

基于 BWO-SVM 的地下水水位预测

王颖慧¹ 李佳丽² 刘弋琿¹

1.河南省地质局地质灾害防治中心, 河南 郑州 450000

2.安阳市建设工程质量安全和消防技术中心, 河南 安阳 455000

[摘要]地下水是全球水循环的重要组成部分,为掌握地下水水位动态变化,文章应用白鲸优化算法(Beluga Whale Optimization, BWO)对支持向量机(Support Vector Machines, SVM)进行参数优化,并利用支持向量机(SVM)的非线性拟合能力对未来5个月的地下水水位变化进行预测。以美国东北部某地下水水位观测井及其周边气象站数据为例,建立标准SVM模型以及BWO-SVM模型对地下水水位数据进行拟合并预测,对比验证了利用BWO-SVM模型方法对地下水水位动态变化进行预测的可行性。

[关键词]地下水水位变化; 预测; 白鲸优化算法; 支持向量机

DOI: 10.33142/ec.v8i6.17239

中图分类号: TH117

文献标识码: A

Groundwater Level Prediction Based on BWO-SVM

WANG Yinghui¹, LI Jiali², LIU Yiyue¹

1. The Prevention and Control Center for the Geological Disaster of He'nan Geological Bureau, Zhengzhou, Hen'an, 450000, China

2. Anyang Construction Engineering Quality and Safety and Fire Protection Technology Center, Anyang, He'nan, 455000, China

Abstract: Groundwater is an important component of the global water cycle. In order to understand the dynamic changes of groundwater level, this article applies the Beluga Whale Optimization (BWO) algorithm to optimize the parameters of Support Vector Machines (SVM), and uses the nonlinear fitting ability of SVM to predict the groundwater level changes in the next five months. Taking a groundwater level observation well and its surrounding meteorological station data in the northeastern United States as an example, a standard SVM model and a BWO-SVM model were established to fit and predict groundwater level data. The feasibility of using the BWO-SVM model method to predict dynamic changes in groundwater level was compared and verified.

Keywords: changes in groundwater level; forecast; white whale optimization algorithm; support vector machine

1 概述

作为最大的淡水资源,地下水资源在水循环中起到了至关重要的作用^[1]。近年来,随着人口增长和气候环境的变化使得地下水需求量也随之增加^[2],因此对地下水水资源进行预测有利于进一步对其进行合理调配^[3]。当前研究方法主要包括基于物理的数值模拟和基于数据驱动的机器学习^[4]。物理模型虽概念明确,却依赖大量难以精确获取的水文地质参数。相比之下,如人工神经网络^[5]、小波分析^[6]、主成分-时间序列模型^[7]、支持向量机(Support Vector Machines, SVM)^[8-10]等机器学习方法则无需物理参数,已广泛应用于地下水水位预测。

为此,本文利用白鲸优化算法^[11](Beluga Whale Optimization, BWO),利用BWO具有求解精度高、收敛速度快的特点^[12-13]优化SVM参数,构建基于BWO-SVM的地下水水位预测模型。

2 基于白鲸优化算法优化支持向量机

2.1 支持向量机(SVM)

假设 $\{(x_1, y_1), \dots, (x_i, y_i)\}$, $(x_i, y_i) \in (R^n \times R)$ 为训练样本。对于地下水监测数据来说,样本中 y_i 为地下水水位, x_i 为

影响地下水水位的因素,支持向量机需要输入训练样本来得到最优的函数关系 $y = f(x)$,再通过这个函数关系来拟合影响地下水水位的因素和地下水水位之间的关系,公式为:

$$f(x) = \omega \cdot \varphi(x) + b \quad (1)$$

式中: $\varphi(x)$ 为非线性映射, ω 为权向量, b 为偏差。

根据结构风险最小化原理,上述函数回归问题等价于最小化代价泛函:

$$\min_{\omega, b, \xi, \xi^*} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (2)$$

$$\text{s.t.} \quad \begin{aligned} \omega^T \varphi(x_i) + b - y_i &\leq \varepsilon + \xi_i \\ y_i - \omega^T \varphi(x_i) - b &\leq \varepsilon + \xi_i^* \end{aligned} \quad i = 1, \dots, N \quad (3)$$

式中: C 为惩罚参数, $C > 0$; ξ 、 ξ^* 为松弛变量, $\xi, \xi^* \geq 0$ 。

用拉格朗日求解,得到非线性模型:

$$f(x) = \sum_{i=1}^N (-\alpha_i + \alpha_i^*) k(x_i, x_j) + b \quad (4)$$

式中: α_i 、 α_i^* 为Lagrange算子; $k(x_i, x_j)$ 为核函数,这里采用RBF核函数:

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (5)$$

2.2 白鲸算法 (BWO)

白鲸算法 (BWO) 是一种元启发式优化算法, 通过模拟白鲸群体的游泳行为、捕食行为、鲸落等行为进行目标优化。该算法具有设计简单、需要调节的参数少且收敛速度快等优点, 通过多个基准函数测试且已经在许多领域得到了应用。本文采用白鲸算法对支持向量机的惩罚系数 C 和核函数参数 σ 进行寻优并建立模型:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \\ \dots & \dots \\ x_{n,1} & x_{n,2} \end{bmatrix} \quad (6)$$

式中: x 为白鲸种群; n 为白鲸种群内个体的个数; 每个白鲸个体都代表一组 C 和 σ , 变量维度为 2, $x_{n,1}$ 为第 n 个白鲸所储存的 C 值, $x_{n,2}$ 为第 n 个白鲸所储存的 σ 值。

白鲸种群将进行游泳行为和捕食行为, 其中游泳行为所代表的阶段为勘探阶段, 捕食行为所代表的阶段为开发阶段。白鲸种群选择何种行为取决于平衡因子 B_f , 当 $B_f > 0.5$ 时白鲸种群进入勘探阶段, 当 $B_f < 0.5$ 时白鲸种群进入开发阶段。平衡因子 B_f 计算公式为:

$$B_f = B_0 \left(1 - \frac{T}{T_{\max}}\right) \quad (7)$$

式中: B_f 为平衡因子; B_0 为 (0, 1) 范围内的随机数; T 为迭代次数; $T_{\max}$ 为最大迭代次数。

当白鲸种群进行游泳行为即进入勘探阶段时, 两条白鲸会进行成对游泳, 位置变换公式如下:

$$\begin{cases} X_{i,j}^{T+1} = X_{i,P_j}^T + (X_{r,P_1}^T - X_{r,P_2}^T)(1 + r_1)\sin(2\pi r_2), j = \text{偶数} \\ X_{i,j}^{T+1} = X_{i,P_j}^T + (X_{r,P_1}^T - X_{r,P_2}^T)(1 + r_1)\cos(2\pi r_2), j = \text{奇数} \end{cases} \quad (8)$$

式中: $X_{i,j}^{T+1}$ 为第 i 个白鲸在 j 维上的新位置 (维度为 2); X_{i,P_j}^T 为 (0, 1) 范围内的随机数; T 为迭代次数; $T_{\max}$ 为最大迭代次数。

当白鲸种群进行捕食行为即进入开发阶段时, 白鲸会通过共享位置来进行合作捕食。在此阶段引入 Levy 游走策略来增强算法收敛性, 位置变换公式如下:

$$X_i^{T+1} = r_3 X_{\text{best}}^T - r_4 X_i^T + C_1 \cdot L_F(X_r^T - X_i^T) \quad (9)$$

式中: X_i^T 为第 i 个白鲸的当前位置; X_i^{T+1} 为第 i 个白鲸的更新位置; 在 X_r^T 为随机白鲸的当前位置; X_r^{T+1} 为该随机白鲸的更新位置; T 为迭代次数; r_3 、 r_4 为 (0, 1) 范围内的随机数; C_1 为用于衡量 Levy 游走强度的随机跳跃度, 公式为:

$$C_1 = 2 \cdot r_4 \cdot (1 - T/T_{\max}) \quad (10)$$

L_F 为 Levy 游走函数, 公式为:

$$L_F = 0.05 \cdot \frac{u \cdot \sigma}{|v|^{1/\beta}} \quad (11)$$

$$\sigma = \left(\frac{\Gamma(1+\beta) \cdot \sin(\pi\beta/2)}{\Gamma((1+\beta)/2) \cdot \beta \cdot 2^{(\beta-1)/2}} \right)^{1/\beta} \quad (12)$$

式中: β 为值为 1.5 的常数; u 、 v 为服从正态分布的随机数。

白鲸死去后沉入海底的现象为鲸落现象, 判断是否有鲸落现象用如下公式:

$$W_f = 0.1 - 0.05T/T_{\max} \quad (13)$$

式中: w_f 为白鲸鲸落概率, 随迭代次数增加而降低, 初始值为 0.1。

当白鲸种群有个体鲸落时, 种群感受到威胁会进行位置变换, 位置变换公式如下:

$$X_i^{T+1} = r_5 X_i^T - r_6 X_r^T + r_7 X_{\text{step}} \quad (14)$$

$$X_{\text{step}} = (u_b - l_b) \exp(-C_2 T/T_{\max}) \quad (15)$$

式中: X_{step} 为鲸落步长; r_5 、 r_6 、 r_7 为 (0, 1) 范围内的随机数; u_b 为变量上限; l_b 为变量下限; C_2 为阶跃因子, 与白鲸鲸落概率和种群内个体个数相关, 公式为:

$$C_2 = 2W_f \cdot n \quad (16)$$

2.3 BWO-SVM 组合模型建模步骤

本文的组合模型通过输入数据进行训练, 通过白鲸优化算法进行惩罚系数 C 和核函数参数 σ 这两个支持向量机参数寻优, 得到最优参数后再利用支持向量机进行计算, 并最终输出其预测结果。图 1 为组合模型建模流程图。

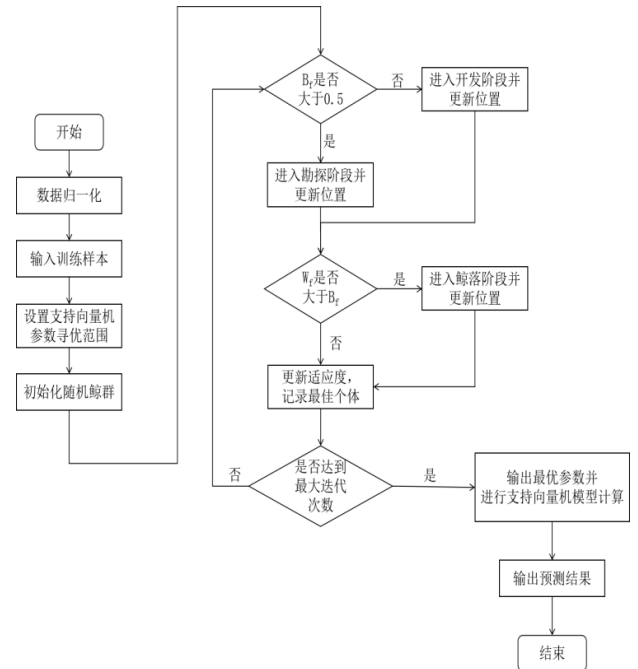


图 1 BWO-SVM 组合模型建模流程图

3 工程实例

本文选择美国东北部并深为 204.83m 的 394037075191501 号地下水水位观测井 (<https://waterdata.usgs.gov/nwis/gw>) 2020 年 1 月至 2024 年 4 月的月平均地下水水位观测数据, 选择该水位监测井周边的 PA_Avondale_2_N 气象站点 (<https://www.ncdc.noaa.gov/cm/qcdatasets.html>) 的月平均气温、月降水总量、月平均太阳辐射、月平均地表温度建立 BWO-SVM 模型。去除掉噪点数据后共得到 2020 年 1 月至 2024 年 4 月的 48 组数据, 选取前 43 组用于拟合, 后 5 组用于预测。

表 1 观测站位置

站点	经度	纬度
394037075191501	75.32° E	39.68° N
PA_Avondale_2_N	75.79° E	39.86° N

输入模型的参变量共 6 个: $\theta - \theta_0$ 、 $\ln \theta - \ln \theta_0$ 、 T 、 P 、 SR 、 GT 。其中: θ_0 为第一个监测日到始测日的天数除以 100, 本次 θ_0 取值为 0; θ 为第一个监测日到始测日的天数除以 100; T 为月平均气温; P 为月降水总量; SR 为月平均太阳辐射; GT 为月平均地表温度。

对训练样本进行标准化, 公式如下:

$$X' = 0.8 \frac{X - X_{\min}}{X_{\max} - X_{\min}} + 0.1 \quad (17)$$

式中: $X_{\min}$ 为每组样本数据的最小值; $X_{\max}$ 为每组样本数据最大值。

白鲸个体数量取 60, 最大迭代次数取 150, 计算得到 SVM 的最佳参数值 $C = 81.5589$ 、 $\sigma^2 = 7.6618$ 。利用复相关系数评价拟合精度, BWO-SVM 组合模型复相关系数

为 0.9916, 拟合精度高于 SVM 模型的 0.9900, 可以看出 BWO-SVM 组合模型的拟合效果较好。

表 2 模型复相关系数对比表

计算方法	SVM	BWO-SVM
复相关系数	0.9900	0.9916

利用训练好的模型对 5 组数据进行预测, 预测结果如表 3、图 3 所示。

表 3 地下水水位实测值和预测值对比表

日期	地下水水位/m	SVM		BWO-SVM	
		预测值/m	相对误差/%	预测值/m	相对误差/%
2023-11	94.35	91.26	-3.28	91.73	-2.78
2023-12	92.93	91.09	-1.98	91.74	-1.29
2024-1	91.85	91.79	-0.07	92.41	0.61
2024-2	91.14	89.82	-1.45	90.79	-0.37
2024-3	90.78	90.77	-0.01	92.12	1.47
平均值			-1.37		-0.49

本文引入均方误差 (MSE)、平均绝对误差 (MAE)、平均绝对百分比误差 (MAPE) 三个误差指标, 公式如下:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (18)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (19)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (20)$$

式中: y 为实测值; $\hat{y}$ 为模型预测值; n 为样本数。

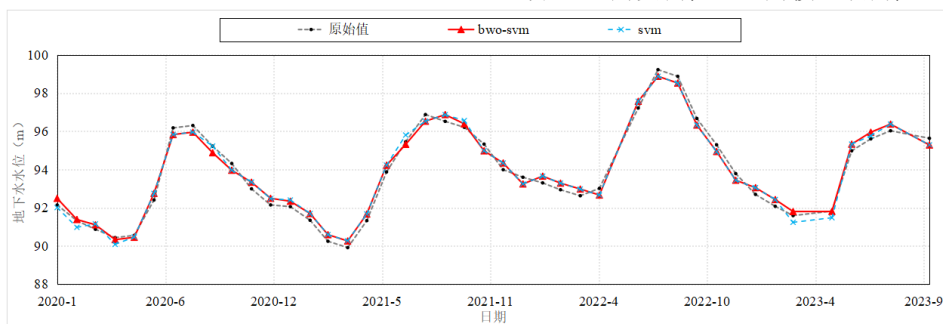


图 2 模型拟合图

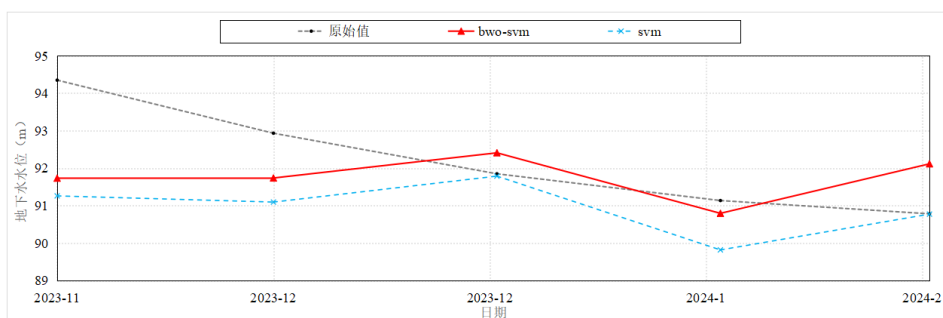


图 3 地下水水位实测值与预测值

表 4 模型误差指标

计算方法	SVM		BWO-SVM	
	拟合	预测	拟合	预测
MSE	0.1107	2.9359	0.1024	2.0997
MAE	0.3240	1.2639	0.3071	1.2101
MAPE	0.0035	0.0136	0.0033	0.0130

由表 4 可以看出 BWO-SVM 模型的拟合、预测的各个误差指标均小于传统 SVM 模型,说明本模型拟合、预测结果较为精确,证明了 BWO-SVM 模型应用于地下水水位预测的可行性。

4 结语

本文建立了支持向量机与白鲸优化算法相结合的地下水水位预测模型,在使用支持向量机的基础上引入白鲸优化算法对支持向量机的参数进行寻优,简化了寻参的过程。在实例中,GWO-SVM 模型与 SVM 模型相比拟合和预测精度较高,证明了其在地下水水位预测的可行性。

[参考文献]

- [1]SHIKLOMANOV I A,RODDA J C.World Water Resources at the Beginning of the 21st Century[M].Cambridge:Cambridge University Press,2003.
- [2]WADA Y,VANBEEK L,VIVIROLI D,et al.Global monthly water stress:2. Water demand and severity of water stress[J].Water Resources Research,2011(47):68-69.
- [3]陈飞,徐翔宇,羊艳,等.中国地下水资源演变趋势及影响因素分析[J].水科学进展,2020,31(6):811-819.
- [4]王光生,杨建青,于钊,等.地下水动态预测的探讨[J].水文,2013,33(3):25-28.
- [5]JEONG J,PARK E,CHEN H,et al.Estimation of groundwater level based on the robust training of recurrent neural networks using corrupted data[J].Journal of Hydrology,2020(582):124512.

[6]WU C,ZHANG X,WANG W,et al .Groundwater level modeling framework by combining the wavelet transform with a long short-term memory data-driven model[J].Science of the Total Environment,2021(783):146948.

[7]张展羽,梁振华,冯宝平,等.基于主成分-时间序列模型的地下水位预测[J].水科学进展,2017,28(3):415-420.

[8]YOON H,JUN S-C,HYUN Y,et al .A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer[J].Journal of Hydrology,2011(396):128-138.

[9]陈海洋,滕彦国,王金生.基于 GA 参数优选的 ε -SVR 地下水位预测方法[J].水资源保护,2011,27(4):15-18.

[10]刘娣,余钟波,吕海深,等.基于 SVM-EnKF 双向数据同化的地下水水位变化预测[J].水文,2023,43(6):58-65.

[11]Changting Z,Gang L,Zeng M.Beluga whale optimization: A novel nature-inspired metaheuristic algorithm[J].Knowledge-Based Systems,2022(251):55-56.

[12]王亚辉,张虎晨,王学兵,等.基于混沌反向学习和水波算法改进的白鲸优化算法 [J]. 计算机应用研究,2024,41(3):729-735.

[13]陈心怡,张孟健,王德光.基于 Fuch 映射的改进白鲸优化算法及应用 [J]. 计算机工程与科学,2024,46(8):1482-1492.

作者简介:王颖慧(1996.1—),毕业院校:河海大学,所学专业:水工结构工程,当前就职单位名称:河南省地质局地质灾害防治中心,职称级别:助理工程师;李佳丽(1996.6—),毕业院校:湖南大学,所学专业:给排水科学与工程,当前就职单位名称:安阳市建设工程质量安全和消防技术中心,职称级别:助理工程师;刘弋琿(1995.5—),毕业院校:中国地质大学(武汉),所学专业:地质工程,当前就职单位名称:河南省地质局地质灾害防治中心,就职单位职务:助理工程师。