

基于机器学习的煤矸石粉混凝土抗压强度预测

Prediction of Compressive Strength of Coal Gangue Powder Concrete based on Machine Learning

赵琦, 兰永强

(延安大学, 延安 716000)

摘要: 为提升煤矸石粉混凝土抗压强度预测的准确性, 构建了 70 组煤矸石粉混凝土抗压强度数据集, 引入 3 种经典机器学习集成算法, 建立 3 个考虑多因素的煤矸石粉混凝土抗压强度预测模型, 研究不同特征组合对模型性能的影响。结果表明: XGBoost 算法模型是煤矸石粉混凝土抗压强度预测的最优模型, 其决定系数 (R^2) 为 0.857, 均方根误差为 2.06; 水泥用量和养护龄期是最重要的两个特征指标; 使用水胶比、水泥用量、煤矸石粉用量、砂用量、碎石用量及养护龄期等 6 个特征时, 3 种不同算法的平均决定系数最高为 0.844, 由此反映了模型较高的准确性。研究成果可为煤矸石粉混凝土的工程应用提供快速可靠的决策依据。

关键词: 煤矸石粉; 混凝土; 抗压强度; 预测; 机器学习

中图分类号: TU528

文献标志码: A

文章编号: 1005-8249 (2025) 04-0157-04

DOI: 10.19860/j.cnki.issn1005-8249.2025.04.028

ZHAO Qi, LAN Yongqiang

(Department of Infrastructure, Yan'an University, Yan'an 716000, China)

Abstract: To improve the accuracy of predicting the compressive strength of coal gangue powder concrete, a dataset consisting of 70 groups of compressive strength data for coal gangue powder concrete was constructed. Three classic machine learning ensemble algorithms were introduced, and three prediction models for the compressive strength of coal gangue powder concrete considering multiple factors were established. The impact of different feature combinations on model performance was studied. The results show that the XGBoost algorithm model is the optimal model for predicting the compressive strength of coal gangue powder concrete, with a coefficient of determination (R^2) of 0.857 and a root mean square error of 2.06; Cement dosage and curing age are the two most important feature indicators; when using 6 features including water-binder ratio, cement dosage, coal gangue powder dosage, sand dosage, gravel dosage, and curing age, the average coefficient of determination of the three different algorithms reaches the highest value of 0.844, which reflects the high accuracy of the model. The research results can provide a fast and reliable decision-making basis for the engineering application of coal gangue powder concrete.

Key words: coal gangue powder; concrete; compressive strength; forecast; machine learning

0 引言

煤矸石粉应用于混凝土中, 是实现煤矸石“变废为宝”的重要处置方式^[1]。快速准确地确定煤矸石粉混凝土的抗压强度值, 可以使煤矸石粉混凝土配合比快速确定, 减少混凝土拌和中的反复试验, 同时有助于促进煤矸石粉混凝土在建设领域更好地

作者简介: 赵琦 (1987—), 女, 硕士, 工程师, 研究方向: 施工管理。

通信作者: 兰永强 (1978—), 男, 本科, 高级工程师, 研究方向: 项目管理。

收稿日期: 2024-07-03

研究与应用

推广应用。近年来学者们对煤矸石磨成粉、作为掺合料取代水泥对混凝土性能的影响进行了研究，取得了丰硕的成果^[2-4]，为煤矸石粉混凝土抗压强度的预测奠定了基础。考虑到影响煤矸石粉混凝土抗压强度的因素较多^[5-6]，且影响作用复杂，传统方法对煤矸石粉混凝土抗压强度预测具有工作量大、花费时间长、难以考虑多种复杂因素影响等缺点。开展快速准确的煤矸石粉混凝土抗压强度预测方法研究具有重要的理论意义和实践价值。

随着人工智能、大模型、云计算等新兴信息技术的不断创新发展，以及在各行各业大量的持续迭代应用，机器学习为混凝土工程领域诸多研究提供了新的视角与方法^[7-8]。在混凝土抗压强度预测方面也得到了广泛的应用与发展，持续推动了混凝土抗压强度预测准确率的提升。例如，丁熠亮等^[9]开展了基于 DBSCAN - AIC - Stacking 的混凝土抗压强度预测研究；周继发等^[10]基于 HEMNG 模型开展了混凝土抗压强度预测；欧阳利军等^[11]采用机器学习方法开展了超高性能混凝土的抗压强度预测研究。在煤矸石粉混凝土抗压强度预测方面也有部分学者开展相关研究。例如，邹品玉等^[6]建立了基于神经网络的煤矸石粉混凝土抗压强度预测模型，模型评估结果表明所建模型较好的预测效果；周双喜等^[12]基于 BP 人工神经网络，以 7 个影响因素为输入参数，28 d 强度为输出参数，对煤矸石粉混凝土的抗压强度进行了预测；崔曙东^[13]基于 15 组煤矸石混合料抗压强度数据，建立了基于遗传算法优化的 SVM 煤矸石混合料无侧限抗压强度的预测模型，模型评价结果显示所提模型对煤矸石混合料无侧限抗压强度具有很好的预测效果。尽管当前研究取得了一定进展，然而基于机器学习的煤矸石粉混凝土抗压强度的研究还较少，同时当前研究在开展预测时采用的数据量较少，且多采用单一机器学习方法，针对同一数据集无法确定最优的预测模型，造成当前预测模型的泛化能力较差，限制了模型的进一步应用。

针对当前研究的不足，构建了 70 组煤矸石粉混凝土抗压强度数据集，建立了基于 3 种机器学习算法的煤矸石粉混凝土抗压强度预测模型，研究不同特征组合对模型性能的影响，采用均方根误差、决定系数 (R^2) 等指标对 3 个模型的回归性能进行综合

评估，以期得到最优的煤矸石粉混凝土抗压强度预测模型，发展并完善煤矸石粉混凝土抗压强度预测方法。

1 数据获取及清洗

1.1 数据获取

构建了含 70 组煤矸石粉混凝土抗压强度数据的数据集^[13-14]，数据集包含水胶比、水泥用量、煤矸石粉用量、砂用量、碎石用量及养护龄期等 6 个影响煤矸石粉抗压强度的因素，具体统计参数见表 1。

表 1 70 组煤矸石粉混凝土抗压强度数据集统计参数
Table 1 Statistical parameters of compressive strength dataset for 70 sets of coal gangue powder concrete

特征	水胶比	水泥用量/kg	煤矸石粉用量/kg	砂用量/kg	碎石用量/kg	养护龄期/d	抗压强度/MPa
最大值	0.63	392.0	137.2	640	1 240	180	44.8
最小值	0.49	197.6	0	577	1 227	7	16.7
平均值	0.56	280.8	67.1	608.5	1 233.5	79	33.2

1.2 数据清洗

煤矸石粉混凝土抗压强度案例中选用的 6 个特征（煤矸石粉混凝土抗压强度的评价指标）均是数值型数据，表 1 所示 6 个特征各自的取值范围差异较大。为计算得到性能最佳的煤矸石粉混凝土抗压强度回归模型，在计算前需要对数据集进行归一化预处理。采用线性函数归一化（Min - Max Scaling）方法对原始数据进行线性变换，使结果映射到 [0, 1] 的范围，从而实现各个特征的数据均处在同一范围。

随机选择数据集中 70% 数量（49 组）的数据作为模型训练集，其余数据作为检验模型性能的测试集。在不同算法模型训练中，采用为随机种子赋固定值的方式，保证不同模型使用的训练集与测试集一致性。为获得回归性能最好的预测模型，在模型训练过程中采用网格搜索技术进行模型超参数的优化。

2 煤矸石粉混凝土抗压强度预测模型

特征组合对于提升模型性能至关重要。结合 6 个特征的物理含义，初步筛选形成了 8 种特征组合，每种特征组合包含特征不同，见表 2。在模型优化构建过程中，分别计算在采用 8 种不同特征组合数据时，GBRT、XGBoost、ExtraTree 三种算法模型的性能，模型性能最优时对应的特征组合即为最优特征组合。

表 2 8 种不同的特征组合
Table 2 8 different feature combinations

特征组合序号	水胶比	水泥用量	煤矸石粉用量	砂用量	碎石用量	养护龄期
F-1	√	√	√	√	√	√
F-2	√	√	√	√		√
F-3	√	√	√			√
F-4		√	√			√
F-5		√				√
F-6	√		√	√	√	√
F-7			√			√
F-8	√	√				√

3 预测模型性能分析

采用不同的特征组合数据训练得到的模型决定系数见表 3，3 种机器学习算法模型的决定系数在 -0.264~0.857 之间（负数表示预测效果不如直接取平均值）。XGBoost 算法模型的决定系数最高，不同特征组合时的决定系数位于 0.776~0.857 之间，均值为 0.812；而 GBRT 算法模型不同特征组合时的决定系数位于 0.700~0.839 之间，均值为 0.800；ExtraTree 算法模型的决定系数最低，不同特征组合时的决定系数位于 -0.264~0.839 之间，均值为 0.718。XGBoost、GBRT 算法模型对特征选择的依赖性不强，而 ExtraTree 算法模型对特征的选择较敏感，采用不同特征组合时其决定系数波动较大，采用特征组合 F-7 时达最小的-0.264。采用特征组合 F-1 时，即使用全部 6 个特征时，不同算法的平均决定系数最高为 0.844，较平均决定系数第二大的特征组合 F-8，回归性能增加 8.8%。因此 F-1 为最优特征组合，在最优特征组合下，XGBoost 算法模型性能最好，其模型决定系数为 0.857，ExtraTree、GBRT 算法模型的决定系数基本一致，分别为 0.839、0.837。

表 3 不同算法不同特征组合时的决定系数
Table 3 Determination coefficients of different algorithms using different feature combinations

特征组合	XGBoost 模型	GBRT 模型	ExtraTree 模型	R ² 平均值
F-1	0.857	0.837	0.839	0.844
F-2	0.828	0.831	0.528	0.729
F-3	0.776	0.835	0.459	0.690
F-4	0.798	0.839	0.340	0.659
F-5	0.825	0.700	0.652	0.726
F-6	0.801	0.823	0.579	0.734
F-7	0.795	0.746	-0.264	0.426
F-8	0.819	0.791	0.718	0.776
R ² 平均值	0.812	0.800	0.481	

根据上述结果，对采用特征组合 F-1 时不同算法的均方根误差、决定系数进行了计算，计算结果见表 4。XGBoost 算法模型的均方根误差 *RMSE* 最小，为 2.06，其决定系数最大，为 0.857；而 GBRT 算法模型与 ExtraTree 算法模型的均方根误差 *RMSE* 和决定系数基本一致，两者的均方根误差 *RMSE* 分别为 2.20、2.19，决定系数分别为 0.837、0.839。因此，综合考虑机器学习的均方根误差 *RMSE* 和决定系数等指标，三种模型中 XGBoost 算法模型性能最好。

表 4 F-1 特征组合不同算法强度预测模型的评估指标
Table 4 Evaluation indicators of different algorithm strength prediction models when using F-1 feature combination

评估指标	XGBoost 模型	GBRT 模型	ExtraTree 模型
<i>RMSE</i>	2.06	2.20	2.19
决定系数	0.857	0.837	0.839

21 个测试集样本的实际值与三种模型预测值见表 5。XGBoost 算法模型预测值与实际值的误差在 -16.71%~10.38% 之间，平均值为 -2.58%；GBRT

表 5 F-1 特征组合测试集实际值与预测值对比情况
Table 5 Comparison between the true values of the test set and the predicted values of different algorithms when using F-1 feature combination

序号	测试值 /MPa	XGBoost 模型		GBRT 模型		ExtraTree 模型	
		预测值 /MPa	误差 /%	预测值 /MPa	误差 /%	预测值 /MPa	误差 /%
1	37.60	37.59	-0.03	38.96	3.62	39.10	3.99
2	33.90	31.70	-6.48	31.84	-6.07	24.40	-28.02
3	37.70	39.54	4.89	40.15	6.50	39.10	3.71
4	42.40	35.31	-16.71	37.39	-11.82	31.70	-25.24
5	31.10	28.20	-9.32	29.44	-5.32	24.40	-21.54
6	34.00	33.60	-1.19	33.08	-2.71	30.00	-11.76
7	39.60	43.24	9.18	43.93	10.94	44.80	13.13
8	43.80	41.89	-4.37	40.50	-7.54	38.30	-12.56
9	27.40	26.83	-2.08	26.86	-1.96	31.70	15.69
10	33.40	34.45	3.13	34.25	2.54	30.80	-7.78
11	21.90	24.17	10.38	22.08	0.82	20.80	-5.02
12	39.10	38.43	-1.71	38.75	-0.88	38.80	-0.77
13	41.10	37.74	-8.19	39.75	-3.29	39.10	-4.87
14	42.90	41.89	-2.36	42.45	-1.05	44.80	4.43
15	36.50	34.02	-6.79	33.66	-7.79	31.40	-13.97
16	38.80	37.96	-2.18	35.83	-7.66	36.90	-4.90
17	29.30	28.90	-1.35	27.02	-7.80	30.00	2.39
18	41.60	39.11	-5.98	41.10	-1.21	39.10	-6.01
19	35.20	34.99	-0.59	35.70	1.42	31.70	-9.94
20	31.60	28.81	-8.82	29.77	-5.80	31.70	0.32
21	38.00	36.67	-3.51	38.69	1.80	39.10	2.89
误差平均值			-2.58		-2.06		-5.04

研究与应用

算法模型预测值与实际值的误差在 $-11.82\% \sim 10.94\%$ 之间, 平均值为 -2.06% ; ExtraTree 算法模型预测值与实际值的误差在 $-28.02\% \sim 15.69\%$ 之间, 平均值为 -5.04% 。因此, 从测试集 21 组样本的预测结果来看, ExtraTree 算法模型的误差较大, XGBoost 算法模型其次, GBRT 算法模型误差最小。

对 F-1 特征组合 6 个特征的重要性进行计算, 计算结果如图 1 所示。对于 XGBoost 算法模型、GBRT 算法模型、ExtraTree 算法模型来说, 最重要的两个特征均是水泥用量和养护龄期, 且这两个特征的重要性之和分别是 0.819、0.977、0.959, 远大于其余特征的重要性。对 XGBoost 算法模型, 水胶比、煤矸石粉用量及砂用量的特征重要性分别为 0.056、0.087、0.037, 其余特征(碎石用量)重要性均小于 0.01; 对于 GBRT 算法模型和 ExtraTree 算法模型, 煤矸石粉用量的特征重要性分别为 0.012、0.029, 其余特征(水胶比、砂用量、碎石用量)重要性均小于 0.01。碎石用量这一特征对三种算法模型均几乎不起作用(重要性均小于 0.01)。上述结果与前述不同特征组合下三种算法模型的决定系数情况吻合。

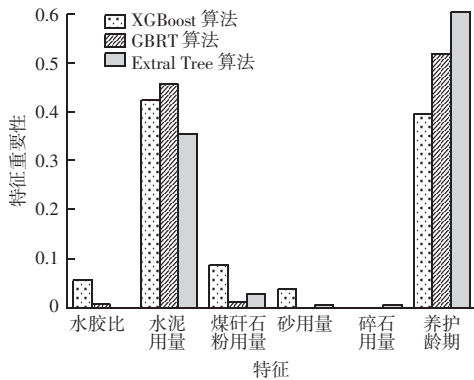


图1 F-1 特征组合不同算法不同特征的重要性
Fig. 1 Shows the importance of different features in different algorithms when using feature combination F-1

4 结论

为提升煤矸石粉混凝土抗压强度预测准确性, 建立了3种机器学习方法的预测模型, 研究了不同特征组合时各预测模型的性能, 获得了不同特征的重要性, 得到以下结论:

(1) 使用全部 6 个特征时, 不同算法的平均决

定系数最高为 0.844, XGBoost 算法模型预测效果最好, 决定系数为 0.857, 均方根误差为 2.06。

(2) 使用全部 6 个特征时, ExtraTree 算法模型对测试集 21 组样本的预测误差范围较大, XGBoost 算法模型预测误差为其次, GBRT 算法模型对应的误差范围最小。

(3) 对于 XGBoost 算法模型、GBRT 算法模型、ExtraTree 算法模型来说, 最重要的两个特征均是水泥用量和养护龄期, 且这两个特征的重要性之和分别是 0.819、0.977、0.959, 远大于其余特征的重要性。

参 考 文 献

[1] 宋强, 杨玉鑫, 许世鹏, 等. 煤矸石混凝土性能及提升研究进展 [J]. 煤炭科学技术, 2025, 53 (2): 407-425.

[2] 张亚鹏, 张凤娇, 曹佳伟, 等. 洗选煤矸石透水混凝土的制备及其性能研究 [J]. 粉煤灰综合利用, 2024, 38 (6): 1-5.

[3] 王建华, 赖益梁, 刘刚, 等. 煤矸石粉制备混凝土用复合掺合料的性能研究 [J]. 粉煤灰综合利用, 2024, 38 (1): 18-21, 51.

[4] 白朝能, 李霖皓, 沈远, 等. 煤矸石作为粗骨料对混凝土力学性能的影响 [J]. 四川建材, 2018, 44 (8): 1-3.

[5] 周美丽, 柯国军. 砂率对煤矸石粉混凝土和易性及抗压强度的影响 [J]. 混凝土, 2016 (8): 133-135.

[6] 邹品玉, 柯国军, 宋百姓, 等. 煤矸石粉混凝土抗压强度试验及其神经网络预测 [J]. 混凝土, 2018 (12): 31-33.

[7] 周继发, 曾晓辉, 郑振华, 等. 基于集成学习的混凝土抗压强度预测模型研究 [J]. 中南大学学报 (自然科学版), 2025, 56 (5): 1981-1992.

[8] 赵佳亮, 达列雄, 郭鸿, 等. 基于 MLP-AdaBoost 模型的混凝土抗压强度预测研究 [J]. 混凝土, 2025 (6): 17-22.

[9] 丁熠亮, 文汉云, 李素若, 等. 基于 DBSCAN-AIC-Stacking 的混凝土抗压强度预测研究 [J]. 混凝土, 2024 (12): 29-34.

[10] 周继发, 曾晓辉, 谢友均, 等. 基于 HEMNG 模型的混凝土抗压强度预测 [J]. 铁道科学与工程学报, 2025, 22 (2): 875-886.

[11] 欧阳利军, 石永超, 丁斌, 等. 基于机器学习预测超高性能混凝土的抗压强度 [J]. 混凝土与水泥制品, 2025 (3): 1-7.

[12] 周双喜, 赵碧华. 神经网络在掺活化煤矸石粉混凝土强度预测中的应用 [J]. 混凝土, 2007 (4): 26-28.

[13] 崔曙东. 基于 GA-SVM 的煤矸石混合料抗压强度预测研究 [D]. 邯郸: 河北工程大学, 2013.

[14] 孟文清, 黄祖德, 崔郡龙, 等. 煤矸石混合料抗压强度的试验研究 [J]. 煤炭工程, 2012 (7): 106-108.