

DOI: 10. 20040/j. cnki. 1000-7709. 2023. 20222434

基于混合策略改进 SSA-SVM 的工业用水量预测

曹敬椿¹, 卢 敏²

(1. 云南省水利水电勘测设计研究院, 云南 昆明 650000; 2. 云南农业大学水利学院, 云南 昆明 650000)

摘要: 为实现工业用水量预测和提高预测精度, 引入混合策略对麻雀搜索算法(SSA)进行优化以提高全局搜索能力, 采用改进后的麻雀搜索算法(ISSA)优化支持向量机(SVM)参数, 建立基于混合策略改进麻雀搜索算法优化的支持向量机模型(ISSA-SVM), 并将该模型应用于宁夏的工业用水量预测中。结果表明, ISSA-SVM模型寻优速度快、训练和预测精度高, 将其应用于工业用水量预测中具有良好的适用性和可行性。

关键词: 支持向量机; 工业用水量预测; 麻雀搜索算法; 自适应惯性权重; 纵横交叉策略

中图分类号: TV213.9; TU991.31; TP183 **文献标志码:** A **文章编号:** 1000-7709(2023)09-0028-04

1 引言

水资源是人类生存和社会经济发展所需的重要资源^[1], 随着我国工业化规模不断壮大, 工业用水量比重不断上升, 因而进行工业用水量预测对促进经济、社会、生态的发展具有重大意义^[2]。目前常用的用水量预测方法主要有时间序列法、灰色预测法、回归分析法、神经网络法及其他组合模型预测方法。近年来机器学习算法在用水量预测中得到了良好的应用, 武风波等^[3]采用遗传算法优化 BP 神经网络模型预测了城市用水量; 赵晓东^[4]将遗传算法和模拟退火算法相结合用于优化支持向量机模型, 并预测了园区短期的需水量。用水量及其影响因素之间存在非线性关系, 而支持向量机(SVM)在解决小样本、非线性问题上具有显著优势, 但对其自身参数 C 和 g 的选取是影响 SVM 性能好坏的关键, 故可采用优化算法优化其参数。麻雀搜索算法(SSA)相较于遗传算法、粒子群优化算法等其他传统算法, 其具有设置参数少、局部搜索能力强等优势, 且在多类基准函数上的求解能力均优于其他传统算法, 但该算法在迭代后期易出现局部最优^[5,6]。为此, 本文引入混合策略改进的麻雀搜索算法(ISSA)对支持向量机参数自动寻优, 建立了 ISSA-SVM 模型, 并将该模型应用于宁夏的工业用水量预测中, 取得了较好的效果, 可供借鉴。

2 基于 ISSA-SVM 的预测模型

2.1 改进的麻雀搜索算法

2.1.1 自适应惯性权重策略

发现者作为帮助麻雀种群获取优质食物的领导核心, 其应具备优秀的搜索机制。当发现者发现优质食物后, 种群中的大量个体会迅速向优质食物位置靠近, 扰乱发现者的搜索进程, 从而导致其搜索能力受干扰, 全局和局部搜索能力均严重下降, 最终陷入局部最优^[7]。为此, 本文引入了自适应惯性权重因子 ω 控制发现者的搜索能力和范围, 防止其陷入局部最优。自适应惯性权重因子 ω 计算公式为:

$$\omega = \exp\left(1 - \frac{t_{\max} + t}{t_{\max} - t}\right) \quad (1)$$

式中, t_{\max} 为最大迭代次数; t 为当前迭代次数。

权重因子 ω 根据迭代次数增加而作出相应的非线性变化, 通过自适应调整以保证发现者能够准确寻找到最优解。迭代前期 ω 较大, 发现者具有较好的全局搜索能力和范围, 算法不易过早收敛。迭代后期 ω 较小, 搜索范围自适应缩小, 发现者专注于某区域寻找最优解, 能够有效保证算法优化精度。本文引入自适应惯性权重因子 ω 得到改进后的发现者位置更新公式为:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \omega & R_2 < S_T \\ X_{i,j}^t + QL\omega & R_2 \geq S_T \end{cases} \quad (2)$$

收稿日期: 2022-11-18, 修回日期: 2022-12-09

作者简介: 曹敬椿(1999-), 男, 硕士研究生, 研究方向为水文与水资源工程, E-mail: 2457902757@qq.com

通讯作者: 卢敏(1972-), 女, 博士、副教授、硕导, 研究方向为水文与水资源工程, E-mail: 1250028815@qq.com

式中, $X_{i,j}^t$ 为 j 维情况下迭代次数运行到 t 次时第 i 只麻雀的位置信息; R_2 为预警值, $R_2 \in [0, 1]$; S_T 为安全值, $S_T \in [0.5, 1]$ (当 $R_2 < S_T$ 时, 发现者处于安全区域, 可进行大范围搜索觅食; 当 $R_2 \geq S_T$ 时, 种群中有麻雀意识到捕食者正在靠近, 发现者需立即飞往其他安全区域进行觅食); Q 为满足正态分布的随机数; L 为元素均为 1 的 $1 \times m$ 矩阵。

2.1.2 纵横交叉策略

在麻雀搜索算法中, 随着迭代次数增加, 麻雀种群会不断向最优个体周围聚集, 部分未知区域无法得到探索, 容易出现局部最优化现象, 从而导致过早收敛^[5,8]。为防止出现局部最优, 将纵横交叉策略用于改进麻雀种群中预警者的搜索方式, 在保证算法收敛速度较快的同时, 提高了全局搜索能力和求解精度。

(1) 横向交叉。横向交叉是指在同一维度下对麻雀种群中的个体进行两两交叉运算, 使个体之间能够交互信息增加多样性以提高全局搜索能力和精度。通过随机选取两个父代个体配对, 并在 d 维度下进行交叉操作。横向交叉后的预警者位置更新公式为:

$$M_{X_{i,d}^t} = s_1 X_{i,d}^t + (1 - s_1) X_{j,d}^t + c_1 (X_{i,d}^t - X_{j,d}^t) \quad (3)$$

$$M_{X_{j,d}^t} = s_2 X_{j,d}^t + (1 - s_2) X_{i,d}^t + c_2 (X_{j,d}^t - X_{i,d}^t) \quad (4)$$

式中, $M_{X_{i,d}^t}$ 、 $M_{X_{j,d}^t}$ 分别为预警者 $X_{i,d}^t$ 、 $X_{j,d}^t$ 在第 d 维横向交叉后产生的子代个体; s_1 、 s_2 分别为 $[0, 1]$ 间的随机数; c_1 、 c_2 均为 $[-1, 1]$ 间的随机数。

将产生的子代个体与父代个体进行对比, 根据适应度值保留最优个体。

(2) 纵向交叉。麻雀种群中的某些个体往往会在迭代后期的某一维度下出现局部最优化现象, 从而导致收敛精度降低, 无法得到全局最优解。故在进行横向交叉运算后, 需对产生的新个体进行纵向交叉运算, 即对所有新个体在不同维度下进行交叉运算, 在保证原有能力不受影响的情况下减小算法陷入局部最优解的可能。纵向交叉后的预警者位置更新公式为:

$$M'_{X_{i,d}^t} = s X_{i,d_1}^t + (1 - s) X_{i,d_2}^t \quad (5)$$

式中, $M'_{X_{i,d}^t}$ 为个体 $X_{i,d}^t$ 在 d_1 和 d_2 维度下纵向交叉产生的子代个体; $s \in [0, 1]$ 为随机数。

将产生的新个体与父代个体进行对比, 保留适应度值较高的个体。

通过纵横交叉操作后, 陷入局部最优的个体

通过学习利用各维度得到的有用信息, 将有更大机会逃离出局部最优的限制, 从而提高算法的全局搜索能力和寻优精度。

2.2 支持向量机

支持向量机 (SVM) 主要用于解决小样本、非线性的高维数问题, 该方法将低维空间中复杂非线性问题映射转变为高维空间中的线性问题, 通过建立回归模型进行求解, 其拟合函数^[9,10]为:

$$f(x) = [\omega^T \varphi(x)] + b \quad (6)$$

式中, ω 为待求问题的权重; $\varphi(x)$ 为非线性映射函数; b 为待求问题的阈值。

根据结构风险最小化原则, 可将支持向量机回归模型的目标函数转化为:

$$\begin{cases} \min \frac{1}{2} \|\omega\|^2 + \frac{1}{2} C \sum_{i=1}^n \xi_i^2 \\ \text{s. t. } y_i (\omega^T x_i + b) \geq 1 - \xi_i \end{cases} \quad (7)$$

构造拉格朗日函数将原有问题转为新对偶问题并引入核函数求解, 得到预测模型的决策函数为:

$$f(x) = \sum_{i=1}^n \alpha_i K(x_i, x) + b \quad (8)$$

式中, n 为样本数; α_i 为拉格朗日乘子; $K(x_i, x)$ 为核函数。

2.3 ISSA-SVM 预测模型构建

建立基于 ISSA-SVM 的工业用水量预测模型, 具体步骤如下。

步骤 1 根据问题的样本数合理选取模型的训练集和测试集, 并进行归一化处理。

步骤 2 初始化麻雀种群数 P 和最大迭代次数 t_{\max} , 设置 SVM 参数范围 (C 、 g 的上下限), 适应度函数采用均方误差 M_{MES} 。

步骤 3 将训练集代入模型, 引入两种策略加快算法寻优能力, 计算个体适应度值并记录当前最优位置信息, 通过全局最优位置信息确定 SVM 最优参数 C 、 g 。

步骤 4 将测试集代入训练好的预测模型, 输出工业用水量的预测结果, 采用均方误差、相对误差等指标对预测值和真实值进行比较, 验证 ISSA-SVM 预测模型的准确性和可靠性。

3 实例应用

3.1 概况

以文献[11]中的宁夏工业用水量预测为例, 对宁夏 2004~2019 年工业用水量及其主要影响因素相关数据进行研究。工业用水量主要影响因素包括制造业增加值占工业增加值比重 (C_1)、能

源消费总量(C_2)、火力发电量(C_3)、柴油消耗量(C_4)、天然气消耗量(C_5)和电力消耗量(C_6)，数据具体见表 1。

表 1 训练样本和测试样本

Tab. 1 Training samples and test samples

年份	C_1 /%	C_2 / 10^8 t	C_3 / 10^8 kW·h	C_4 / 10^4 t	C_5 / 10^8 m ³	C_6 / 10^8 kW·h	工业用水 量/ 10^8 m ³
2004	0.60	0.16	243.99	68.94	6.77	272.83	3.22
2005	0.55	0.18	295.19	63.04	6.63	303.00	3.45
2006	0.60	0.22	267.31	78.14	7.94	377.85	3.47
2007	0.60	0.26	432.45	86.70	8.99	439.78	3.52
2008	0.58	0.29	449.28	96.51	11.00	439.62	3.33
2009	0.58	0.38	454.60	98.57	11.98	462.96	3.68
2010	0.53	0.46	551.36	105.82	15.48	547.00	4.12
2011	0.52	0.54	909.42	104.04	18.58	724.54	4.64
2012	0.52	0.58	954.70	111.32	20.48	741.80	4.86
2013	0.53	0.60	1011.60	117.91	19.57	811.20	5.01
2014	0.52	0.59	1041.60	123.38	17.88	848.75	4.98
2015	0.53	0.56	1015.83	122.75	20.65	878.00	4.35
2016	0.54	0.52	953.56	124.03	22.40	886.91	4.39
2017	0.53	0.57	1144.39	127.55	22.27	978.30	4.52
2018	0.53	0.60	1367.77	117.63	20.96	1064.85	4.34
2019	0.53	0.58	1443.87	120.58	25.35	1084.00	4.43

3.2 模型参数设置

将制造业增加值占工业增加值比重等主要影响因素作为 ISSA-SVM 预测模型的输入变量，工业用水量作为输出变量。根据表 1 对训练样本和测试样本进行划分，以宁夏 2004~2015 年的数据作训练集用于 ISSA-SVM 预测模型的参数寻优，2016~2019 年的数据作测试集用于验证该模型合理可靠性。

利用混合策略改进后的 SSA 算法对 SVM 参数进行寻优，设置初始种群数 $P=5$ ，最大迭代次数 $t_{max}=100$ ，麻雀种群中发现者比例为 0.2，加入者比例为 0.8，预警值取 0.8， C 和 g 的取值范围均设为 $[0.01, 1.000]$ ，SVM 核函数采用径向基函数。

通过将训练集中的 12 组数据输入模型进行训练并与真实值进行比较，结果见图 1。由图 1 可知，预测值和真实值的均方误差 M_{MSE} 为 0.003，拟合度系数为 0.994，满足精度要求。

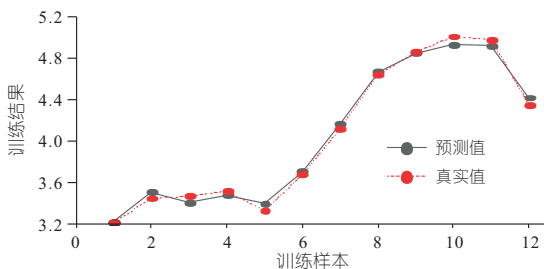


图 1 训练集训练结果

Fig. 1 Training set training result

对 ISSA 算法和初始 SSA 算法优化 SVM 的适应度值作对比，ISSA 算法和 SSA 算法的适应度值曲线见图 2。由图 2 可知，SSA 算法不仅收敛速度较慢，还陷入了局部最优，而 ISSA 算法收敛速度快、寻优精度高，其在迭代到 15 次时模型适应度已基本趋于平稳且最优，ISSA-SVM 模型优于 SSA-SVM 模型，ISSA 算法有效改善了初始算法的缺点。

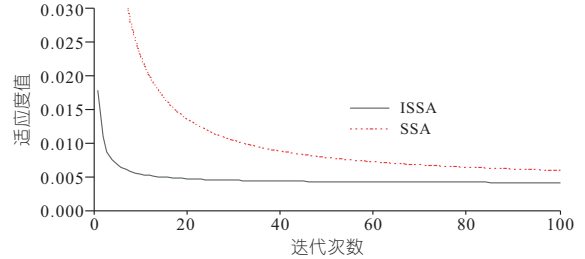


图 2 适应度值变化曲线图

Fig. 2 Change curve of fitness value

经过混合策略改进的 SSA 算法优化得到 SVM 最优参数 $C=171.7672$ 、 $g=0.0100$ ，将最优参数带入 SVM 得到训练好的 ISSA-SVM 预测模型。

3.3 预测结果与分析

将测试集代入训练好的 ISSA-SVM 模型中进行预测并与真实值对比，测试集预测结果见图 3。由图 3 可知，预测值和真实值的均方误差 M_{MSE} 为 0.001，拟合度系数为 0.811。

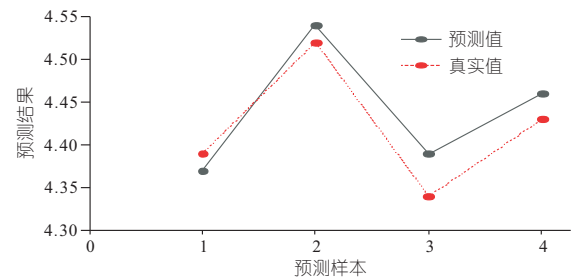


图 3 测试集预测结果图

Fig. 3 Test set prediction result diagram

分别采用 ISSA-SVM 模型和文献[11]中的模型预测了宁夏 2004~2019 年的工业用水量，结果见表 2。由表 2 可知，从训练集来看，ISSA-SVM 模型相对误差绝对值最大值为 2.10%，最小值为 0.21%，平均相对误差为 1.19%，说明该模型拟合精度较高；从测试集来看，ISSA-SVM 相对误差绝对值最大值为 1.15%，最小值为 0.44%，平均相对误差为 0.68%，略优于文献[11]中的模型计算结果，说明该模型具有较好的预测能力。综上可知，ISSA-SVM 模型不仅预测精度高，且 ISSA-SVM 模型寻优速度快、综合性能好，可用于工业用水量预测中。

表 2 不同模型预测结果对比

Tab. 2 Comparison of prediction results of different models

年份	真实值 /10 ⁸ m ³	文献[11]模型 预测值/10 ⁸ m ³	相对误 差/%	ISSA-SVM 预 测值/10 ⁸ m ³	相对误 差/%
2004	3.22	3.25	0.94	3.22	0.22
2005	3.45	3.41	-1.17	3.51	1.74
2006	3.47	3.39	-2.21	3.41	-1.73
2007	3.52	3.52	-0.12	3.48	-1.14
2008	3.33	3.44	3.36	3.40	2.10
2009	3.68	3.68	0.17	3.71	0.82
2010	4.12	4.14	0.39	4.17	1.21
2011	4.64	4.65	0.28	4.67	0.65
2012	4.86	4.81	-1.05	4.85	-0.21
2013	5.01	4.94	-1.49	4.93	-1.60
2014	4.98	4.87	-2.27	4.92	-1.20
2015	4.35	4.49	3.22	4.42	1.61
平均相对误差/%			1.39		1.19
2016	4.39	4.32	-1.68	4.37	-0.46
2017	4.52	4.51	-0.20	4.54	0.44
2018	4.34	4.40	1.30	4.39	1.15
2019	4.43	4.42	-0.16	4.46	0.68
平均相对误差/%			0.83		0.68

4 结论

为实现工业用水量预测和提高预测精度,本文提出将混合策略改进麻雀搜索算法用于优化支持向量机参数,建立了基于 ISSA-SVM 工业用水量预测模型。实例应用结果表明,ISSA-SVM 模型寻优速度快、训练和预测精度高,可用于工业用水量预测。

Industrial Water Consumption Prediction Improved by SSA-SVM Based on Mixed Strategy

CAO Jing-chun¹, LU Min²

- (1. Yunnan Institute of Water Resources and Hydropower Survey, Design and Research, Kunming 650000, China;
2. College of Water Resources and Hydraulic Engineering, Yunnan Agricultural University, Kunming 650000, China)

Abstract: In order to realize the prediction of industrial water consumption and improve the prediction accuracy, a mixed strategy was introduced to optimize the sparrow search algorithm (SSA) for improving the global search ability. The improved sparrow search algorithm (ISSA) was used to optimize the parameters of support vector machine (SVM). A support vector machine model (ISSA-SVM) based on hybrid strategy and ISSA was established, and the prediction of industrial water consumption in Ningxia was taken as an example. The results show that the ISSA-SVM model has the characteristics of fast optimization speed and high precision, and it has good applicability and feasibility in industrial water consumption prediction.

Key words: support vector machine; industrial water consumption forecast; sparrow search algorithm; adaptive inertia weight; criss-cross strategy

参考文献:

- [1] 王浩, 王建华, 胡鹏. 水资源保护的新内涵:“量-质-域-流-生”协同保护和修复[J]. 水资源保护, 2021,37(2):1-9.
- [2] 周申蓓, 刘亚灵, 郑士鹏, 等. 工业用水量测算方法及应用[J]. 水利经济, 2017,35(1):40-44,47,77.
- [3] 武风波, 赵盼, 吕茜彤. 基于 GA-BP 神经网络的城市用水量预测[J]. 现代电子技术, 2020,43(8):147-150.
- [4] 赵晓东. 基于 GASA-SVR 的园区短期需水预测方法[D]. 邯郸:河北工程大学, 2019.
- [5] 段玉先, 刘昌云. 基于 Sobol 序列和纵横交叉策略的麻雀搜索算法[J]. 计算机应用, 2022,42(1):36-43.
- [6] 薛建凯. 一种新型的群智能优化技术的研究与应用[D]. 上海:东华大学, 2020.
- [7] 回立川, 陈雪莲, 孟嗣博. 多策略混合的改进麻雀搜索算法[J]. 计算机工程与应用, 2022,58(16):71-83.
- [8] 曾琮, 黄强, 陈德, 等. 基于改进纵横交叉算法的电网最优潮流计算[J]. 中国电力, 2021,54(9):9-16.
- [9] 胡军, 邱俊博. 多策略灰狼算法优化 SVM 的尾矿坝地下水位预测[J]. 矿冶工程, 2021,41(3):24-27.
- [10] BIN Z, HONGWEI J, CHEN X, et al. Learning performance of weighted distributed learning with support vector machines[J]. IEEE transactions on cybernetics, 2021.
- [11] 闫文晶. 工业用水量预测研究[D]. 郑州:华北水利水电大学, 2022.