

基于 GA-SA-SVR 模型的输电边坡危险性预测

段国勇^{1,2}, 韩亮^{1,2}, 王彦海^{1,2}, 吕军旗^{1,2}, 郑武略³

(1. 湖北省输电线路工程技术研究中心(三峡大学), 湖北宜昌 443002;

2. 三峡大学电气与新能源学院, 湖北宜昌 443002;

3. 中国南方电网有限责任公司超高压输电公司广州局, 广东广州 510600)

摘要:输电边坡的危险性预测对于国内输电线路安全运营意义重大。该文依托某输电公司输电线路区域的边坡隐患排查及状态评估数据,对数据库进行筛选,以杆塔边缘距离、边坡高度、坡度、周边土地情况、岩土性质以及植被情况这六项作为输入特征值,危险系数作为输出标签建立支持向量回归(Support Vector Regression, SVR)预测模型,并采用遗传(Genetic Algorithm, GA)和模拟退火(Simulated Annealing algorithm, SA)的单独优化算法和组合优化算法分别对SVR模型进行优化,并设置鱼鹰、猎食者等优化算法作为对照组。结果表明:组合算法的优化效果要优于单一算法的优化效果,遗传-模拟退火组合算法(GA-SA)的优化效果在准确率和拟合程度上更有优势,测试集 R^2 为0.937 5, MSE值为0.001 2, 适应度函数 $f(x)$ 值为0.072 4。该模型预测性能较好,相较原方法更加客观智能。

关键词:输电边坡;危险性预测;支持向量回归;模拟退火算法;组合算法

中图分类号:TP18;TN911-34

文献标识码:A

文章编号:1673-629X(2024)12-0207-06

doi:10.20165/j.cnki.ISSN1673-629X.2024.0282

Hazard Prediction of Transmission Slope Based on GA-SA-SVR Model

DUAN Guo-yong^{1,2}, HAN Liang^{1,2}, WANG Yan-hai^{1,2}, LYU Jun-qi^{1,2}, ZHENG Wu-lue³

(1. Hubei Transmission Line Engineering Technology Research Center, Three Gorges University, Yichang 443002, China;

2. School of Electrical and New Energy, Three Gorges University, Yichang 443002, China;

3. China Southern Power Grid Co., Ltd., Guangzhou Bureau, EHV Transmission Company, Guangzhou 510600, China)

Abstract: Hazard evaluation of transmission slopes is of great significance for the safe operation of transmission lines in China. Relying on the data of transmission line slope hazard investigation and condition assessment of a transmission company, the database was screened, and the six items of distance from the edge of the tower, height of the slope, slope, surrounding land, geotechnical properties, and vegetation were used as the input eigenvalues, and the hazard coefficients were used as the output labels to establish a prediction model using Support Vector Regression (SVR). The Genetic Algorithm (GA) and Simulated Annealing (SA) of the individual optimization algorithm and the combination of optimization algorithms are used to optimize the SVR model, respectively, and set up OOA, HPO and other optimization algorithms as a control group. The results show that the optimization effect of the combination algorithm is better than that of the single algorithm optimization, and the optimization effect of the genetic-simulated annealing combination algorithm (GA-SA) is more advantageous in terms of accuracy and degree of fit, with an R^2 of 0.937 5 for the test set, an MSE value of 0.001 2, and a fitness function $f(x)$ value of 0.072 4. The model has better prediction performance and is more objective and intelligent compared to the original method.

Key words: transmission slopes; hazard prediction; support vector regression; simulated annealing algorithm; combinatorial algorithm

0 引言

随着输电通道的建设,越来越多的输电塔建在了山坡以及丘陵地区,由边坡失稳造成输电塔基础的沉

降及倾斜事件对输电运维造成了不小损失,且地质灾害预警预测对于保障居民安全和降低损失至关重要^[1]。

收稿日期:2024-06-22

修回日期:2024-10-25

基金项目:国家自然科学基金联合基金项目(U2034203, U22A20600);南方电网广州局输电线路塔基区域边坡灾害危险性评价与分级研究技术服务科研项目(SDHZ2022341)

作者简介:段国勇(1986-),男,讲师,博士,从事输电线路力学与岩土工程研究;通讯作者:韩亮(1998-),男,硕士研究生,研究方向为输电工程与算法评价。

国内外研究人员在边坡稳定性的相关方向得到了许多成果。陈新民等^[2]以某±800 kV 特高压直流输电线路为背景,用三维有限差分数值模拟的方式对塔基区域边坡稳定性进行了模拟与分析;Alkahtani 等^[3]用一系列实验模型测试以及有限元分析的方法分析和评估了无粘性边坡对输电塔基座的极限承载力。在用智能优化算法优化支持向量回归模型方面,史耀凡等^[4]利用主成分分析法、遗传算法优化的支持向量机模型对地表下沉系数进行了预测;贾荣谷等^[5]用 GA 优化的支持向量机模型实现了对边坡工程实例稳定性精准的预测;张建涛等^[6]用鲸鱼优化算法对超参数进行优化并结合随机森林模型对边坡稳定性进行了预测;此外,还有蜂群^[7]、麻雀^[8]、鲸鱼^[9]、狮群^[10]等算法也可用于优化支持向量回归模型。综上所述,许多学者在借助 SVR 模型或者其他机器学习模型对边坡的稳定性预测中取得了成果,而将 SVR 模型用于输电塔-边坡体系的稳定性预测的研究尚不多见,前者大多是将边坡本体的特征参数作为各自预测模型的输入特征,而后者需要将输电线路边坡运维相关的影响因子作为输入特征来构建预测模型。

综上,该文基于南方电网某输电公司对线路边坡隐患排查及状态评估所采集的样本数据,建立 SVR 预测模型来预测边坡状态,并采用遗传和模拟退火组合

算法对模型进行优化。

1 数据整理

1.1 数据库处理

南方电网某输电处理司着重对历年易发生边坡塌方区域以及新建线路区域开展了边坡稳定性评价,依据健康度评价方法对各项影响因子进行扣分,并以扣分总和作为危险系数对边坡危险程度进行评价。

对比边坡灾害发生区域,由原健康度评价方法所总结的数据存在比一般工况更加严苛的情况没有列入隐患边坡范畴,这表明健康度评价法存在主观评价的局限性且不够保守,故对样本数据进行筛选,用更加客观科学的数学模型对边坡危险性进行评价。

为方便进行特征值输入,对数据库进行以下筛选及整理,重点关注坡高 25 m 及以上、坡度 40°及以上、最近距离在 3 m 以内的工况,对扣分太低的工况进行剔除,对“周边土地情况”以及“植被情况”两项汇总的数据进行离散化处理,将输入特征总结为边坡与杆塔基础最近距离、边坡高度、边坡坡度、周边土地情况、岩土性质以及植被情况这六项,输出标签为危险系数,数据库整理完后最终得到 624 组样本数据,如表 1 所示,以此来代入预测模型进行危险性评价。

表 1 样本数据

序号	线路名称	周边土地及基面情况	植被情况		杆塔边缘距离/m	边坡高度/m	边坡坡度/°	岩土性质	危险系数
			隐患描述	植被类型					
1	xxxx 线 1718#	C	植被不完整,但稳定	2.0	28.0	21	60	土质边坡	116
2	xxxx 线 1255#	C	植被不完整,但稳定	2.0	26.3	15	70	土质边坡	104
3	xxxx 线 1417#	D	植被不完整且不稳定	3.0	5	4	70	岩土混合边坡	152
...
623	xxx 线 391#	A	植被完整稳定	1.0	0.1	30	30	土质边坡	84
624	xxx 线 441#	A	植被完整稳定	1.0	2.0	2	50	岩土混合边坡	68

1.2 数据相关性分析

相关性系数有助于识别变量之间的冗余性,若两个变量高度相关,则提供的信息可能有重复的部分,这

样模型的解释性就会降低。故对各个特征值进行相关性分析,基于皮尔逊公式得到相关性矩阵,并总结于表 2 中。具体公式为:

表 2 相关性矩阵

	周边土地情况	植被情况	距离	高度	坡度	岩土性质
周边土地情况	1.000 000	0.417 723	0.417 180	0.089 077	0.424 033	-0.072 021
植被情况	0.417 723	1.000 000	0.176 112	0.043 600	0.308 152	0.039 637
距离	0.417 180	0.176 112	1.000 000	-0.097 407	0.328 826	0.025 528
高度	0.089 077	0.043 600	-0.097 407	1.000 000	0.246 009	-0.396 692
坡度	0.424 033	0.308 152	0.328 826	0.246 009	1.000 000	0.068 319
岩土性质	-0.072 021	0.039 637	0.025 528	-0.396 692	0.068 319	1.000 000

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} \quad (1)$$

式中, x 、 y 为分析的变量, n 为分析变量的总数, 以上矩阵中若两个变量之间的相关系数满足 $|r| > 0.8$, 则认为变量之间属于强相关性^[11]。从上述结果得出, 所有变量之间的皮尔逊系数绝对值均在 0.5 以下, 认定各个变量之间有较好的独立性。

2 支持向量回归模型

对比决策树^[12]、随机森林^[13]、LSTM^[14] 等其他机器学习算法, 利用 SVM 进行危险性预测具有准确、高效、可靠的优点, 文献[15]提出了六种应用于边坡稳定性预测的分类问题的综合人工智能方法, 并对六种方法进行对比。得到了优化后的 SVM 模型在预测结果可靠性上具有最优值的结论。

基于 scikit-learn 库构建 SVR 模型, SVR 可视作 SVM 在回归问题上的延伸和变种, SVR 的目标是找到一个超平面, 超平面两侧的样本点要尽可能相似, 如图 1 所示。

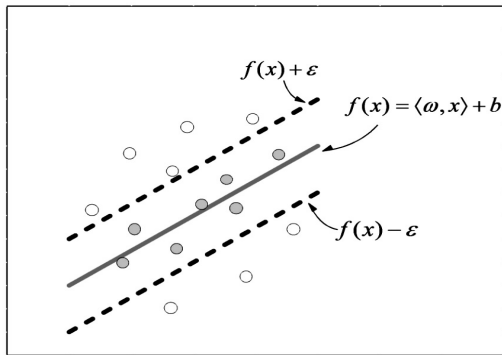


图 1 支持向量回归原理

SVR 的核心思想是最小化支持向量与超平面之间的间隔, 同时限制间隔不超过一个给定的容差值, 可表示为式 2, 约束条件可表示为式 3, 即以 $f(x)$ 为中心在最大容忍的间隔区内, 落入宽度为间隔区的训练样本可被认为是正确预测的。

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (2)$$

$$\begin{cases} y_i - (\omega \cdot x_i + b) \leq \varepsilon + \xi_i \\ (\omega \cdot x_i + b) - y_i \leq \varepsilon + \xi_i^* \end{cases} \quad (3)$$

其中, y 为实际值, ω 为决策面法向量, ε 为误差, ξ_i 和 ξ_i^* 为松弛变量, C 为惩罚参数。径向基(RBF)核函数可将数据映射到高维特征空间, 表达式为:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (4)$$

其中, x_i 和 x_j 为样本点, γ 为核函数参数。

对数据库进行处理后, 分别按照样本总量的 80% 和 20% 抽取训练集和测试集, 将 SVR 模型中的四个参数(分别为 γ : 核函数参数, tol: 模型停止迭代的误差容忍, C : 惩罚参数以及 ε : 目标变量的容忍度)作为一个数组, 并在每个参数取值范围内对数组进行 random 编码得到初始种群, 算法优化的目的即为得到最佳参数组合, 使得将模型训练至数据库与模型的拟合程度最好以及准确度最高。

3 基于遗传-模拟退火组合算法的 SVR 参数优化

遗传算法(GA)的特点是把握搜索过程的能力较强、收敛速度较快, 但存在过早收敛和局部寻优能力较差等问题; 而模拟退火算法(SA)能避免在收敛过程中陷入局部最优解, 并且局部搜索能力较强, 但不足之处是对于初始值相关参数比较敏感, 并且降温以及跳出局部最优解过程比较缓慢, 算法的收敛效率被降低。因此, 对上述两种算法取长补短, 可在 GA 算法的搜索过程中融合 SA 算法的思想以此对 SVR 参数进行优化, 提高模型性能。

为突出遗传-模拟退火组合算法的优越性, 设置其他三组相关的优化算法进行对比, 分别是单独用遗传算法优化 SVR 模型(GA-SVR)、单独用模拟退火算法优化 SVR 模型(SA-SVR)、用模拟退火-遗传组合算法优化 SVR 模型(SA-GA-SVR)。

其中 GA-SA-SVR 优化的步骤为: 将四个待优化参数的数组作为染色体形式, 每个基因在待优化参数取值范围内生成随机数, 将以上随机数数组作为初始化种群; 交换基因信息后, 计算种群的适应度值并排序, 再将排序后的新种群进行 SA 操作, 由产生函数从当前种群中产生位于解空间的新解, 重新计算适应度值后不仅接受最优解, 还以 Metropolis 准则以一定概率 P 跳出局部最优选择其他较差解, 直到迭代至遗传框架的最大迭代次数。

$$P(\text{accept}) = \begin{cases} 1 & [f(x)_{\text{old}} \geq f(x)_{\text{new}}] \\ \exp\left[\frac{f(x)_{\text{old}} - f(x)_{\text{new}}}{t}\right] & [f(x)_{\text{old}} < f(x)_{\text{new}}] \end{cases} \quad (5)$$

式中, t 代表当前的温度参数。组合算法主要参数设定为: 核函数为 RBF 高斯径向基核函数, 最大迭代次数 generation = 100, 种群规模 population size = 20, 变异率 mutation rate = 0.1, 初始温度 $T_0 = 1$, 降温速率 $\alpha = 0.9$, 终止温度 $T_i = 0.01$, 具体算法流程如图 2 所示。

而 SA-GA-SVR 组合算法则是在 SA 算法框架中加入遗传算法步骤, 具体流程为分别在每个待优化参

数取值范围内生成随机数,组成随机数组作为原始种群先进行模拟退火操作,产生新解后不仅接受最优解,还在一定概率 P 上接受较差解,再执行遗传算法,将新

的种群进行交叉变异得到新解,直到降至最低温度表示迭代完成。算法初始相关参数与 GA-SA-SVR 算法一致,流程如图 3 所示。

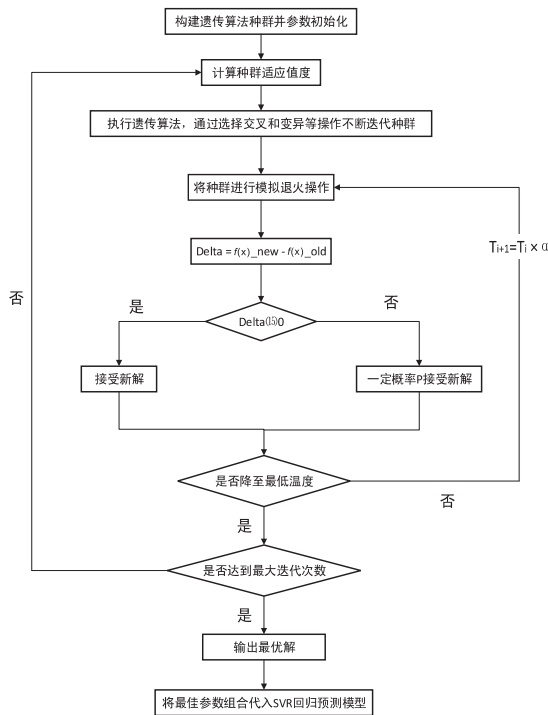


图 2 GA-SA-SVR 算法流程

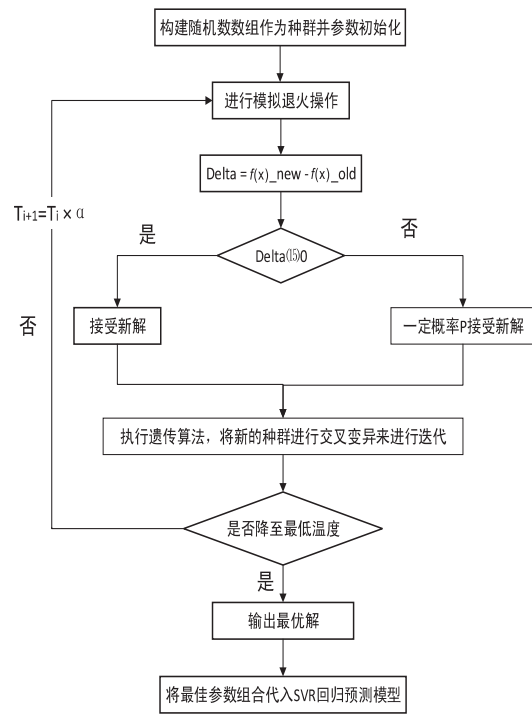


图 3 SA-GA-SVR 算法流程

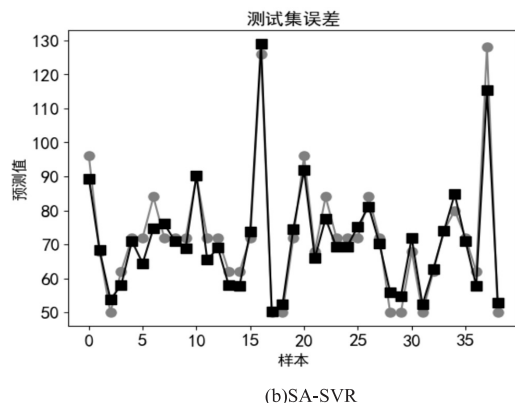
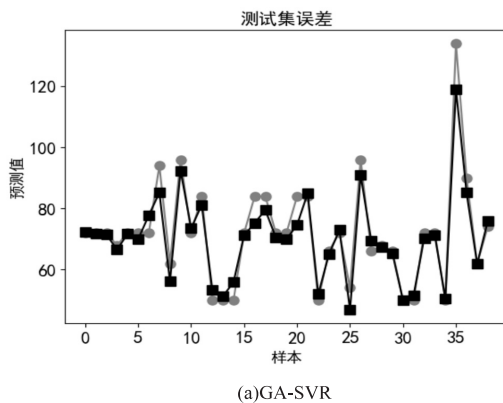
4 结果对比与分析

K 折交叉验证法是对样本测试集与训练集以 1 : (K - 1) 的比例进行划分,搭配为几种不同的训练集和测试集组合,交叉验证可以在有限的样本数据中考虑更多的耦合关系,在一定程度上规避陷入局部最优值和模型过度拟合的问题。 R^2 (mean) 即为 K 折交叉验证后测试集的平均拟合系数。为便于比较,将适应度函数设定为 $f(x) = 1 - R^2$ (mean), $f(x)$ 越小,表示模型整体平均拟合程度越高。其他两个评价指标分别为测试集和训练集当前的拟合系数 R^2 以及测试集的均方误差 (MSE)。 R^2 的取值范围为 [0, 1], 越趋近于 1, 代表模型拟合效果越好, MSE 的值越小代表预测值和真实值的差异越小,计算公式如下:

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (\bar{y} - y_i)^2} \tag{7}$$

其中, \hat{y}_i 与 y_i 分别代表真实值和预测值。将优化结果进行对比,结果分别是测试集真实值与预测值对比图 (图 4) 以及模型拟合图 (图 5), 适应度曲线对比图 (图 6), 并且额外设萤火虫-麻雀 (FA-SSA) 组合优化算法^[16]、麻雀 (SSA) 算法、猎食者 (HPO) 算法^[17]、鱼鹰 (OOA) 算法^[18] 作为对照组, 将指标对比结果总结于表 3 中。



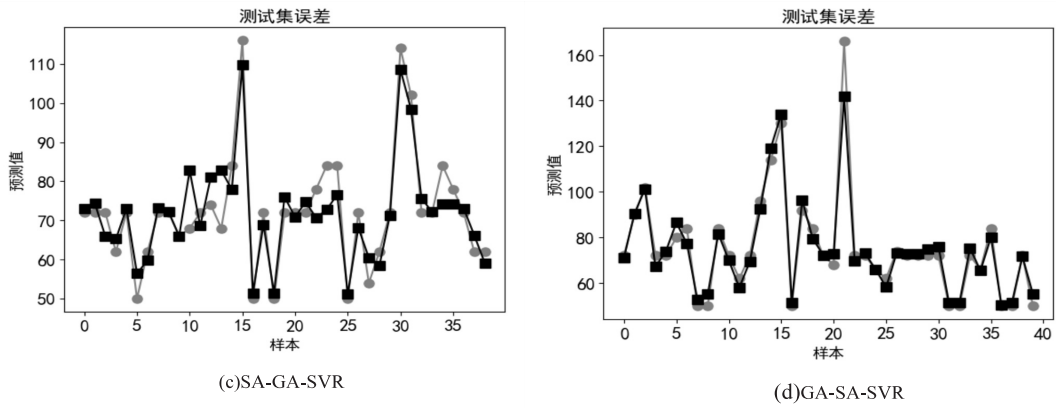


图 4 预测值与真实值误差对比

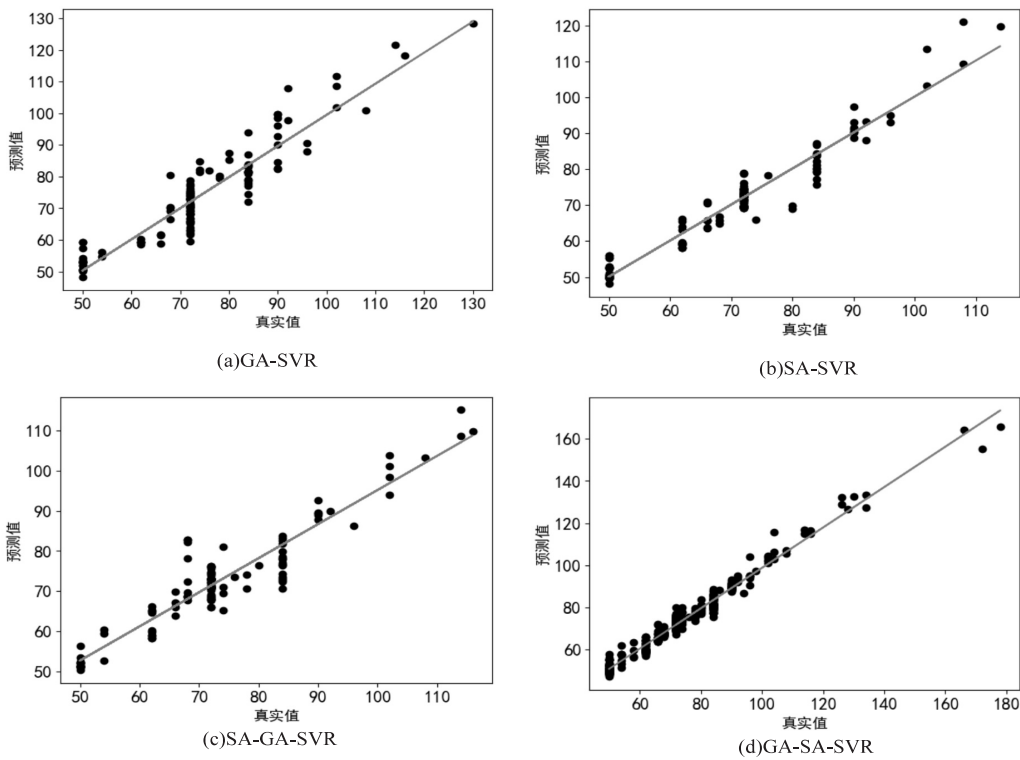


图 5 数据拟合图

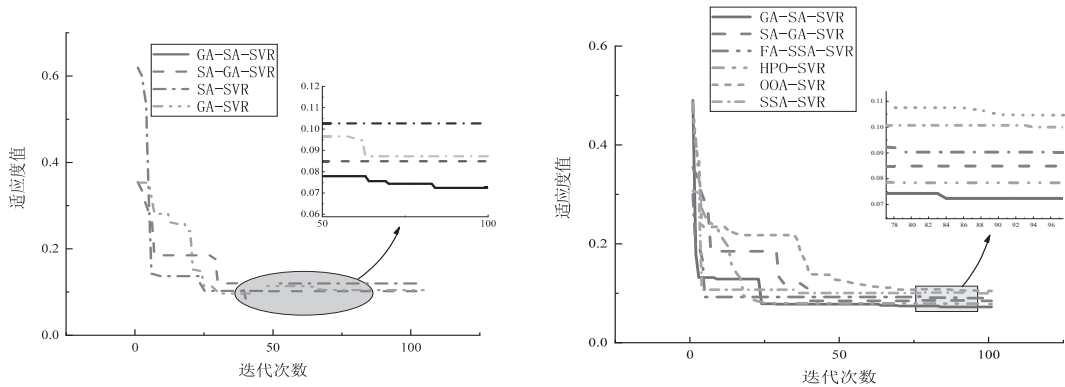


图 6 适应度值迭代曲线

结合适应度值迭代曲线,在 GA、SA 及其组合算法中,GA-SA-SVR 组合算法前期搜索速度最快并且后期也有小幅度收敛,符合遗传算法前期收敛速度快,模拟退火算法局部搜索能力强的特点,适应度值 $f(x)$ 最

终迭代至了 0.072 4,代表数据拟合效果最好。图 6 中圆形标记点代表真实值,方形标记点代表预测值,结合两点的吻合程度判断准确率,将均方误差 (MSE) 进行对比,其中 GA-SA-SVR 组合算法的 MSE 值最小,为

0.001 2, SA-SVR 算法次之, GA-SVR 算法最大, 代表 GA-SA-SVR 组合算法优化的预测模型准确率最高。在数据拟合图中, 标记点与拟合线离散化程度越高, 拟合线与 1:1 线偏差越大, 代表模型预测能力越差, 其中 GA-SA-SVR、SA-SVR 两种组合算法拟合效果较好, GA-SVR 拟合效果偏差, 从数据来看, GA-SA-SVR 模型测试集 R^2 值为 0.937 5, 数据与拟合线离散化程度最小。

表 3 相关指标对比

算法	R^2 (测试集)	R^2 (训练集)	MSE	$f(x)$
GA-SVR	0.788 5	0.930 7	0.004 2	0.073 2
SA-SVR	0.889 6	0.964 8	0.002 2	0.102 6
SA-GA-SVR	0.803 1	0.943 9	0.003 9	0.084 9
GA-SA-SVR	0.937 5	0.962 0	0.001 2	0.072 4
FA-SSA-SVR	0.851 8	0.957 7	0.003 1	0.092 3
SSA-SVR	0.230 7	0.058 0	0.025 1	0.100 5
HPO-SVR	0.953 2	0.921 1	0.001 7	0.078 4
OOA-SVR	0.676 6	0.954 9	0.006 3	0.104 7

在算法对比组中, HPO 优化算法的效果也很优秀, 对比指标接近 GA-SA 算法; 基于萤火虫改进的麻雀 (FA-SSA) 算法和麻雀 (SSA) 算法相对比, 也呈现出了更加优良的优化效果, 辅证了组合优化算法效果更好的结论。

综合以上分析得出结论, GA 和 SA 组合算法的优化效果要优于单一算法的优化效果, 其中 GA-SA-SVR 组合算法从拟合程度和准确率上更优。此模型对边坡进行客观值输入及分类即可得到危险系数, 评价方法更具科学性。

5 结束语

(1) 本课题依托某输电公司线路边坡隐患排查及状态评估数据, 建立边坡隐患数据库并代入 SVR 模型预测危险系数。此模型避免了原方法人为扣分的主观性, 且能更加智能、客观地对危险工况进行预测, 更有利于规避风险。

(2) 分别对遗传算法 (GA)、模拟退火算法 (SA)、模拟退火-遗传算法 (SA-GA)、遗传-模拟退火算法 (GA-SA)、萤火虫-麻雀 (FA-SSA) 组合优化算法、麻雀 (SSA) 算法、猎食者 (HPO) 算法、鱼鹰 (OOA) 算法的优化效果进行对比; 组合算法的优化效果要优于单一算法的优化效果, 其中遗传-模拟退火组合算法 (GA-SA) 的优化效果在准确率和拟合程度上更好, 测试集 R^2 为 0.937 5, MSE 值为 0.001 2, 适应度函数 $f(x)$ 值为 0.072 4。模型性能优于其他算法, 能更准确地得到预测值。

参考文献:

- [1] 刘传正, 陈春利. 中国地质灾害防治成效与问题对策[J]. 工程地质学报, 2020, 28(2): 375-383.
- [2] 陈新民. 特高压直流输电线路塔基边坡的稳定性分析[J]. 工程技术研究, 2020, 5(18): 119-120.
- [3] ALKAHTANI M, NAGGAR M H E. Bearing capacity of power transmission tower footings near cohesionless slopes[J]. Innovative Infrastructure Solutions, 2021, 6(4): 187.1-187.23.
- [4] 史耀凡, 栾元重, 于水, 等. 基于 PCA-GA-SVM 模型的地表下沉系数预测[J]. 矿业研究与开发, 2022, 42(2): 65-69.
- [5] 贾荣谷, 王学祥, 李育红, 等. 基于 GA-SVM 的边坡稳定性预测模型研究[J]. 地质灾害与环境保护, 2024, 35(2): 85-90.
- [6] 张建涛, 刘志祥, 张双侠, 等. 基于 WOA-RF 的边坡稳定性预测模型[J]. 高压物理学报, 2024, 38(3): 191-202.
- [7] 胡军, 王凯凯, 董建华. 基于 BCABC-SVM 的边坡稳定性预测[J]. 沈阳工业大学学报, 2016, 38(2): 222-227.
- [8] 金爱兵, 张静辉, 孙浩, 等. 基于 SSA-SVM 的边坡失稳智能预测及预警模型[J]. 华中科技大学学报: 自然科学版, 2022, 50(11): 142-148.
- [9] 郭辉, 付接递, 李振东, 等. 基于改进鲸鱼算法优化 SVM 参数和特征选择[J]. 吉林大学学报: 工学版, 2023, 53(10): 2952-2963.
- [10] 吴程昊, 莫路锋. 基于改进狮群算法的支持向量机参数优化方法[J]. 现代电子技术, 2022, 45(14): 79-83.
- [11] 王团辉, 王超, 吴顺川, 等. 基于 MISSA-SVM 模型的边坡稳定性预测及应用[J]. 中国安全科学学报, 2024, 34(4): 135-144.
- [12] JIAN Z, ENMING L, SHAN Y, et al. Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of casehistories[J]. Safety Science, 2019, 118: 505-518.
- [13] LUDWIG S, RAUSCH A, DEUTSCHER V, et al. Predicting problem-solving success in an office simulation applying N-grams and a random forest to behavioral process data[J]. Computers & Education, 2024, 218: 105093.
- [14] ZHU X, CHEN G, NI C, et al. Hybrid CNN-LSTM model driven image segmentation and roughness prediction for tool condition assessment with heterogeneous data[J]. Robotics and Computer-Integrated Manufacturing, 2024, 90: 102796.
- [15] 吕鹏. 基于信息技术的深凹露天矿高陡边坡稳定性综合分析研究[D]. 北京: 北京科技大学, 2019.
- [16] 张梅, 高犁, 陈万利. 基于 FASSA-SVM 的充电桩故障预测算法研究[J]. 电子测量技术, 2022, 45(12): 48-53.
- [17] 张晓虎, 宁环宇. 基于 ReliefF 和 HPO-SVM 的变压器故障检测方法[J]. 电工技术, 2023(17): 1-5.
- [18] 陈曦明, 张军伟, 张冉, 等. 一种改进鱼鹰优化算法及其应用[J]. 重庆理工大学学报: 自然科学, 2024, 38(3): 122-133.