

DOI: 10.20040/j.cnki.1000-7709.2023.20221947

基于 HHO-SVM 的水质预测模型及应用

宋治岑, 张顺平, 卢 敏

(云南农业大学水利学院, 云南 昆明 650201)

摘要: 支持向量机(SVM)在小样本模拟预测中具有优势,但其惩罚因子 C 和核函数参数 γ 的选取存在主观性,因此采用哈里斯鹰算法(HHO)对 SVM 的 C 、 γ 进行优化,建立基于 HHO-SVM 的水质预测模型,并应用于滇池草海西苑隧道断面的水质预测研究中。结果表明,HHO-SVM 水质预测模型的预测精度高于基于遗传算法的支持向量机模型(GA-SVM)和基于鲸鱼算法的支持向量机模型(WOA-SVM),验证了 HHO 算法优化 SVM 参数的可行性,表明 HHO-SVM 可用于水质预测中。

关键词: 哈里斯鹰算法;支持向量机;水质预测;应用

中图分类号: X52;TP18

文献标志码: A

文章编号: 1000-7709(2023)08-0070-03

1 引言

据统计,我国平均每 2~3 d 就会发生一起水污染事故,这使本来就不充足的可利用水资源越来越短缺,严重影响了我国的用水安全^[1]。水质预测能有效分析和了解河流、湖泊水环境质量的变化,对缓解水域环境污染、保护水资源和社会经济发展具有重要的现实意义^[2]。目前水质预测模型分为机理性水质预测模型和非机理性水质模型两大类^[3]。其中最常见非机理性水质预测方法有灰色理论法、人工神经网络和支持向量机^[4-6]。支持向量机(SVM)在处理小样本、非线性、高维复杂等问题中具有优势,但其惩罚因子 C 和核函数参数 γ 的选取存在主观性。当前大多采用智能算法对这两个参数进行优化,如遗传算法(GA)、海鸥算法(SOA)、鲸鱼算法(WOA)等^[7-9]。然而以上智能算法寻优时易陷入局部最优解,达不到最优结果^[10]。因此,本文尝试采用哈里斯鹰算法(HHO)优化支持向量机的惩罚因子 C 和核函数参数 γ ,并以此建立水质预测模型进行研究,旨在为优化支持向量机的参数和水质预测提供一种新方法。

2 基于 HHO-SVM 的水质预测模型

2.1 支持向量机的参数

支持向量机中的惩罚因子 C 和核函数参数 γ 对预测结果影响很大。其中惩罚因子 C 越小越易欠拟合, C 越大越易出现过拟合;径向基核函数参数 γ 越大越会出现过训练现象,而 γ 越小越会影响测试数据集的精确度^[11]。因此,利用哈里斯鹰算法优化 SVM 的 C 、 γ 两个参数。径向基(RBF)核函数公式为:

$$K(x_i, y_i) = \exp(-\gamma \|x - x_i\|)^2 \quad (1)$$

2.2 哈里斯鹰算法原理

哈里斯鹰算法(HHO)是通过观察哈里斯鹰独特的群体捕食行为提出的一种元启发式算法^[12]。该算法通过模拟哈里斯鹰群狩猎行为获得函数最优值,其优点为操作简单、调整的参数少且收敛能力强^[13]。哈里斯鹰算法具体有 3 个阶段。

(1)全局搜索阶段。哈里斯鹰群体内部分散度很高,个体通过长时间的等待和观察追踪猎物,以两种策略探测猎物。

$$X(t+1) = \begin{cases} X_{\text{rand}}(t) - r_1 | X_{\text{rand}}(t) - 2r_2 X(t) | & q \geq 0.5 \\ [X_r(t) - X_m(t)] - r_3 [l_b + r_4 (u_b - l_b)] & q < 0.5 \end{cases} \quad (2)$$

$$X_m(t) = \frac{1}{N} \sum_{i=1}^N X_i(t) \quad (3)$$

式中, X_{rand} 为群体中随机个体; X_r 为最优个体; X_m 为种群平均位置; q 、 r_1 、 r_2 、 r_3 、 r_4 均为 $[0,1]$

收稿日期: 2022-09-18, 修回日期: 2022-10-09

作者简介: 宋治岑(1995-),男,硕士研究生,研究方向为农业水土工程,E-mail: szc023@qq.com

通讯作者: 卢敏(1972-),女,副教授,研究方向为农业水土工程,E-mail: 1250028815@qq.com

随机数; u_b 、 l_b 分别为种群上下界; N 为种群数量。

(2) 全局搜索到局部开发的转换阶段。其公式为:

$$E = 2E_0(1 - t/T) \quad (4)$$

式中, E 为猎物的逃逸能力; E_0 为 $[-1, 1]$ 随机数; t 为迭代次数; T 为最大迭代次数。

(3) 局部开发阶段。哈里斯鹰基于猎物逃逸行为及自己的追逐策略, 进化为 4 种攻击策略。

① 轻柔围攻。当逃脱概率 $r \geq 0.5$ 且 $|E| \geq 0.5$ 时, 猎物仍有充足能力逃生, 这时哈里斯鹰使用轻柔围攻策略不断在猎物周围徘徊, 进而突袭:

$$X(t+1) = \Delta X(t) - E |JX_r(t) - X(t)| \quad (5)$$

$$\Delta X(t) = X_r(t) - X(t) \quad (6)$$

$$J = 2(1 - r_5) \quad (7)$$

式中, ΔX 为最优和当前个体差值; r_5 为 $[0, 1]$ 随机数; J 为猎物跳跃距离。② 强硬围攻。当 $r \geq 0.5$ 且 $|E| < 0.5$ 时, 猎物精力不充足未能逃脱, 哈里斯鹰选择直接攻击猎物:

$$X(t+1) = X_r(t) - E |\Delta X(t)| \quad (8)$$

③ 渐进式快速俯冲的轻柔围攻。 $r < 0.5$ 且 $|E| \geq 0.5$ 时, 哈里斯鹰进攻前渐进式快速俯冲的轻柔围攻:

$$Y = X_r(t) - E |JX_r(t) - X(t)| \quad (9)$$

此次更新的策略无效时, 执行第二种策略:

$$Z = Y + \text{SLF}(D) \quad (10)$$

$$X(t+1) = \begin{cases} Y & \text{fitness}(Y) < \text{fitness}(X(t)) \\ Z & \text{fitness}(Z) < \text{fitness}(X(t)) \end{cases} \quad (11)$$

式中, D 为空间维度; S 为 $1 \times D$ 的随机向量; $\text{LF}()$ 为 Levy 飞行函数; $\text{fitness}()$ 为适应度函数。④ 渐进式快速俯冲的强硬围攻。 $r < 0.5$ 且 $|E| < 0.5$ 时, 哈里斯鹰在突袭前采取强硬围攻, 缩小与猎物的平均距离:

$$Y = X_r(t) - E |JX_r(t) - X_m(t)| \quad (12)$$

$$Z = Y + \text{SLF}(D) \quad (13)$$

$$X(t+1) = \begin{cases} Y & \text{fitness}(Y) < \text{fitness}(X(t)) \\ Z & \text{fitness}(Z) < \text{fitness}(X(t)) \end{cases} \quad (14)$$

2.3 基于 HHO-SVM 水质预测模型的构建

先选用支持向量机 (SVM) 建立水质预测模型, 其中 SVM 的核函数采用径向基 (BRF) 核函数, 再利用哈里斯鹰算法 (HHO) 对支持向量机的惩罚因子 C 和核函数参数 γ 寻优, 在此基础上建立基于哈里斯鹰算法的支持向量机水质预测模型 (HHO-SVM)。具体步骤如下。

步骤 1 确定训练集和测试集, 设置参数 C 和 γ 的取值范围, 并进行数据归一化处理。

步骤 2 确定待优化适应度函数, 公式为:

$$\text{fitness}(C, \gamma) = \sum_{i=1}^l \left| \frac{x_i^* - x_i}{x_i^*} \right| \quad (15)$$

式中, x_i^* 为第 i 个高锰酸盐指数实测值; x_i 为第 i 个高锰酸盐指数预测值。

步骤 3 设置哈里斯鹰种群数为 N 、最大迭代次数为 T 。设当前迭代次数为 1, 初始化哈里斯鹰种群位置。

步骤 4 计算每头鹰的适应度值, 计算鹰群的适应度, 确定最优个体位置。

步骤 5 判断 $|E|$ 和 r 。若 $r \geq 0.5$ 且 $|E| \geq 0.5$ 利用式 (5) 更新位置, 若 $r \geq 0.5$ 且 $|E| < 0.5$ 利用式 (8) 更新位置, 若 $r < 0.5$ 且 $|E| \geq 0.5$ 利用式 (11) 更新位置, 若 $r < 0.5$ 且 $|E| < 0.5$ 利用式 (14) 更新位置。

步骤 6 计算更新的哈里斯鹰群个体适应度值。若新鹰群个体适应度优于前代, 新的个体位置替代原位置, 否则保留原位置。

步骤 7 判断是否达到 T , 若是, 则输出当前最优个体位置; 若否, 则重新判断猎物 E 、 r 。

步骤 8 把 HHO 算法优化的最优个体位置即最优 C 、 γ 代入 SVM 水质模型中进行建模预测。

3 应用

选取滇池草海西苑隧道断面 2013~2015 年的 120 组水质实测周数据进行分析, 包括 pH 值、溶解氧 (DO)、高锰酸盐 (COD_{Mn})、氨氮 (NH₃-N) 4 个指标, 数据来源于中国环境监测总站的水质自动监测周报。

3.1 模型的输入与输出

先采用熵权法^[14]算出 4 个水质指标的权重, 确定模型的输入、输出因子, 以此求得 pH 值、DO、NH₃-N、COD_{Mn} 的权重分别为 0.302 8、0.210 5、0.046 4、0.440 2。由此可知, 高锰酸盐指数为 0.440 2, 在 4 个水质指标中权重最大。同时根据中国环境监测总站的水质自动监测周报, 高锰酸盐指数为该断面主要污染指数。所以建立模型预测高锰酸盐指数。

对该断面下一周的高锰酸盐指数与之前几周的进行相关性分析, 选取相关性显著的前 1~5 周高锰酸盐指数作为影响因子, 预测下一周的高锰酸盐指数, 以前 100 组作为训练集, 后 20 组作为测试集。本文通过所建 HHO-SVM 模型预测高锰酸盐指数, 并与 WOA-SVM、GA-SVM 水质模型预测结果进行对比验证。

3.2 参数设置

将归一化后的影响因子作为 HHO-SVM、

WOA-SVM、GA-SVM 三个模型的输入,将高锰酸盐指数作为输出。设置惩罚因子 C 、核函数参数 γ 取值区间为 $[0.1, 100]$,不敏感系数 $\epsilon = 0.1$,种群数目 $N = 20$,最大迭代次数 $T = 100$ 。三种模型连续运行 20 次后,得到的最优 C 、 γ 结果见表 1。

表 1 三种模型的参数设置和最优 C 、 γ 值比较

Tab. 1 Parameter setting and optimal C , γ values of three models

指数	GA-SVM	WOA-SVM	HHO-SVM
C	40.345 3	19.476 5	28.851 2
γ	0.250 1	0.101 0	0.103 2

3.3 结果与分析

利用 HHO-SVM、WOA-SVM、GA-SVM 水质预测模型对滇池草海西苑隧道断面的高锰酸盐指数进行预测。预测值见图 1,具体预测值和其相对误差见表 2。

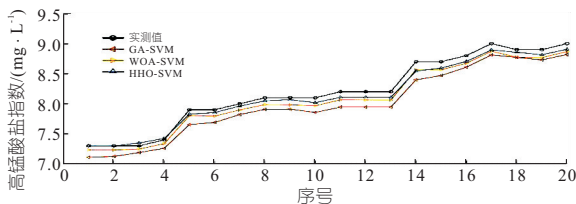


图 1 高锰酸盐预测结果对比

Fig. 1 Prediction of COD_{Mn}

表 2 高锰酸盐指数预测值和相对误差对比

Tab. 2 Prediction and relative error of COD_{Mn}

序 号	实 测 值	GA-SVM		WOA-SVM		HHO-SVM	
		预 测 值	相 对 误 差	预 测 值	相 对 误 差	预 测 值	相 对 误 差
1	7.3	7.110 0	-2.602 7	7.232 1	-0.930 1	7.300 7	0.009 6
2	7.3	7.123 8	-2.413 7	7.231 5	-0.938 4	7.298 9	-0.015 1
3	7.3	7.188 0	-1.534 2	7.246 3	-0.735 6	7.348 2	0.660 3
4	7.4	7.2598	-1.894 6	7.336 6	-0.856 8	7.423 3	0.314 9
5	7.9	7.653 6	-3.119 0	7.801 8	-1.243 0	7.826 0	-0.936 7
6	7.9	7.688 8	-2.673 4	7.794 0	-1.341 8	7.853 6	-0.587 3
7	8.0	7.820 7	-2.241 3	7.897 5	-1.281 3	7.961 8	-0.477 5
8	8.1	7.906 3	-2.391 4	7.986 0	-1.407 4	8.048 1	-0.640 7
9	8.1	7.909 7	-2.349 4	7.981 5	-1.463 0	8.066 4	-0.414 8
10	8.1	7.856 4	-3.007 4	7.967 3	-1.638 3	8.018 8	-1.002 5
11	8.2	7.944 1	-3.120 7	8.065 2	-1.643 9	8.107 6	-1.126 8
12	8.2	7.944 8	-3.112 2	8.062 0	-1.682 9	8.109 4	-1.104 9
13	8.2	7.944 3	-3.118 3	8.061 1	-1.693 9	8.105 3	-1.154 9
14	8.7	8.397 7	-3.474 7	8.563 9	-1.564 4	8.543 0	-1.804 6
15	8.7	8.468 6	-2.659 8	8.564 4	-1.558 6	8.592 1	-1.240 2
16	8.8	8.607 3	-2.189 8	8.672 7	-1.446 6	8.704 3	-1.087 5
17	9.0	8.812 7	-2.081 1	8.873 0	-1.411 1	8.894 2	-1.175 6
18	8.9	8.769 8	-1.462 9	8.773 3	-1.423 6	8.856 0	-0.494 4
19	8.9	8.729 0	-1.921 3	8.766 6	-1.498 9	8.813 9	-0.967 4
20	9.0	8.819 2	-2.008 9	8.868 1	-1.465 6	8.906 5	-1.038 9

注:实测值、预测值单位均为 mg/L ;相对误差单位为%。

由图 1 可知,HHO-SVM 模型的预测值相比 GA-SVM、WOA-SVM 模型更接近实测值。由表 2 可知,GA-SVM、WOA-SVM、HHO-SVM 的相对误差绝对值最大值分别为 3.474 7%、1.693 9%、1.804 6%。3 种模型的预测值绝大多数都小于实测值,说明 3 种模型的预测结果相比实测值更加乐观。

为进一步分析 3 种模型预测的精准度,用均

方误差 M_{MSE} 、平均绝对值误差 M_{MAE} 和平均绝对百分比误差 M_{MAPE} 评价模型预测结果。结果表明,GA-SVM、WOA-SVM、HHO-SVM 的均方误差分别为 0.043 1、0.013 5、0.006 1;平均绝对值误差分别为 0.202 3、0.112 8、0.068 3;平均绝对百分比误差分别为 2.47%、1.36%、0.81%。由此可知,HHO-SVM 的 M_{MSE} 、 M_{MAE} 、 M_{MAPE} 明显小于 GA-SVM、WOA-SVM 模型,说明 HHO-SVM 模型预测精度高于 GA-SVM、WOA-SVM 模型,可认为其回归预测效果更好。而 WOA-SVM 模型的 3 种误差均小于 GA-SVM 模型,说明 WOA-SVM 模型的预测效果比 GA-SVM 模型更好。

4 结 论

针对 SVM 的预测结果受参数选取主观性影响较大的问题,利用 HHO 算法对 SVM 的惩罚因子 C 和核函数参数 γ 进行寻优,提出 HHO-SVM 水质预测模型,并与 WOA-SVM、GA-SVM 水质模型进行对比验证。实例结果表明,HHO-SVM 模型预测结果比 WOA-SVM、GA-SVM 模型更为精确,验证了 HHO 算法优化 SVM 参数可行,表明 HHO-SVM 可用于水质预测中。

参 考 文 献:

- [1] 李杰. 漳河上游河流水质评价及预测[D]. 邯郸: 河北工程大学, 2018.
- [2] 王兆波. 长春市典型水库水质综合评价及预测研究[D]. 长春: 长春工程学院, 2018.
- [3] 吴岳玲. 水质综合评价及预测研究进展[J]. 安徽农业科学, 2020, 48(2): 23-26.
- [4] 王健, 向峰, 邱飞, 等. 水质预测模型研究进展[J]. 环境科学导刊, 2018, 37(4): 63-67.
- [5] 岳丹丹, 梁海涛, 王天慧, 等. BP 神经网络模型在西鞍山铁矿地下水水质评价中的应用[J]. 环境监测管理与技术, 2016, 28(4): 23-26.
- [6] 陈荣. 东江湖流域水质变化趋势分析与水质评价[D]. 长沙: 中南林业科技大学, 2016.
- [7] 马创, 王尧, 李林峰. 基于遗传算法与支持向量机的水质预测模型[J]. 重庆大学学报, 2021, 44(7): 108-114.
- [8] 孙毅, 石墨, 单葆国, 等. 基于粒子群优化支持向量机的电能替代潜力分析方法[J]. 电网技术, 2017, 41(6): 1767-1771.
- [9] 张学坤. 基于 SOA-SVM 和 SOA-BP 模型的溶解氧预测[J]. 人民珠江, 2021, 42(4): 99-104.
- [10] 周有荣, 王凯. 改进鲸鱼算法优化混合核支持向量机在径流预测中的应用[J]. 中国农村水利水电, 2020(7): 50-53.

4 结论

a. 基于“幸福河”的概念,以齐齐哈尔市为例,构建了幸福河评价体系和评价模型,划分 5 个评价等级,对绰尔河、雅鲁河、音河、阿伦河、诺敏河、润津河 6 条中小型河流进行了幸福河评价,确定齐齐哈尔市河流整体幸福等级为Ⅲ级。

b. 水质达标率不稳定,可从监测体系、农业污染等方面进行整治;生物多样性出现问题,可从机构主治、自然恢复等两个方面改善。

参考文献:

[1] 习近平. 在黄河流域生态保护和高质量发展座谈会上的讲话[J]. 中国水利, 2019(20): 1-3.

[2] 左其亭,郝明辉,马军霞,等. 幸福河的概念、内涵及判断准则[J]. 人民黄河,2020,42(1):1-5.

[3] 唐克旺. 对“幸福河”概念及评价方法的思考[J]. 中

国水利, 2020(6):15-16.

[4] 靳春玲,李燕,贡力,等. 基于 UMT 模型的幸福河绩效评价及障碍因子诊断[J]. 中国环境科学,2022,42(3):1466-1476.

[5] 水利部河长办. 河湖健康评价指南(试行)[EB/OL]. <http://hhs.mwr.gov.cn/tzgg/202008/P020200820401648281363.pdf>.

[6] 黄德才,胥琳. AHP 法中判断矩阵的比例标度构造法[J]. 控制与决策,2002,17(4):484-486.

[7] 安慧,孟益,蒙锦涛. 基于 AHP-GEM-FCA 模型的水电 PPP 项目脆弱性评价[J]. 水电能源科学,2018,36(9):143-147.

[8] 颜山凯,黄显峰,方国华,等. 基于熵权 TOPSIS 法的梯级水电站补偿效益分摊模型研究[J]. 水电能源科学,2021,39(9):76-80,28.

[9] 齐齐哈尔市人民政府. 2021 年齐齐哈尔市国民经济和社会发展统计公报[EB/OL]. <http://www.flej.gov.cn/yuetongjishuju/10382.html>.

Comprehensive Evaluation of Happiness River Based on Combined Weighting TOPSIS in Qiqihar Areas

HAO Xu-dong¹, WANG Li-quan¹, LI Tie-nan², ZHANG Ning², MENG Hao-jie¹

(1. School of Hydraulic and Electric-power, Heilongjiang University, Harbin 150080, China;

2. Heilongjiang Province Hydraulic Research Institute, Harbin 150080, China)

Abstract: Based on the understanding of the concept of happiness river, taking 6 rivers with moderate river scale in Qiqihar Areas as an example, the TOPSIS evaluation model is constructed by combining the proportional scale AHP and AEM to carry out the evaluation on the happiness river, whose goal was divided into 5 guiding layers and 20 indicator layers. The evaluation results show that the overall happiness of the rivers in Qiqihar Areas is grade Ⅲ, which the happiness of Yalu River is grade Ⅳ, the happiness of Alen River, Nuomin River, Yin River and Runjin River is grade Ⅲ, and the happiness of Chaoer River is grade Ⅱ. In addition, the common problems of poor evaluation grade in the results were analyzed and the treatment suggestions were put forward.

Key words: happiness river; combination weighting approach; TOPSIS; Qiqihar Areas

(上接第 72 页)

[11] 宋杰,吕谋,郝晨西,张士官. 基于 PSO-SVM 模型的供水管网漏损诊断[J]. 水电能源科学,2020,38(7): 122-125.

[12] ALI ASGHAR HEIDARI, SEYEDALI MIRJALI-LI, HOSSAM FARIS, et al. Harris hawks optimization: Algorithm and applications[J]. Future genera-

tion computer systems,2019,97:849-872.

[13] 吴丁杰,温立书. 一种基于哈里斯鹰算法改进的 BP 神经网络[J]. 网络安全技术与应用,2022(1):38-40.

[14] 职璐爽,薛惠锋. 基于熵权法的城市水资源脆弱性研究——以广东省为例[J]. 水土保持通报,2018,38(5):322-329.

Water Quality Prediction Model Based on HHO-SVM and Its Application

SONG Zhi-cen, ZHANG Shun-ping, LU Min

(School of Water Conservancy, Yunnan Agricultural University, Kunming 650201, China)

Abstract: Support Vector Machine (SVM) has advantages in small sample simulation prediction, but there is subjectivity in the selection of penalty factor C and kernel function parameter γ in SVM. Therefore, the Harris Hawks Optimization (HHO) algorithm was used to optimize C and γ in the SVM. And then the HHO-SVM mode was established to predict water quality in the Xiyuan tunnel section of Lake Dianchi Caohai. The results show that the prediction accuracy of the water quality prediction model based on HHO-SVM is higher than that of the SVM based on genetic algorithm (GA-SVM) and the SVM based on whale optimization algorithm (WOA-SVM). It is proved that the HHO is feasible to optimize the parameters in SVM, and HHO-SVM can be used in water quality prediction.

Key words: Harris Hawks Optimization; Support Vector Machine; water quality prediction; application