离式	Ⅳ级	5<~6	127.68	112.11	40.36	0.77	2.33	92.69	25.64	139.6	4.08	44.68
8	分离式	Ⅴ级	1<~2	207.66	135.5	60.7	1.73	3.85	115.85	35.9	154.68	5.78	67.14
9	连拱式	Ⅴ级	2<~3	219.76	70.7	72.07	1.96	3.95	117.61	43.88	178.21	5.73	75.62
10	分离式	Ⅴ级	0~1	207.83	87.12	60.7	1.73	3.05	115.85	35.81	138.96	5.21	63.87
11	分离式	Ⅴ级	0~1	292.84	147.3	91.02	2.83	4.78	133.43	52.1	201.84	7.66	95.17
12	分离式	Ⅴ级	5<~6	207.66	135.5	60.7	1.73	3.85	115.85	35.9	231.64	6.51	68.16
13	分离式	Ⅳ级	3<~4	127.68	112.11	40.36	0.77	2.33	92.69	25.64	115.95	3.91	44.42
14	连拱式	Ⅴ级	5<~6	211.82	70.7	65.69	1.67	4.15	116.21	40.14	234.11	5.88	70.91
15	分离式	Ⅲ级	2<~3	169.22	0	30.66	0.92	2.54	114.32	18.52	161.51	3.63	37.08
16	分离式	Ⅴ级	1<~2	216.50	154.87	68.17	2.03	3.96	117.26	40.18	165.63	6.82	76.09
17	分离式	Ⅳ级	3<~4	199.82	0	54.43	1.71	3.05	114.32	32.07	185.71	5.08	58.15
18	连拱式	Ⅴ级	4<~5	219.76	84.84	72.07	1.96	3.94	117.61	44.12	233.82	6.13	76.64
19	分离式	Ⅲ级	1<~2	167.12	0	27.78	0.89	2.38	113.35	17.16	145.51	2.94	33.87
20	分离式	Ⅳ级	4<~5	127.68	112.11	40.36	0.77	2.33	92.69	25.64	132.4	4.00	44.57

$$f_{\text{obj}} = \sum_{j=1}^{T}\left[G_j w_j + \frac{1}{2}\left(H_j + \lambda\right)w_j^2\right] + \gamma T \tag{1}$$

式中:T为一棵树上的叶子总数;G_j、H_j分别为叶子结点j所包含样本的一阶偏导数累加之和和二阶偏导数的累加之和;w_j为叶子结点j的权重;γ、λ为惩罚系数.

将式(1)写成关于叶子结点权重w_j的一元二次函数求解得到最优的权重w_j^*和目标函数f_{obj}^*公式:

$$w_j^* = -\frac{G_j}{H_j + \lambda} \tag{2}$$

$$f_{\text{obj}}^* = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j + \lambda} + \gamma T \tag{3}$$

2.1.2 SVR算法

支持向量回归(support vector regression,SVR)的基本思想是寻找一个超平面,使所有训练样本离该超平面的误差最小,为减小样本点的边界误差,根据经验在回归函数两侧定义一个间距为e的间隔带,只计算落在间隔带外侧的样本点的损失.引入松弛变量以及惩罚系数C,以最小化总损失和最大化间隔优化拟合问题,并利用拉格朗日函数将其转化为对偶问题形式[17],如式(4)所示.

$$\max_{\alpha,\alpha^*}\left[-\frac{1}{2}\sum_{i,j=1}^{n}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)\left(x_i x_j\right) - e\sum_{i=1}^{n}\left(\alpha_i + \alpha_i^*\right) + y_i\sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)\right] \tag{4}$$

$$\text{s.t.}\begin{cases}\sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right) = 0\\ 0 \leqslant \alpha_i, \alpha_i^* \leqslant C, i = 1, 2, \cdots, n\end{cases}$$

式中:α_i、α_i^*为拉格朗日算子.

由式(4)求出α_i和α_i^*,根据泛函相关理论,引入满足Mercer条件的核函数$K(x_i, x_j)$,得到回归函数[18],如式(5)所示.

$$f(x) = \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)K\left(x_i - x_j\right) + b \tag{5}$$

2.1.3 Ridge算法

岭回归(Ridge regression)实质上是一种改良的最小二乘估计法,通过放弃最小二乘法的无偏性,以损失部分信息、降低精度为代价获得回归系数更为符合实际、更可靠的回归方法,对病态数据的拟合要强于最小二乘法[19].如式(6)所示,在原先的代价函数$J_\beta(\beta)$的最小二乘估计中加入对系数值的惩罚,将原先无法求广义逆的情况变成可以求出其广义逆,使得问题稳定并得以求解.

$$J_{\beta}(\beta)=\arg\min_{\beta}\left\{\left[\sum_{i=1}^{N}\left(y_i-\beta_0-\sum_{j=1}^{p}x_{ij}\beta_j\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2\right]\right\} \tag{6}$$

式中：β为回归参数；λ为岭参数（惩罚系数）；β_j为总计p个特征中第j个特征的系数.

2.2　改进Stacking集成算法

传统Stacking集成算法是一个两层的模型结构，第一层由多种不同的算法组成，这些算法也被称为基学习器，第二层只由一种算法构成，该算法称为元学习器[20]，基本框架如图2所示. 在训练阶段，通常运用K折交叉验证的思想将训练集划分为多个子数据集，将各基学习器K折交叉验证的预测结果输入第二层进行元学习器的训练. 在测试阶段，各基学习器在K折交叉验证时建立的模型对测试集进行预测，将同一基学习器不同模型的预测结果取均值输入元学习器得到最终的测试结果.

传统Stacking算法的元学习器仅采用基学习器的预测结果进行训练，这可能导致一定程度的数据信息泄露问题[21]，且在组合多个模型预测结果时也容易对异常值敏感，这均会导致预测结果与实际结果误差较大. 因此，为提高传统传统Stacking算法的预测准确性，本文主要做了以下几点改进：1）基学习器交叉验证得到的各训练模型通过残差赋权的方式组合，将各组合基学习器对原始训练集的预测结果作为元学习器输入的一部分；2）将原始训练集数据作为元学习器输入的一部分；3）元学习器测试时，采用原始测试集和各组合基学习器对原始测试集的预测结果作为元学习器的输入. 以XGBoost、SVR和Ridge为基学习器，改进Stacking算法的隧道建设碳排放预测流程如图3所示.

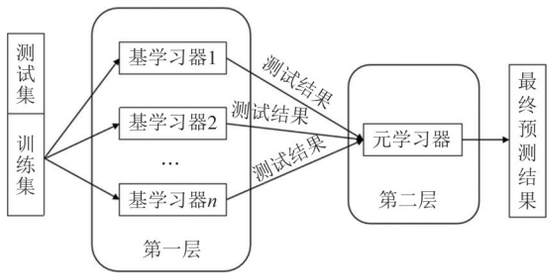


图2　Stacking集成算法基本框架

Fig.2　The basic framework of Stacking integration algorithms

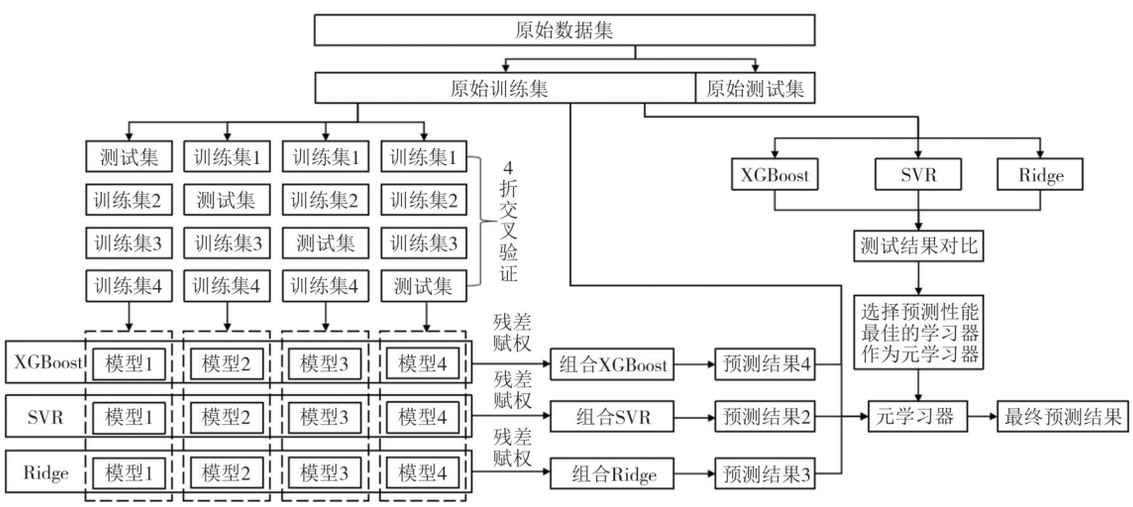


图3　改进Stacking算法的预测流程图

Fig.3　Prediction flowchart of the improved Stacking algorithm

1）对原始数据集进行归一化处理，结果如表3所示，并将原始数据集随机划分为原始训练集和原始测试集，利用原始训练集对XGBoost、SVR和Ridge三种单一基学习器进行训练，选择其中误差最小的基学习器作为元学习器.

2）进行4折交叉验证，将原始训练集划分为4份，依次选择其中一份作为测试集，剩余3份作为训练集输入各基学习器中，因此对于三种基学习器均会训练得到4个模型.

3）将原始训练集输入各基学习器的4个模型，计算得到各基学习器的4个模型的残差，同一基学习器的4个模型通过残差赋权组合得到该基学习器的组合基学习器.

4）将各组合基学习器对原始训练集的预测结果

以及原始训练集输入元学习器中进行训练,最后通过将各组合基学习器对原始测试集的预测结果以及原始测试集输入元学习器得到最终预测结果.

表3 输入特征赋值及归一化处理结果

Tab.3 Input feature assignments and normalization processing results

输入参数	样本特征	赋值	归一化结果
结构形式	连拱式	1	0
	分离式	2	1
围岩级别	Ⅲ级	3	0
	Ⅳ级	4	0.5
	Ⅴ级	5	1
隧道总长度	0~1 km	1	0
	1<~2 km	2	0.2
	2<~3 km	3	0.4
	3<~4 km	4	0.6
	4<~5 km	5	0.8
	5<~6 km	6	1
开挖面积	单位为m^2的数值	原数值	$(x-127.68)/165.16$
总碳排放量	单位为tCO_{2eq}的数值	原数值	$(x-33.68)/62.85$

2.3 预测结果评价方法

均方根误差(E_{RMSE})、平均绝对误差(E_{MAE})以及决定系数(R^2)可用来衡量模型的预测性能.其中E_{RMSE}为预测值与实际值的平方差的期望值的平方根,E_{RMSE}越小,模型越稳定;E_{MAE}为绝对误差的平均值,E_{MAE}越小,反映模型的精确度越高;R^2为回归平方和与总平方和的比值,R^2越接近1,反映模型的拟合效果越好.三者的计算公式如式(7)~式(9)所示.

$$E_{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2} \tag{7}$$

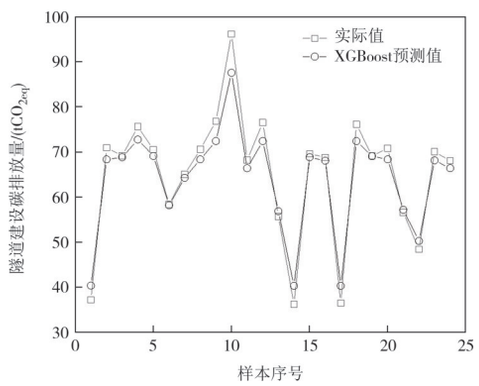
$$E_{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \tag{8}$$

$$R^2=1-\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}_i\right)^2} \tag{9}$$

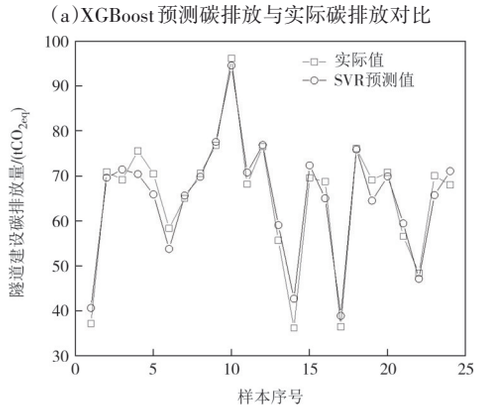
式中:y_i为隧道建设碳排放的实际值;$\hat{y}_i$为隧道建设碳排放的实际值;$\bar{y}_i$为隧道建设碳排放的平均值;n为样本数量.

3 单一基学习器预测结果

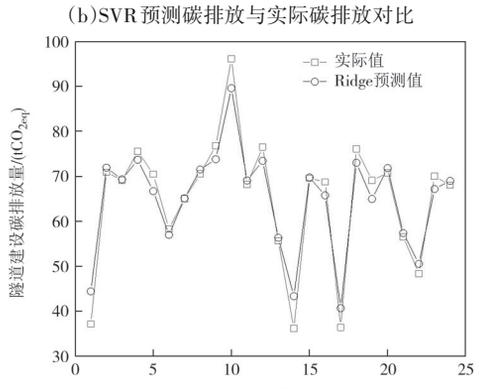
基学习器面向的是更高维的数据,需要选择泛化能力较强的学习器作为Stacking算法的基学习器.将120个隧道建设碳排放样本按4∶1的比例随机划分为原始训练集和原始测试集,将原始训练集输入XGBoost、SVR和Ridge基学习器中进行学习,结果如图4所示,并计算原始测试集的均方根误差(E_{RMSE})、平均绝对误差(E_{MAE})以及决定系数(R^2),如表4所示.可以看出,3种单一基学习器的均方根误差(E_{RMSE})和平均绝对误差(E_{MAE})小于3.277 0和2.693 0,且决定系数(R^2)大于0.944 1,说明3种单一基学习器均可以得到较为准确的隧道建设碳排放预测结果,其中XGBoost的拟合效果最优,均方根误差(E_{RMSE})、平均绝对误差(E_{MAE})以及决定系数(R^2)分别为2.957 0、2.302 2和0.954 5.



(a)XGBoost预测碳排放与实际碳排放对比



(b)SVR预测碳排放与实际碳排放对比



(c)Ridge预测碳排放与实际碳排放对比

图4 单一基学习器预测碳排放与实际碳排放对比

Fig.4 Comparison between the predicted carbon emissions and the actual carbon emissions of single base learner predict

表4　单一基学习器的评价指标

Tab.4　Evaluation indicators for single base learners

基学习器	均方根误差(E_{RMSE})	平均绝对误差(E_{MAE})	决定系数(R^2)
XGBoost	2.957 0	2.302 2	0.954 5
SVR	3.160 3	2.693 0	0.948 0
Ridge	3.277 0	2.517 5	0.944 1

基学习器差异性的大小对Stacking算法最终的预测性能有重要影响,选择本质不同的学习器可以在一定程度上减小误差[22]. 因此,计算3种基学习器对隧道建设碳排放预测误差的Pearson相关系数,相关性分析结果如表5所示. 可以看出3种学习器的误差相关性低于0.786 8,学习器之间具有一定的差异性,说明XGBoost、SVR和Ridge三种学习器作为Stacking集成算法的基学习器是合适的.

表5　基学习器的预测误差相关性分析

Tab.5　Prediction error correlation analysis of base learners

基学习器	XGBoost	SVR	Ridge
XGBoost	1	0.408 0	0.786 8
SVR	0.408 0	1	0.687 2
Ridge	0.786 8	0.687 2	1

4　改进Stacking算法预测结果

根据前述分析,3种单一基学习器中XGBoost的预测碳排放与实际碳排放的误差较小,因此选择XGBoost作为元学习器. 以XGBoost、SVR和Ridge为基学习器,XGBoost为元学习器,利用原始训练集训练改进Stacking模型,并对测试集数据进行预测,测试结果如图5所示. 并可由此分析预测结果与实际结果的相关关系,如图6所示. 可以看出,改进Stacking算法的预测值与实际值的决定系数(R^2)为0.981 7,说明改进Stacking算法的隧道建设碳排放预测准确度是可以保证的,推荐用于隧道建设碳排放预测.

为进一步验证改进Stacking算法的预测准确度,将改进Stacking算法的均方根误差(E_{RMSE})、平均绝对误差(E_{MAE})以及决定系数(R^2)与传统Stacking算法、XGBoost、SVR和Ridge进行对比,如图7所示. 可以看出,5种模型的隧道建设碳排放预测结果的精度由高到低排序依次为:改进Stacking>传统Stacking>XGBoost>SVR>Ridge. 对比改进Stacking算法与3种单一基学习器,改进Stacking算法的均方根误差

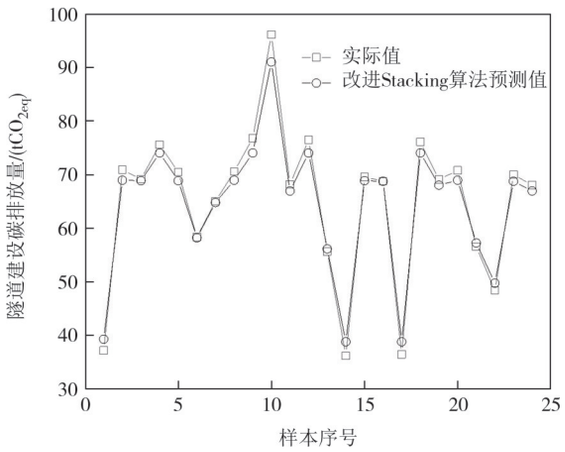


图5　改进Stacking算法预测值与实际值对比

Fig.5　Comparison between the predicted value and the actual value of the improved Stacking algorithm

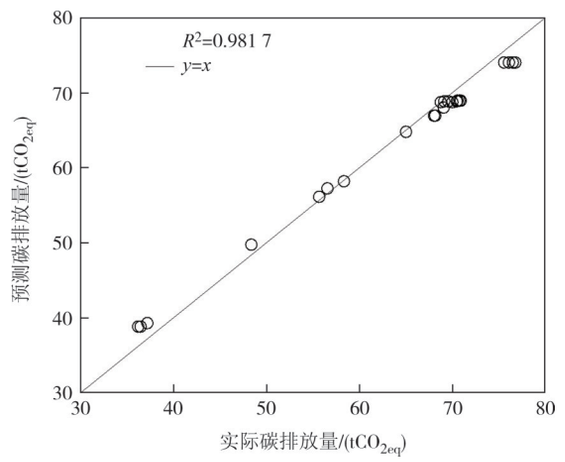


图6　改进Stacking算法预测值与实际值的相关关系

Fig.6　Correlation between the predicted value and the actual value of the improved Stacking algorithm

(E_{RMSE})和平均绝对误差(E_{MAE})较单一基学习器中拟合效果最佳的XGBoost分别降低了36.60%和33.73%,决定系数(R^2)提升了2.85%. 对比改进Stacking算法和传统Stacking算法,改进Stacking算法的均方根误差(E_{RMSE})和平均绝对误差(E_{MAE})分别降低了18.83%和21.23%,决定系数(R^2)提升了0.97%. 这说明本文提出的对传统Stacking算法的改进在隧道建设碳排放的预测效果上是有效的,改进Stacking算法更适合用于隧道建设碳排放预测.

改进Stacking算法不仅优于3种单一基学习器,也优于传统Stacking算法,可对其进行如下讨论:

1)传统Stacking算法仅将基学习器的预测结果作为元学习器的输入,这可能丢失了原始数据集中的部分信息,而改进Stacking算法将原始训练集数据作为元学习器输入的一部分,保留了这部分信息.

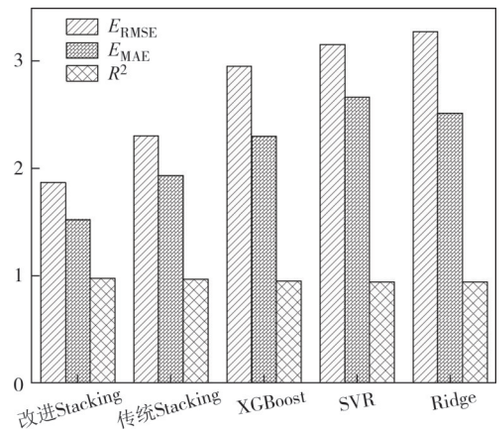


图7 不同模型的评价指标对比

Fig.7 Comparison of evaluation indicators of different models

2)传统Stacking算法将K折交叉验证中的K个测试集的测试结果作为元学习器的输入,但每折的测试数据不同,基学习器在不同折之间的性能也会不一样,可能导致噪声放大,改进Stacking算法通过残差赋权的方式将各基学习器的多个模型进行组合,元学习器在训练时则会更加关注具有较大残差的样本,降低对噪声的敏感性.

3)在测试阶段,传统Stacking算法将各基学习器在K折交叉验证时建立的模型对原始测试集的测试结果取平均值输入元学习器得到最终测试结果,而改进Stacking算法将测试集数据和组合基学习器对原始测试集的预测结果作为元学习器的输入,保留了原始测试集中的信息,降低了对噪声的敏感性,提高模型的预测性能.

5 结 论

1)以勐绿高速公路为工程依托,考虑隧道总长度、围岩级别等12个特征因素,根据生命周期评价方法从隧道建设的材料生产阶段、材料运输阶段和建造阶段3个阶段综合考虑,建立了120个隧道建设的碳排放量样本.

2)采用XGBoost、SVR和Ridge三种单一基学习器对隧道建设碳排放量进行预测,结果表明3种单一基学习器均可以得到较为准确的隧道建设碳排放预测结果,且基学习器间存在着一定的差异性,其中XGBoost基学习器的预测精度最高.

3)将改进Stacking算法的均方根误差、平均绝对误差和决定系数与传统Stacking算法和单一基学习器进行比较,结果表明改进Stacking算法不仅优于3种单一基学习器,还优于传统Stacking算法,对隧道建设碳排放的预测效果更加准确,推荐用于隧道建设碳排放预测.

4)虽然本文的改进方法取得了一定的成果,但仍存在一些不足,样本量对预测准确性具有重大影响,如何获得更多样本是关键问题.基学习器与元学习器的选取对Stacking算法的预测性能具有一定的影响,后续工作可考虑采用多种不同的学习器组合进行预测分析.

参考文献

[1] 《中国公路学报》编辑部. 中国交通隧道工程学术研究综述·2022[J]. 中国公路学报, 2022, 35(4): 1-40.
Editorial Department of China Journal of Highway and Transport. Review on China's traffic tunnel engineering research: 2022[J]. China Journal of Highway and Transport, 2022, 35(4): 1-40. (in Chinese)

[2] MILIUTENKO S, ÅKERMAN J, BJÖRKLUND A. Energy use and greenhouse gas emissions during the life cycle stages of a road tunnel - the Swedish case norra länken[J]. European Journal of Transport and Infrastructure Research, 2012(1): 39-62.

[3] 郭亚林,孙文昊, 许洪伟,等. 地铁隧道盾构及钻爆法施工碳排放强度对比[J]. 现代隧道技术,2023, 60(3): 14-23.
GUO Y L, SUN W H, XU H W, et al. Comparison between shield method and drill and blast method regarding carbon emission intensity in subway tunnel construction[J]. Modern Tunnelling Technology, 2023, 60(3): 14-23. (in Chinese)

[4] 林宇亮,熊锦江,邢浩,等. 基于XGBoost-SVR组合模型的高速公路建造碳排放量预测方法研究[J]. 中南大学学报(自然科学版), 2024, 55(7): 2588-2599.
LIN Y L, XIONG J J, XING H, et al. Research on carbon emission prediction method of expressway construction based on XGBoost-SVR combined model [J]. Journal of Central South University (Science and Technology), 2024, 55(7): 2588-2599. (in Chinese)

[5] GUO C, XU J F, YANG L, et al. Life cycle evaluation of greenhouse gas emissions of a highway tunnel: a case study in China[J]. Journal of Cleaner Production, 2019, 211: 972-980.

[6] 苏征宇,张帆,韦逸清,等. 公路隧道建设碳排放量化分析计算及数字化[J]. 现代隧道技术,2022,59(增刊1):115-120.
SU Z Y, ZHANG F, WEI Y Q, et al. Calculation and digitization of carbon emission analysis for road tunnel construction [J]. Modern Tunnelling Technology, 2022, 59(Sup.1): 115-120. (in Chinese)

[7] XU J F, GUO C, CHEN X F, et al. Emission transition of greenhouse gases with the surrounding rock weakened-a case study of tunnel construction[J]. Journal of Cleaner Production, 2019, 209: 169-179.

[8] XU J F, GUO C, YU L. Factors influencing and methods of predicting greenhouse gas emissions from highway tunnel

construction in southwestern China [J]. Journal of Cleaner Production,2019,229: 337-349.

[9] 林宇亮, 左伟俊, 邢浩, 等. 基于组合赋权-VIKOR 法的山区施工便道方案优选模型[J]. 中南大学学报(自然科学版), 2024,55(2): 445-456.
LIN Y L, ZUO W J, XING H, et al. Optimization model on construction scheme of pioneer road in mountainous area based on combination weight and VIKOR method[J]. Journal of Central South University (Science and Technology), 2024, 55(2): 445-456.(in Chinese)

[10] 刘凯华,郑佳凯,谢维力,等. 基于机器学习的再生混凝土配合比设计方法[J]. 湖南大学学报(自然科学版), 2023,50(9): 88-96.
LIU K H,ZHENG J K,XIE W L,et al. Mixture design method of recycled aggregate concrete based on machine learning [J]. Journal of Hunan University (Natural Sciences), 2023, 50(9): 88-96.(in Chinese)

[11] 齐婉婉, 孙锐, 郑桐, 等. 融合XGBoost和SHAP的地表峰值加速度预测分析模型[J]. 岩土工程学报, 2023, 45(9): 1934-1943.
QI W W,SUN R,ZHENG T,et al. Prediction and analysis model for ground peak acceleration based on XGBoost and SHAP[J]. Chinese Journal of Geotechnical Engineering, 2023,45(9):1934-1943.(in Chinese)

[12] ZHANG D M, ZHANG J Z, HUANG H W, et al. Machine learning-based prediction of soil compression modulus with application of 1D settlement[J]. Journal of Zhejiang University-Science A (Applied Physics & Engineering), 2020, 21 (6): 430-444.

[13] 陈嘉,季雪,阙云,等. 改进型VGG算法对小样本路面破损的分类识别[J]. 湖南大学学报(自然科学版),2023,50(3): 206-216.
CHEN J, JI X, QUE Y, et al. Classification recognition of pavement disaster with small sample size based on improved VGG algorithm[J]. Journal of Hunan University (Natural Sciences), 2023, 50(3): 206-216.(in Chinese)

[14] HUANG L Z, BOHNE R A, BRULAND A, et al. Environmental impact of drill and blast tunnelling: life cycle assessment[J]. Journal of Cleaner Production, 2015, 86: 110-117.

[15] 郭春, 徐建峰, 张佳鹏. 隧道建设碳排放计算方法及预测模型[J]. 隧道建设(中英文), 2020, 40(8): 1140-1146.
GUO C, XU J F, ZHANG J P. Calculation methods and prediction models of carbon emission of tunnel construction[J]. Tunnel Construction, 2020, 40(8): 1140-1146.(in Chinese)

[16] ZHANG H, ZHOU Y H, HUANG Z Y, et al. Multiparameter identification of bridge cables using XGBoost algorithm [J]. Journal of Bridge Engineering, 2023, 28(5): 04023016.

[17] FRANKLIN J. The elements of statistical learning: data mining, inference and prediction [J]. The Mathematical Intelligencer, 2005, 27(2): 83-85.

[18] 闫林君, 陈慧鑫, 鲍学英, 等. 山区铁路隧道工程资源环境影响效应评估研究[J]. 铁道科学与工程学报, 2024, 21(4): 1613-1623.
YAN L J,CHEN H X,BAO X Y,et al. Evaluation and research on the resource and environmental impact effects of railway tunnel engineering in mountainous areas[J]. Journal of Railway Science and Engineering,2024,21(4): 1613-1623.(in Chinese)

[19] 戚家南, 邹伟豪, 李智杰, 等. 数据驱动的UHPC与钢筋界面黏结强度研究[J]. 中国公路学报, 2023, 36(9): 61-72.
QI J N, ZOU W H, LI Z J, et al. Study on interface bond strength between UHPC and rebar using data driven method [J]. China Journal of Highway and Transport,2023,36(9):61-72.(in Chinese)

[20] 江宝得,李秀春,罗海燕,等. 异质集成学习在滑坡易发性评价中的对比研究[J]. 土木工程学报, 2023, 56(10): 170-179.
JIANG B D,LI X C,LUO H Y,et al. A comparative analysis of heterogeneous ensemble learning methods for landslide susceptibility assessment[J]. China Civil Engineering Journal, 2023,56(10):170-179.(in Chinese)

[21] 王毅, 方志策, 牛瑞卿, 等. 基于深度学习的滑坡灾害易发性分析[J]. 地球信息科学学报, 2021, 23(12): 2244-2260.
WANG Y,FANG Z C,NIU R Q,et al. Landslide susceptibility analysis based on deep learning[J]. Journal of Geo-Information Science, 2021, 23(12): 2244-2260.(in Chinese)

[22] 马良玉,耿妍竹,梁书源, 等. 基于Stacking多模型融合的风电机组齿轮箱油池温度异常预警[J]. 中国电机工程学报,2023, 43(增刊1):242-251.
MA L Y,GENG Y Z,LIANG S Y,et al. Anomaly warning of wind turbine gearbox oil pool temperature based on Stacking fusion of multiple models[J]. Proceedings of the CSEE,2023,43(Sup.1): 242-251.(in Chinese)