

基于GA-BP算法的高炉矿渣 - 粉煤灰混凝土抗压强度预测

董建军¹, 谢洪阳¹, 戴宜文¹, 邓勇¹, 邓明伟²

¹南昌航空大学土木建筑学院, 江西 南昌

²江西建设职业技术学院, 江西 南昌

收稿日期: 2022年1月3日; 录用日期: 2022年1月16日; 发布日期: 2022年2月7日

摘要

为更准确地预测高炉矿渣 - 粉煤灰混凝土抗压强度, 在MATLAB平台上通过遗传算法对BP神经网络的初始权值和阈值进行改进, 建立了抗压强度预测的GA-BP模型。将人工神经网络(BP)、随机森林(RF)、支持向量机(SVM)、极限学习机(ELM)和多元非线性回归(MnLR)的预测结果进行对比分析, GA-BP模型的预测精度和模型稳定性等方面优势明显。从而为高炉矿渣 - 粉煤灰混凝土质量评估提供指导, 具有重要的实用价值。

关键词

高炉矿渣 - 粉煤灰混凝土, 抗压强度, 遗传算法, 预测

Compressive Strength Prediction of Blast Furnace Slag-Fly Ash Concrete Based on GA-BP Algorithm

Jianjun Dong¹, Hongyang Xie¹, Yiwen Dai¹, Yong Deng¹, Mingwei Deng²

¹School of Civil Engineering and Architecture, Nanchang Hangkong University, Nanchang Jiangxi

²Jiangxi Construction Vocational and Technical College, Nanchang Jiangxi

Received: Jan. 3rd, 2022; accepted: Jan. 16th, 2022; published: Feb. 7th, 2022

Abstract

In order to predict the compressive strength of blast furnace slag-fly ash concrete more accurately,

文章引用: 董建军, 谢洪阳, 戴宜文, 邓勇, 邓明伟. 基于 GA-BP 算法的高炉矿渣-粉煤灰混凝土抗压强度预测[J]. 统计学与应用, 2022, 11(1): 27-37. DOI: 10.12677/sa.2022.111004

a GA-BP model for compressive strength prediction was developed by improving the initial weights and thresholds of BP neural network through genetic algorithm on MATLAB platform. The prediction results of artificial neural network (BP), random forest (RF), support vector machine (SVM), extreme learning machine (ELM) and multiple nonlinear regression (MnLR) were compared and analyzed, and the GA-BP model has obvious advantages in terms of prediction accuracy and model stability. Thus, it provides guidance for the quality assessment of blast furnace slag-fly ash concrete, which has important practical value.

Keywords

Blast Furnace Slag-Fly Ash Concrete, Compressive Strength, Genetic Algorithm, Prediction

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

混凝土抗压强度是衡量混凝土质量的关键指标之一，也是工程设计、施工及验收的重要依据。准确监测或评定出建筑物结构的混凝土抗压强度，对建筑物结构的安全性具有深远的意义。对混凝土抗压强度进行评定的主要方法是先对混凝土试件进行 28 天养护[1] [2]，然后再进行测试，或利用无损检测技术对混凝土强度进行评定[3]。这类评定方法的局限性在于会影响施工进度和施工质量。对于普通混凝土强度，一般可以用水灰比为主要因素的线性函数来进行描述和预测，而对于高掺量的高炉矿渣 - 粉煤灰混凝土来说，由于组分的增加，水化反应的机理还不完全明确，影响因素复杂并具有交互作用，表现为特定的高维非线性规律[4]。

神经网络和机器学习可以挖掘到数据的深层规律，是解决非线性问题的有效手段之一，并且具有容错、联想、推测、记忆、自适应、自学习和处理复杂模式的功能，已有许多学者将神经网络和机器学习应用于混凝土抗压强度的预测。马高等[5]建立了 BP 神经网络模型预测 CFRP 约束混凝土的抗压强度；胡毅等[6]提出了一种基于随机森林(RF)的混凝土抗压强度预测方法；曹斐等[7]以支持向量机(SVM)算法为理论基础，提出一种基于马氏距离的加权型 SVR (MWSVR)的人工智能算法对混凝土强度进行预测；Zaher Mundher Yaseen 等[8]将极限学习机(ELM)模型用于轻质泡沫混凝土抗压强度的预测；许开成等[9]利用 SPSS 软件的逐步回归分析法、多元非线性回归法(MnLR)建立锂渣混凝土的强度预测模型。

为更加准确地预测高炉矿渣 - 粉煤灰混凝土的抗压强度，本研究以 UCI 的 Concrete Compressive Strength 数据集为基础，借助 MATLAB 软件，采用遗传算法优化 BP 神经网络的初始权值及阈值，以水泥、高炉矿渣粉、粉煤灰、水、减少剂、粗集料、细集料、置放天数这 8 项指标作为输入参数，以立方体抗压强度作为输出参数，建立了基于 GA-BP 神经网络的高炉矿渣 - 粉煤灰抗压强度预测模型。最后与神经网络(BP)、随机森林(RF)、支持向量机(SVM)、极限学习机(ELM)和多元非线性回归(MnLR)的预测结果进行对比分析，验证所提出的 GA-BP 模型在高炉矿渣 - 粉煤灰混凝土强度预测中的可行性，为此类混凝土强度预测提供新思路。

2. 方法及原理

BP 神经网络是一种多层前馈神经网络，该网络的特点是信号前向传递，误差反向传播。在前向传递

中,输入信号从输入层经隐含层逐层处理,直至输出层。每一层的神经元状态只影响下一层神经元状态。如果输出层得不到期望输出,则转入反向传播,根据预测误差调整网络权值和阈值,从而使BP神经网络预测输出不断逼近期望输出。BP神经网络的拓扑结构如图1所示。

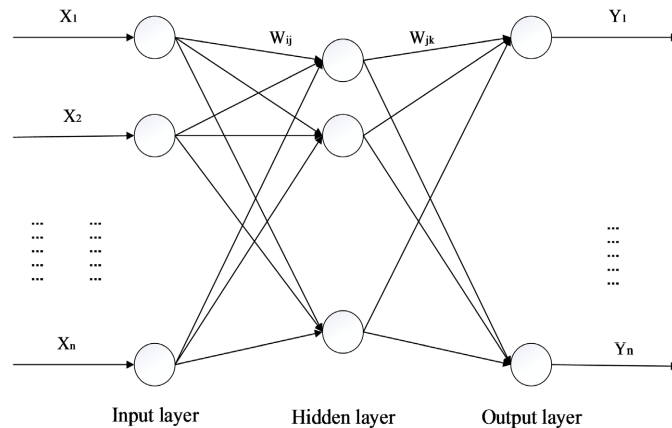


Figure 1. BP neural network topology
图 1. BP 神经网络拓扑结构图

图1中, X_1, X_2, \dots, X_n 是BP神经网络的输入值, Y_1, Y_2, \dots, Y_n 是BP神经网络的预测值, W_{ij} 和 W_{jk} 为BP神经网络权值。BP神经网络可以看成是一个非线性函数,网络输入值和预测值分别为该函数的自变量和因变量。

然而对于大数据量、高维数及多隐藏层节点等条件,各类神经网络训练过程很有可能会遇到大量的局部极值,不仅严重影响了收敛速度,而且有可能导致训练误差收敛于局部最优解而不是全局最优解,严重影响了神经网络的性能。遗传算法(Genetic Algorithm, GA)是一种模拟生物遗传及进化机制的自适应并行寻优算法,它具有全局搜索、并行性高、泛化能力强等优点。将遗传算法应用于BP神经网络,可以增强网络训练收敛的精确性,并且充分发挥遗传算法的全局寻优特性[10][11]。

GA-BP神经网络,即采用遗传算法对BP神经网络的初始权值和阈值进行改进,得到全局最优区域,接着用最优权值和阈值进行BP神经网络训练直到训练结束,并搜索最优适应度对应的参数。图2是GA-BP神经网络建立的过程。

2.1. GA 优化过程

遗传算法优化BP神经网络的要素包括种群初始化、适应度函数、选择操作、交叉操作和变异操作。

1) 种群初始化

个体编码方法为实数编码,每个个体均为一个实数串,由输入层与隐含层连接权值、隐含层阈值、隐含层与输出层连接权值以及输出层阈值4部分组成。个体包含了神经网络全部权值和阈值,在网络结构已知的情况下,就可以构成一个结构、权值、阈值确定的神经网络。

2) 适应度函数

根据个体得到BP神经网络的初始权值和阈值,用训练数据训练BP神经网络预测系统输出,把预测输出和期望输出之间的误差绝对值和E作为个体适应度值F,计算公式为

$$F = k \left(\sum_{i=1}^n \text{abs}(y_i - o_i) \right) \quad (1)$$

式中, n 为网络输出节点; y_i 为BP神经网络第 i 个节点的预测输出; k 为系数。

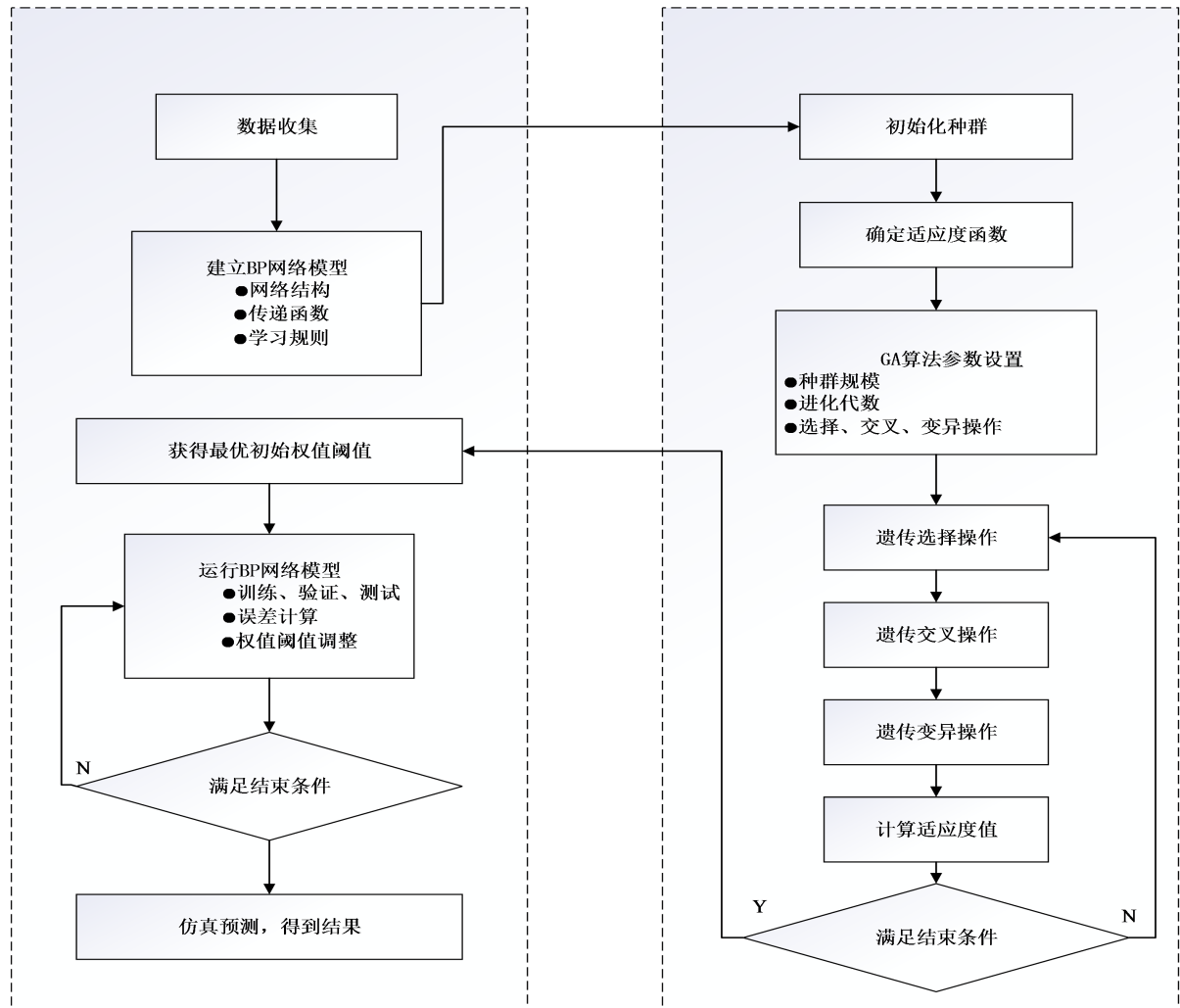


Figure 2. The establishment process of GA-BP neural network
图 2. GA-BP 神经网络的建立过程

3) 选择操作

基于适应度比例的选择策略，每个个体 i 的选择概率 p_i 为

$$f_i = k/F_i \tag{2}$$

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j} \tag{3}$$

式中， F_i 为个体 i 的适应度值，由于适应度值越小越好，所以在个体选择前对适应度值求倒数； k 为系数； N 为种群个体数目。

4) 交叉操作

由于个体采用实数编码，所以交叉操作方法采用实数交叉法，第 k 个染色体 a_k 和第 l 个染色体 a_l 在 j 位的交叉操作方法如下：

$$a_{kj} = a_{kj}(1-b) + a_{lj}b \tag{4}$$

$$a_{ij} = a_{ij}(1-b) + a_{kj}b \quad (5)$$

式中, b 是[0,1]间的随机数。

5) 变异操作

选取第 i 个个体的第 j 个基因 a_{ij} 进行变异, 变异操作方法如下:

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{\max}) * f(g) & r > 0.5 \\ a_{ij} + (a_{\min} - a_{ij}) * f(g) & r \leq 0.5 \end{cases} \quad (6)$$

式中, a_{\max} 为基因 a_{ij} 的上界; a_{\min} 为基因 a_{ij} 的下界; $f(g) = r_2(1 - g/G_{\max})^2$; r_2 为一个随机数; g 为当前迭代次数; G_{\max} 是最大进化次数; r 为[0,1]间的随机数。

2.2. 模型数据库描述与分析

记前 8 个输入参数分别为自变量 X_1 、 X_2 、 X_3 、 X_4 、 X_5 、 X_6 、 X_7 、 X_8 , 输出自变量抗压强度为 Y_1 。为了对数据库进行详细的描述, 统计分析了所有变量的最小值、平均值、中值、最大值、标准差、偏度和峰度, 如表 1 所示。

Table 1. Statistical parameters of input and output variables

表 1. 输入和输出变量的统计参数

	Unit	Min	Max	Median	Average	Std	Sk	Kur
Cement (X_1)	Kg/m ³	102.00	540.00	272.90	281.20	104.46	0.51	-0.52
Blast Furnace Slag (X_2)	Kg/m ³	0.00	359.40	22.00	73.90	86.24	0.80	-0.51
Fly Ash (X_3)	Kg/m ³	0.00	200.10	0.00	54.20	63.97	0.54	-1.33
Water (X_4)	Kg/m ³	121.80	247.00	185.00	181.60	21.35	0.07	0.12
Superplasticizer (X_5)	Kg/m ³	0.00	32.20	6.40	6.20	5.97	0.91	1.41
Coarse Aggregate (X_6)	Kg/m ³	801.00	1145.00	968.00	972.90	77.72	-0.04	-0.60
Fine Aggregate (X_7)	Kg/m ³	594.00	992.60	779.50	773.60	80.14	-0.25	-0.10
Age (X_8)	d	1	365	28	-	-	-	-
Concrete compressive strength (Y_1)	MPa	2.33	82.60	34.44	35.82	16.70	0.42	-0.31

-代表无实际意义; Std-代表偏度; Kur-代表峰度。

皮尔逊相关系数可以分析自变量之间以及自变量与因变量之间的相关性, 利用 Rstudio 软件进行计算绘图, 其结果如图 3 所示。由图 3 可以看出, 各输入变量之间相关性不高, 不存在多重共线性问题。

在 1030 组粉煤灰 - 矿渣混凝土数据中, 随机抽取 70% 的数据作为训练样本, 30% 的数据作为测试样本。GA-BP 模型与其他对照组模型均以此标准设置训练样本和测试样本。

2.3. 模型的性能指标

本文选取五个统计量值用于评估 MARS 模型和对照模型的预测精度, 即决定系数(R^2)、均方误差(RMSE)、平均值误差(MAE)、平均绝对偏差百分比误差(RMAE)和效益系数(E), 计算公式见表 2。

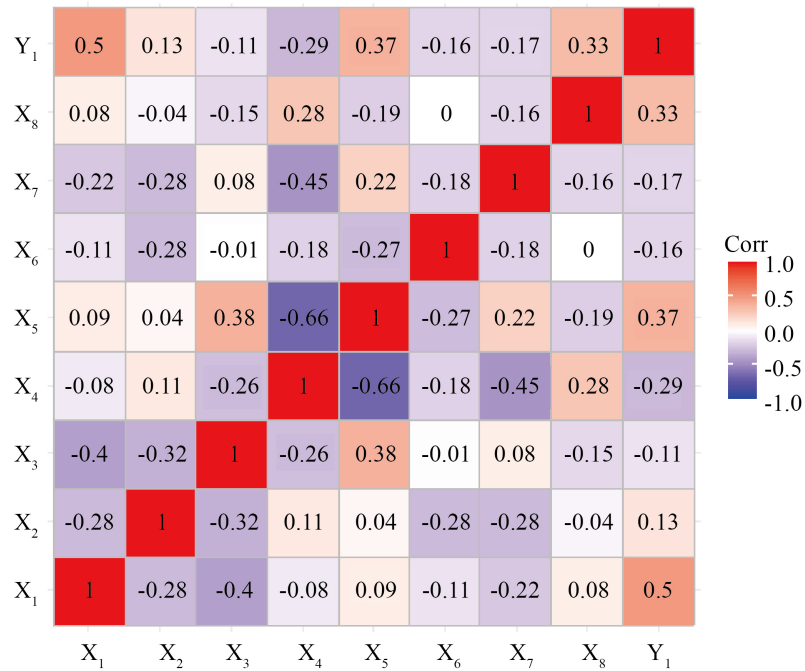


Figure 3. Heat map of correlation among variables
图 3. 各变量的相关性热图

Table 2. Statistical index calculation formula
表 2. 统计指标计算公式

Measure	Calculation
Coefficient of determination (R ²)	$R^2 = 1 - \frac{\frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}$
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - y_i)^2}$
Mean absolute error (MAE)	$MAE = \frac{1}{N} \sum_{i=1}^N abs(Y_i - y_i)$
Root mean absolute error (RMAE)	$RMAE = 100 \cdot \frac{MAE}{\bar{y}}$
Coefficient of efficiency (E)	$E(\%) = 1 - \frac{\sum_{i=1}^N (Y_i - y_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$

y_i 是实际目标值, \bar{y} 是 y_i 的平均值; Y_i 是预测值, \bar{Y} 是 Y_i 的平均值; N 是模型数据数目。

3. 模型验证

3.1. 训练 MARS 模型及结果分析

本文从模型数据库中随机选取 721 组数据构建 MARS 模型的训练样本, 剩余的 309 组数据则为测试

样本, 基于水泥含量、高炉矿渣粉含量、粉煤灰含量、水含量、减少剂含量、粗集料含量、细集料含量、龄期共 8 个特征, 利用 MATLAB 软件建立粉煤灰-矿渣混凝土预测的 GA-BP 模型, 参考相关文献[12], 输入节点个数设为 8, 输出节点个数为 1, 基因的上下界为 ± 1 , 变量个数为 16, 最大遗传 19 代。因为训练样本和测试样本数据较多, 为了展示拟合的效果, 两组分别选取 25 组数据进行可视化绘图, 如图 4、图 5 所示。

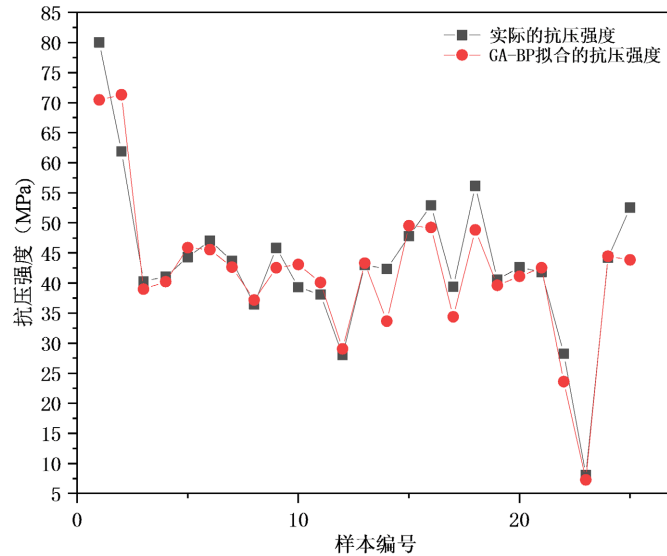


Figure 4. GA-BP model training sample fitting effect

图 4. GA-BP 模型训练样本拟合效果图

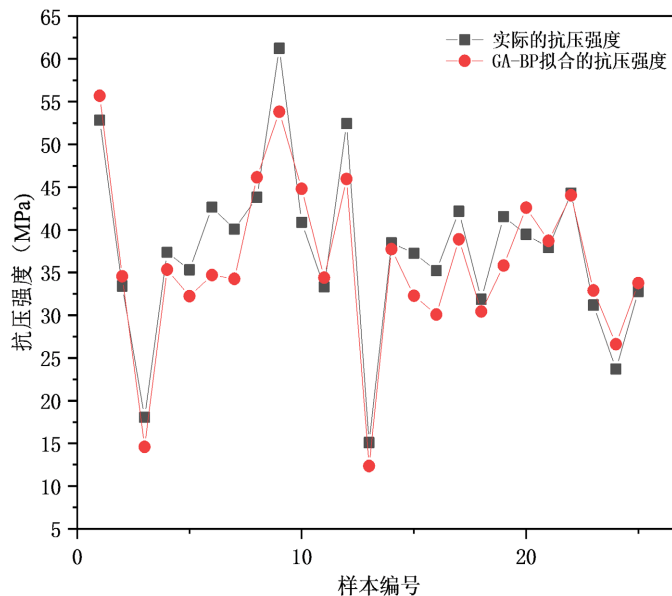


Figure 5. GA-BP model test sample fitting effect

图 5. GA-BP 模型测试样本拟合效果图

由图 4、图 5 可知, 预测值和实际值接近, 即 GA-BP 模型预测结果准确, 误差较小, 拟合精度高。

3.2. 模型性能比较

将 GA-BP 模型与各种模型的预测性能进行比较, 包括 BP、RF、SVM、ELM 和 MnLR 等非线性模型, 其拟合结果如下所示。

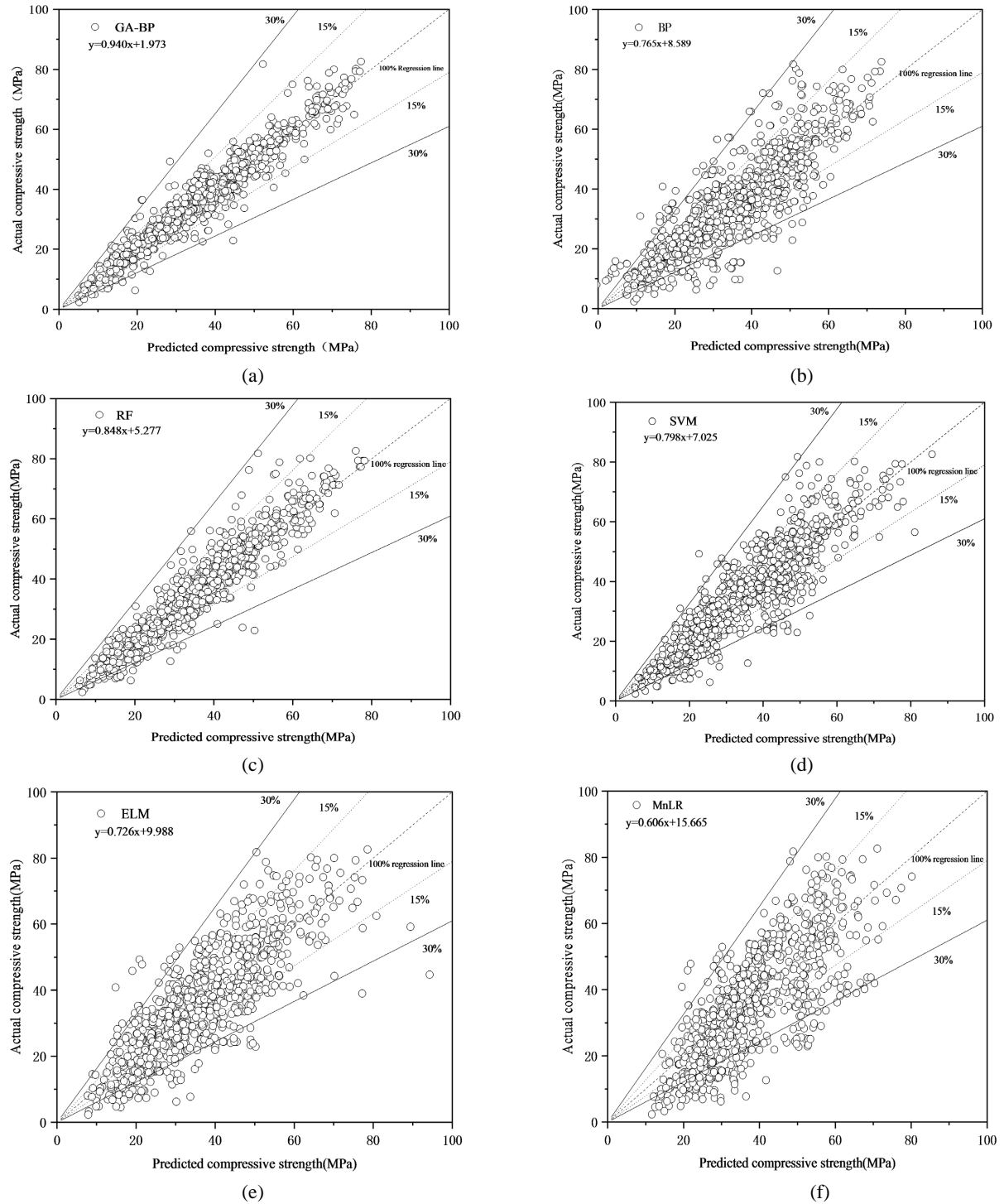


Figure 6. Scatter plot of predicted and actual values of each model

图 6. 各个模型预测值与实际值的分布散点图

由图 6 可得, GA-BP 预测模型的大部分散点数据集中在 100% 回归线及其周围, 而 BP、RF、SVM、ELM 和 MnLR 散点数据离散性很大。将图 6 中的数据根据 $y = a_0x + a_1$ 的形式进行拟合, 与对照组的五个模型相比, MARS 模型的 a_0 值更接近于 1, a_1 值更接近于 0。验证了 GA-BP 模型的预测精度更高, 误差更小。

3.3. 模型的统计指标

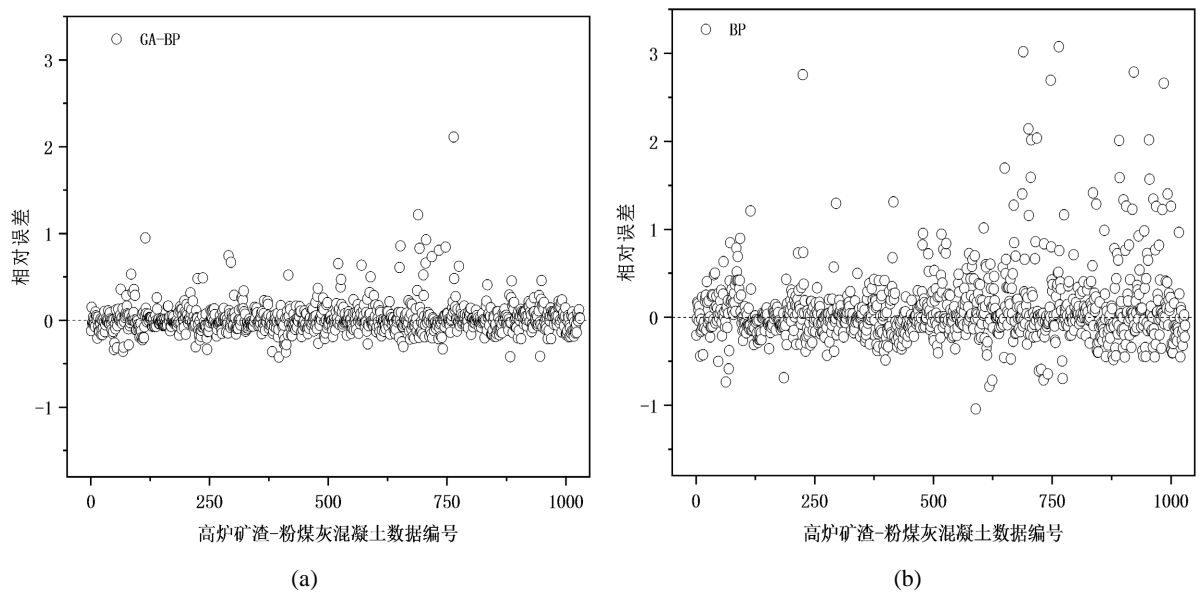
根据前文所述的评价指标公式, 计算结果见表 3。

Table 3. Results of R^2 , RMSE, MAE, RMAE and E (%) for different models
表 3. 不同模型的 R^2 、RMSE、MAE、RMAE 和 E (%) 结果

Model	R^2	RMSE	MAE	RMAE	E (%)
GA-BP	0.940	4.085	2.842	7.935	94.016
BP	0.765	40.503	6.983	19.496	70.946
RF	0.848	13.022	3.606	10.066	90.659
SVM	0.798	26.380	5.376	15.009	81.077
ELM	0.726	40.503	6.983	19.496	70.946
MnLR	0.606	10.482	8.273	23.096	60.591

由表 3 可得, GA-BP 预测模型比其他模型的 R^2 、E (%) 值高, RMSE、MAE、和 RMAE 值小, 即 MARS 和 RF 在该数据库上预测性能优越。

由图 7 可得, GA-BP 模型的相对误差大部分集中在 0 线周围, 其余模型的相对误差波动很大。GA-BP 的预测性能稳定, 明显优于其余模型。



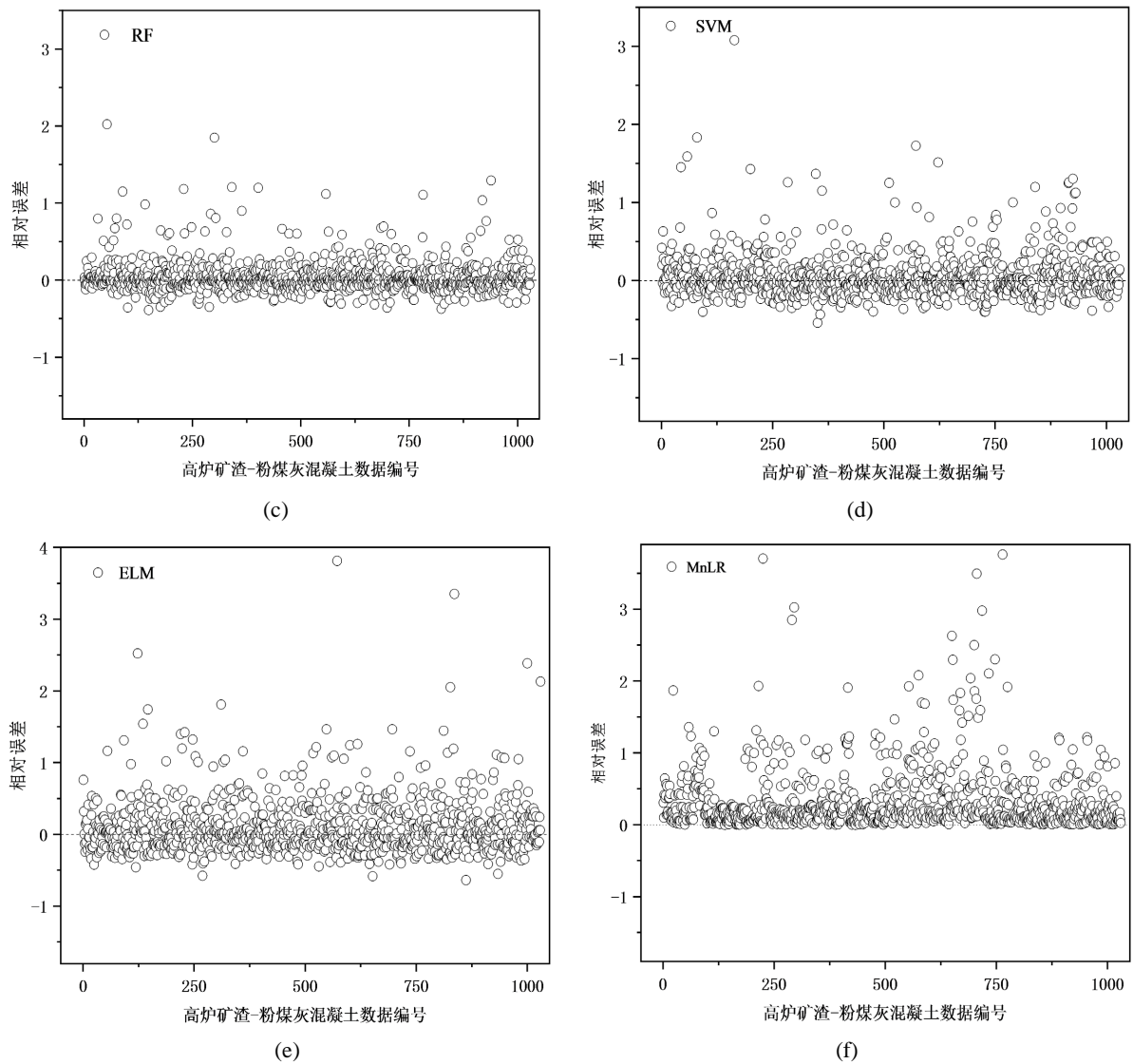


Figure 7. The relative error variation of each model
图 7. 每个模型的相对误差

4. 结论

准确预测或评定高炉矿渣-粉煤灰混凝土抗压强度具有重要的工程意义。由于影响混凝土抗压强度的因素有很多，而且这些影响间往往相互不独立，影响了对抗压强度预测或评定的精确度。因此，在评定或预测模型建立之前，有必要对各因素指标进行相关性验证，选择最能表征抗压强度的因素指标参与评定和建模，这样有助于提高评定结果的准确性和所建模型的预测精度。本文利用 GA-BP 算法建立高炉矿渣-粉煤灰混凝土的抗压强度预测模型，与人工神经网络(BP)、随机森林(RF)、支持向量机(SVM)、极限学习机(ELM)和多元非线性回归(MnLR)的预测结果进行对比分析，GA-BP 模型在 R^2 、RMSE、MAE、RMAE 和 E 值上都优于其他模型，进一步说明该模型的可靠性和准确性，可很好地应用于工程实际之中。

参考文献

- [1] Hussain, Y. and Rolph C. (2015) Coriolis Mass Flowmeter. *Flow Measurement & Instrumentation*, **3**, 241-246.

-
- [2] 宋明刚, 樊尚春. Coriolis 质量流量测量技术的发展综述[J]. 测控技术, 2000, 19(10): 1-2+9.
- [3] 段慧明, 史振东. 介绍 2008 版《科里奥利质量流量计》检定规程[J]. 中国计量, 2009(4): 121-122.
- [4] 何晓凤, 周红标. 基于 GA-BP 的混凝土抗压强度预测研究[J]. 淮阴工学院学报, 2011, 20(3): 21-25.
- [5] 马高, 刘康. 基于 BP 神经网络 CFRP 约束混凝土抗压强度预测[J]. 湖南大学学报(自然科学版), 2021, 48(9): 88-97.
- [6] 胡毅, 张陆山, 袁福银, 李铁军, 吴贤国, 邓婷婷. 基于随机森林的混凝土强度预测研究[J]. 施工技术, 2020, 49(17): 89-94.
- [7] 曹斐, 周彧, 王春晓, 任梦宇, 周峰. 一种改进的支持向量回归的混凝土强度预测方法[J]. 硅酸盐通报, 2021, 40(1): 90-97.
- [8] Yaseen, Z.M., Deo, R.C., Hilal, A., *et al.* (2018) Predicting Compressive Strength of Lightweight Foamed Concrete Using Extreme Learning Machine Model. *Advances in Engineering Software*, **115**, 112-125.
<https://doi.org/10.1016/j.advengsoft.2017.09.004>
- [9] 许开成, 毕丽苹, 陈梦成. 基于 SPSS 回归分析的锂渣混凝土抗压强度预测模型[J]. 建筑科学与工程学报, 2017, 34(1): 15-24.
- [10] 任谢楠. 基于遗传算法的 BP 神经网络的优化研究及 MATLAB 仿真[D]: [硕士学位论文]. 天津: 天津师范大学, 2014.
- [11] 胡立黎. 基于神经网络和遗传算法的无粘结部分预应力高强混凝土梁延性研究[D]: [硕士学位论文]. 西安: 西安理工大学, 2006.
- [12] 杨杰, 占君, 张继传. MATLAB 神经网络 30 例[M]. 北京: 电子工业出版社, 2014.