

陈艳辉, 魏霖静. 基于 WOA-BP 神经网络的兰州地区降水量预测[J]. 智能计算机与应用, 2026, 16(1): 97-102. DOI: 10.20169/j. issn. 2095-2163. 24032501

基于 WOA-BP 神经网络的兰州地区降水量预测

陈艳辉, 魏霖静

(甘肃农业大学 信息科学技术学院, 兰州 730070)

摘要: 降水量不仅仅对农产品的种植至关重要, 与人们的日常生活也息息相关。本文基于 1951~2022 年兰州地区的降水量数据进行研究, 使用鲸鱼优化算法对 BP 神经网络模型进行改进, 对兰州地区降水量进行预测, 计算模型选用了评价指标 MAE、MSE, 并与 BP 神经网络模型评价指标进行对比。结果表明, WOA-BP 神经网络模型较未优化的 BP 神经网络模型的预测结果更准确, 更适用于兰州地区降水量的预测。

关键词: 降水量预测; BP 神经网络; 鲸鱼优化算法

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2026)01-0097-06

Prediction of precipitation in Lanzhou region based on WOA-BP neural network

CHEN Yanhui, WEI Linjing

(College of Information Science and Technology, Gansu Agricultural University, Lanzhou 730070, China)

Abstract: Precipitation is not only important for agricultural production, but also is closely related to people's daily lives. Based on the precipitation data of Lanzhou from 1951 to 2022, this paper uses the Whale Optimization Algorithm to improve the BP neural network model, predicts the precipitation of Lanzhou, calculates the model evaluation indexes MAE and MSE, and compares them with the evaluation indexes of BP neural network model. The research show that the results of the WOA-BP neural network model is more accurate than the unoptimized BP neural network model, and is more suitable for the prediction of precipitation in Lanzhou.

Key words: precipitation forecast; BP neural network; Whale Optimization Algorithm

0 引言

众所周知, 降雨量多变性的显著增强可能会引起旱涝灾害问题, 国内也不可避免地面临同样的挑战。在自然条件中, 大面积持续高强度降雨是形成洪涝灾害的根本原因。洪涝的发生不仅对国家经济产生影响, 而且还可能危及人们的生命安全。如果能够对降雨量进行科学预测, 就可以使人们提前采取应对措施, 从而在很大程度上减少各项受灾损失。

1 研究背景

降雨量指的是一定时段内降落到地面、水平面上单位面积的雨水深度, 以 mm 计算, 可以直观地表

示降雨的多少。降水量不仅仅包括雨水, 也包括其他水形态的降落, 例如雪。降水的多少不仅仅给当地居民带来不便, 对当地农产品种植也有很大的影响。降水量是区域水资源量计算的基础依据, 也是一种重要的灾害数据, 因此寻找一种准确有效的方法来预测降水量是十分必要的。

降水属于自然界的一种天气状态, 但是自然界的各种气候情况具有随机性, 所以对于降水量的预测也是比较困难的。目前, 神经网络模型得到了快速发展, 并在处理非线性的问题上展现出了多方面的优越性, 现在已有诸多学者使用神经网络模型对降水量进行预测, 并且获得不错的结果。

针对降水量预测的研究, Esther 等学者^[1]通过

基金项目: 2021 年度兰州市人才创新创业项目(2021-RC-47); 2022 年度科技部国家外专项目(G2022042005L); 2023 年甘肃省高等学校产业支撑项目(2023CYZC-54); 2023 年甘肃省重点研发计划项目(23YFWA0013)。

作者简介: 陈艳辉(1999—), 女, 硕士研究生, CCF 会员, 主要研究方向: 智能计算。

通信作者: 魏霖静(1977—), 女, 博士(后), 教授, CCF 会员(57876D), 主要研究方向: 智能计算, 农业信息化, 生物信息学。Email: 916277964@qq.com。

收稿日期: 2024-03-25

集成 AdaBoost Regressor (ABR) 和狮群优化 (LSO) 算法,提出了一种混合的长期(提前1个月)降水预报模型(ABR-LSO),对该地区时间序列降水数据进行了预测,并采用平均绝对误差(MAE)和均方根误差(RMSE)等统计指标来评价预测模型的效果。

针对降水量预测的研究,甘亚星^[2]采用线性回归、R/S分析、小波分析、MK突变、滑动t检验等方法,分别对石羊河流域内及周围10个气象站点1969~2018年共50年的雨季(6~9月)中不同等级降雨及总降雨量进行年内、年际趋势变化、突变、周期情况的特征分析。然后建立BP网络模型预测雨季总降雨量,接着利用遗传算法优化BP网络模型,再次对降雨量进行预测。凌铭等学者^[3]提出基于支持向量机(SVM)、完全集合经验模态分解(CEEMDAN)和双向长短期记忆神经网络(BiLSTM)的降水预测耦合模型,对鄱阳湖流域景德镇站和赣县站2个典型站点的日降水量数据进行预测,并与各种传统模型组合的预测结果进行了对比,实验表明提出的模型结果与实测结果基本一致且精度最高。

根据对国内外文献的研究分析可知,目前在降水量的预测方面,对兰州地区进行预测的还不多见。基于此,本文将对兰州地区的降水量进行预测研究,填补了兰州地区降水量研究文献较少的不足,对兰州地区的农作物种植、居民日常生活及出行等方面均有重要价值。

2 数据预处理

本文研究所使用的数据为1951年1月2日到2022年12月31日的日降水量历史气象数据,共有26297条数据。本研究所使用的数据都来自国家气象局,来源可靠。本文选取3种气象因子:气温、相对湿度、日照时数,主要研究对象为降水量。

国家气象局的数据集格式为TXT。为了方便数据整理,对数据集格式进行转换并保存为EXCEL格式,同时将提取所需要的兰州市(代码为52889)的逐日降水量及气象数据。

在数据预处理的过程中,本实验所使用的数据集也存在数值缺失、异常等问题。分析可知,这些情况可能出于各种原因,包括人为错误或技术故障。

有以下3种对于缺失值的处理方法:

(1)剔除。如果存在缺失或者异常情况,删除当天数据。例如,1951年1月1日数据,当天平均气温缺失,而且没有前一天的数据进行参考,所以对这一天的数据进行剔除处理。

(2)就近补齐。使用最接近的数据进行填充。这种方法可能会导致关键信息的丢失,特别是在数据出现突然变化的情况下。因此,对于本数据集来说不可用。

(3)均值补齐。数据的时间序列性质要求采用更连续的方法来填补缺失值,比如用类似时间段的平均值来替代缺失数据。在进行数据预处理时,发现有几天的数据缺少日照时数,就可根据最近一周内和缺少数据当天气象情况(气温、相对湿度、降水量等)相近的日照时数取出来进行平均处理,来填补缺失值。

对于异常值,降水量数据中出现了一些需要着重处理的数据。例如,30XXX、31XXX、32XXX、32700。查阅降水量数据文档,文档内有对这些异常值的解释:30XXX为雪量(仅包括雨夹雪,雪暴),31XXX为雨和雪的总量,32XXX为纯雾露霜,32700表示降水“微量”。根据异常值的说明,32700取0mm,其他的异常值除以1000之后取余数。

对实验数据进行数值处理之后,直接用于神经网络之前,还需要对所有的实验数据进行归一化。数据归一化起着关键的作用,主要是防止小数值数据被吞噬,做到统一量纲。本实验将所有的数据进行线性归一化,转换成[0,1]之间的数值。

使用的线性归一化,即最小-最大规范化,是对原始数据的线性变换,将数据值映射到[0,1]之间。具体公式如下:

$$x_1 = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

其中, x_1 表示标准化后的数值; x 表示原始数值。

线性归一化保留了原始数据中存在的关系,是消除量纲和数据取值范围影响的最简单的方法。

3 研究方法及创新点

对于降水量的预测,本文选取了BP神经网络模型,这也是一种经典的网络模型。除了使用基础的BP神经网络对降水量进行预测,还使用较新型的鲸鱼优化算法(WOA)对BP神经网络进行优化,后续则使用优化后的模型对兰州日降水量进行预测,将2个模型的预测结果进行模型指标评价。根据指标评价结果,证明优化后的模型更适用于兰州地区降水量的预测。

(1)BP神经网络模型。BP(Back Propagation)神经网络是1986年由Rumelhart和McClelland等学者提出的概念,是一种按照误差逆向传播算法训练

的多层前馈神经网络,也是目前应用最广泛的神经网络模型之一^[4-7]。

BP 神经网络是一种多层网络,只需要借助样本数据,就可以对网络进行权值训练,实现输入到输出的高度非线性映射,BP 神经网络的计算过程分为正向计算过程与反向计算过程。

BP 神经网络模型在结构上具有输入层、隐藏层、输出层三个部分,每层都设有若干个神经元。输入层的神经元个数是由输入的数据决定的,输出层由期望输出的数据决定,隐藏层的神经元个数不定,一般会根据具体情况来决定。网络拓扑结构如图 1 所示。

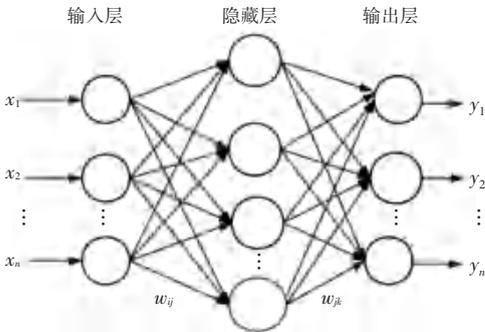


图 1 BP 神经网络拓扑图

Fig. 1 BP neural network topology diagram

在 BP 神经网络模型中,输入层的节点个数与输出层的节点个数都是确定的,只有隐藏层的节点个数不确定。隐藏层的节点个数对于整个神经网络模型的性能好坏具有很大的影响。根据已有的研究成果,对于隐藏层节点个数的选取,会用到如下公式:

$$h = \sqrt{m + n} + a \quad (2)$$

其中, m 表示输入层的节点数量; n 表示输出层的节点数量; a 表示 1 ~ 10 之间的调节常数; h 表示隐藏层节点数量。

从输入层通过隐藏层转向输出层,逐层处理,每一层神经元的状态仅仅影响下一层神经元的状态,这就是正向传播过程。如果经过正向传播过程,在输出层不能得到预期的输出,就会进入反向传播过程。此时将误差信号沿原来的路线返回,通过修改各神经元的权值,使误差信号最小,从而达到预期目标。

(2) 鲸鱼优化算法。鲸鱼优化算法 (Whale Optimization Algorithm, WOA) 是在 2016 年提出的一种新型群体智能算法^[8-10]。该计算方法是通过对自然界中座头鲸群体狩猎行为进行模拟。在自然界中,鲸鱼群体狩猎时有 3 种策略:搜索觅食、包围猎物 and 捕

捉猎物,所以,鲸鱼优化算法也同样包含这 3 个步骤。

鲸鱼优化算法在搜索空间产生若干鲸鱼个体,鲸鱼群体会根据目前群体中的最优个体更新每只鲸鱼的位置,通过循环迭代来获得最优值。鲸鱼优化算法的 3 个种群更新机制相互独立,因此其寻优阶段的全局探索和局部开发过程得以分别运行及控制。此外,鲸鱼优化算法不需要人为地设置各种控制参数值,提高了算法的使用效率并降低了应用难度。与其它群体智能优化算法相比,WOA 算法结构新颖,控制参数少,在许多数值优化问题的求解中表现出较好的寻优性能,并且优于蚁群算法和粒子群算法等智能优化算法^[11-12]。

假设鲸鱼种群的规模为 N , d 表示空间维度,第 i 只鲸鱼在空间中的位置为 $X_i = (x_{i1}, \dots, x_{id})$ 。鲸鱼优化算法在各个策略中的计算步骤如下。

步骤 1 包围猎物。猎物在群体中的最优位置,即对应全局最优解,其他鲸鱼则会向这个位置逐步靠拢包围,利用下列公式更新位置:

$$D = |CX^*(t) - X(t)| \quad (3)$$

$$X(t+1) = X^*(t) - AD \quad (4)$$

其中, t 表示目前迭代次数; $X^*(t)$ 表示目前猎物所在的位置、目前个体位置; D 表示鲸鱼与猎物之间的距离; A 、 C 表示位置系数,分别用如下公式来进行定义:

$$A = 2ar_1 - \alpha \quad (5)$$

$$C = 2r_2 \quad (6)$$

其中, r_1, r_2 表示 $[0, 1]$ 中的随机数; α 表示收敛因子,定义公式如下:

$$\alpha = 2 - \frac{2t}{T} \quad (7)$$

其中, T 表示最大迭代次数。

步骤 2 发泡网攻击。在 WOA 算法中,发泡网攻击步骤中,有螺旋包围和收缩包围两种捕食方法。

① 收缩包围。这种捕食行为与上述包围猎物行为的数学模型几乎完全相同,唯一的区别就是 A 的取值范围。收缩包围指的就是当前位置的鲸鱼个体向当前最佳位置的鲸鱼个体位置靠近^[13-15]。

② 螺旋式位置更新。鲸鱼通过螺旋游动的方式逼近猎物进行狩猎,更新与猎物之间的距离,当前鲸鱼个体向最佳鲸鱼个体靠近。

WOA 算法中模拟鲸鱼螺旋攻击猎物建立方程为:

$$X(t+1) = \begin{cases} X^*(t) - AD, & p < 0.5 \\ D^t e^{bl} \cos(2\pi l) + X^*(t), & p \geq 0.5 \end{cases} \quad (8)$$

其中, $D^i = |X_p(t) - X(t)|$ 表示鲸鱼与猎物之间的距离; b 表示限定螺旋线的长度; l 表示 $[-1, 1]$ 区间的随机数。在不断优化过程中, 为了模拟鲸鱼在沿着螺旋线包围同时收缩包围的路径向猎物靠近, 设置更新概率 p 为 0.5。

步骤3 搜索猎物。为了提高算法的全局搜索能力, 避免陷入局部最优, $A \geq 1$ 时, 算法将会随机搜寻猎物的位置, 该阶段的表达式如下:

$$D = |\bar{X}(t) - X(t)| \quad (9)$$

$$X^*(t+1) = \bar{X}(t) - AD \quad (10)$$

其中, $\bar{X}(t)$ 表示鲸鱼群体中随机选择的鲸鱼位置。

WOA 算法先是对一组解进行随机初始化, 然后在每次迭代中, 根据随机选择的搜索代理或者到目前为止获得的最优解来更新其位置。WOA 算法流程如图 2 所示。

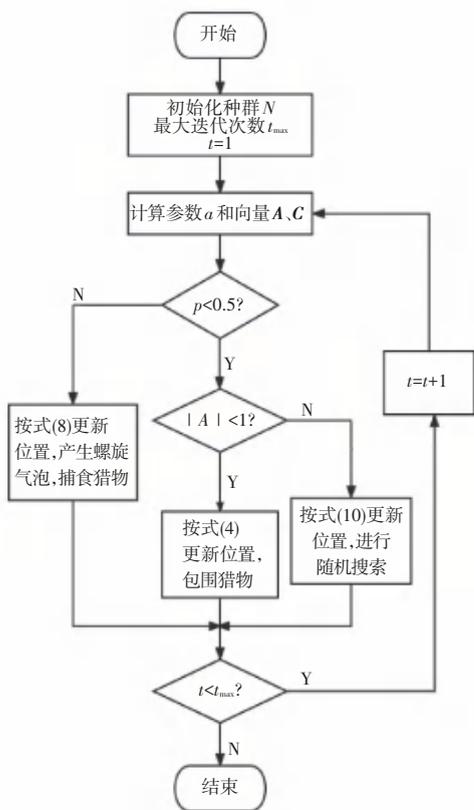


图 2 WOA 算法流程图

Fig. 2 Flowchart of WOA algorithm

(3) 本文的创新点为:

① 本文使用了较新型的鲸鱼优化算法 WOA 对 BP 神经网络模型进行优化。

② 本文研究了兰州地区的降水量, 改善相关文献较少的情况, 对兰州地区的农产品种植、居民生活

具有重要指导意义。

4 WOA-BP 神经网络预测

BP 神经网络是一种前向反馈的神经网络模型, 通过不断调整网络中的权重和偏置, 以最小化预测误差。因其具有较强的非线性建模能力和适应性, 可以有效地处理多变量和非线性关系。然而, BP 神经网络在训练过程中容易陷入局部最优解, 并且收敛速度较慢。为了解决这些问题, 本文利用鲸鱼优化算法 (WOA) 对 BP 神经网络进行优化。

使用鲸鱼优化算法 (WOA) 优化 BP 神经网络进行降水量预测的步骤如下:

(1) 数据准备。收集实验数据, 并对实验数据进行预处理。对预处理后的实验数据划分为训练集和测试集, 本实验中每种数据有 26 297 个, 其中随机选取 2 万个作为训练集, 剩下的作为测试集。

(2) 初始化网络。初始化 BP 神经网络的结构和参数, 输入层节点数为 3 个 (平均气温、平均相对湿度、日照时数), 输出层节点数为 1 个 (降水量)。隐藏层节点数根据经验公式计算范围来确定。接着对每种情况都进行 1 000 次迭代, 得到各种情况下的最优训练集均方误差, 选取均方误差最小情况下的隐藏层神经元个数作为 BP 神经网络模型的隐藏层神经元个数。本次实验中选取的隐藏层神经元个数为 8。隐藏层激活函数使用 tansig , 输出层激活函数使用 purelin , 学习速率为 0.01, 训练目标最小误差为 0.000 01, 最小性能梯度为 0.000 001, 最高失败次数为 32, 其余参数权重和偏置随机为初始值。

(3) WOA 优化。本实验使用鲸鱼算法对 BP 神经网络进行权重和偏置优化。鲸鱼算法通过迭代来搜索解空间中的最优解, 从而找到最佳的权重和偏置配置。在每一次迭代中, 鲸鱼算法根据当前的解空间位置和适应度值, 调整权重和偏置的值。相关参数设置为: 种群规模 $N = 32$, 最大迭代次数 $T_{\max} = 128$, 通过不断训练、验证和测试实现预测效果最优。

(4) 训练网络。将 WOA 优化后的最优权值和偏置赋值给 BP 神经网络, 对神经网络进行训练。训练过程中, 网络根据输入数据进行前向传播和误差计算, 然后通过反向传播算法调整权重和偏置的值。这个过程会不断迭代, 直到达到实验预设的目标之一 (迭代次数 1 000, 最小误差 0.000 01, 最小性能梯度 0.000 001 或最高失败次数 32)。

(5) 数据预测。使用训练好的网络对测试集数据进行预测。优化后的 BP 神经网络根据输入数据

进行前向传播,输出预测结果。将训练结果进行反归一化,并将其与真实值进行平均绝对误差 (Mean Absolute Error, MAE)、均方误差 (Mean Squared Error, MSE) 计算。对未优化过的 BP 神经网络同样进行 MAE、MSE 评价指标计算作为对照组,证明优化后的 BP 模型更适用于兰州地区降水量预测。

使用鲸鱼优化算法 (WOA) 对权值偏置优化过程中得到的 WOA 进化曲线如图 3 所示。

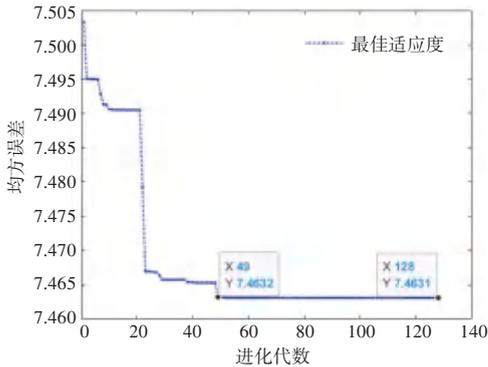


图 3 WOA 进化曲线图

Fig. 3 Evolution curve graph of WOA

在第 49 次迭代时,使用的权值偏置应用于 BP 神经网络后的均方误差为 7.463 2,此后再进行迭代,均方误差并未变小,证明此时已获得最大迭代次数内的最优权值阈值。

建立了 WOA-BP 神经网络预测模型,对于 WOA-BP 的训练结果,具体见图 4。

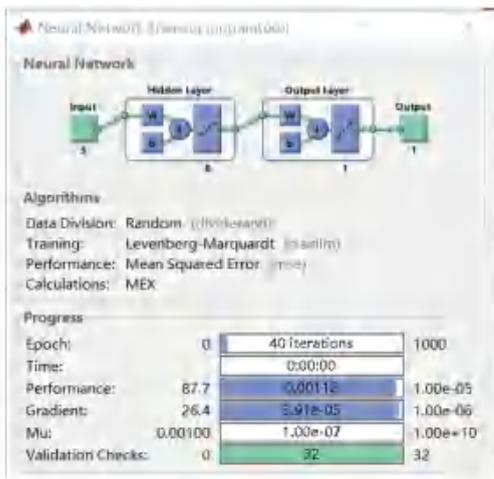


图 4 WOA-BP 模型训练结果图

Fig. 4 Training results of WOA-BP model

WOA-BP 模型迭代网络性能如图 5 所示。由图 5 可知,使用 WOA 优化后的最优权值偏置得到的 BP 神经网络模型实际共迭代了 40 次,后面 32 次迭代没有获得更好的网络性能,所以停止了迭代。在第 8 次迭代时,训练集和测试集的均方误差

(MSE) 已达最优值,此后的迭代并没有更新。

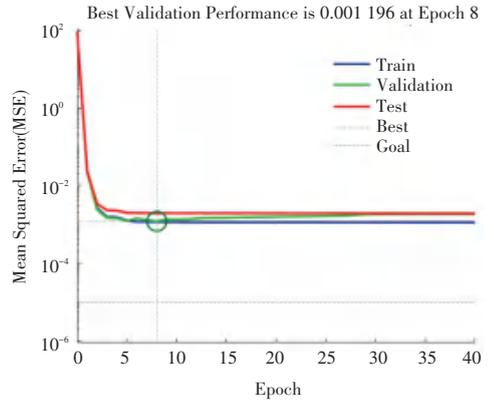


图 5 WOA-BP 模型迭代网络性能图

Fig. 5 Iterative network performance diagram of WOA-BP model

本文中的模型属于回归预测模型。对于回归预测模型常见的评价指标主要有平均绝对误差 (MAE)、均方误差 (MSE)、均方根误差 (RMSE)。

因为均方根误差 RMSE 真实值位于分母位置,所以不能用于真实值数据为 0 的情况。本模型用于兰州地区日降水量预测,因为存在日降水量为 0 (不降水) 的情况,该指标并不适用于本文研究。

为进一步证明 WOA-BP 神经网络模型精度较 BP 神经网络模型效果要更好,表 1 列出了 2 种模型误差值和预测误差指标的对比效果。

表 1 2 种神经网络模型评价指标对比表

Table 1 Comparison of evaluation indexes of two neural network models

模型	MAE	MSE
BP	0.97	7.94
WOA-BP	0.97	7.88

通过对表 1 中指标数值进行对比,可以发现 WOA-BP 神经网络模型的预测结果比 BP 神经网络要好很多。因为对于未优化的 BP 神经网络,训练次数达 1 000 次之多,所以对于模型的评价已经很好了。将 BP 神经网络进行 WOA 优化后,取得最好的权值偏置结果,再进行 WOA-BP 神经网络训练,只需要 8 次就可以达到最优。将 2 个模型的 MAE, MSE 指标进行对比后可知,虽然 MAE 的结果都是 0.97,但是 WOA-BP 模型的结果较 BP 模型结果小了 0.08。均方误差 (MSE) 对于离群值更敏感,证明 WOA-BP 网络较 BP 网络在离群值方面预测结果更好。总体而言,WOA-BP 模型较 BP 模型预测结果更好,更适用于兰州地区降水量的预测。

5 结束语

降水量不仅仅对农产品的生产有着显著影响,

也与人们的日常生活密切相关。本文采用深度学习方法,建立了基于BP神经网络的降水量预测模型,并使用鲸鱼优化算法对BP神经网络模型的权值偏置进行改进寻优,以期获得更好的预测效果。实验结果表明,WOA-BP模型具有较好的预测效果,更适用于兰州地区降水量的预测。

参考文献

- [1] ESTHER S P, KUMUDHA R, YAFIT C, et al. Evaluation of a novel hybrid lion swarm optimization-AdaBoost Regressor model for forecasting monthly precipitation[J]. *Sustainable Computing: Informatics and Systems*, 2023, 39:100884.
- [2] 甘亚星. 石羊河流域雨季降雨特征分析及降雨量的预测[D]. 武汉:华中科技大学, 2020.
- [3] 凌铭, 肖丽英, 赵嘉, 等. 基于SVM-CEEMDAN-BiLSTM模型的日降水量预测[J]. *人民珠江*, 2023, 44(9):61-68.
- [4] 王海燕, 鲁岳, 陈丽娜. 基于BP神经网络的酒泉市霜冻天气预测研究[J]. *河南科技*, 2021, 40(5):144-146.
- [5] WANG Yuping, LIU Haiyan, WEI Fei, et al. Cooperative coevolution with formula-based variable grouping for large-scale global optimization[J]. *Evolutionary Computation*, 2018, 26(4): 569-596.
- [6] 褚鼎立, 陈红, 王旭光. 基于自适应权重和模拟退火的鲸鱼优化算法[J]. *电子学报*, 2019, 47(5):992-999.
- [7] 冯豪博, 胡桥, 赵振轶. 基于精英族系遗传算法的AUV集群路径规划[J]. *系统工程与电子技术*, 2022, 44(7):2251-2262.
- [8] 李晴雯. 军民融合背景下应急物资预配置优化研究[D]. 秦皇岛:燕山大学, 2023.
- [9] LI Y, MA T, CHEN P Y, et al. Autonomous underwater vehicle optimal path planning method for seabed terrain matching navigation[J]. *Ocean Engineering*, 2017, 133:107-115.
- [10] WU Jiehong, SONG Chengxin, MA Jian, et al. Reinforcement learning and particle swarm optimization supporting real-time rescue assignments for multiple autonomous underwater vehicles[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(7): 6807-6820.
- [11] 李广强, 董文超, 朱大庆等. 基于改进鲸鱼优化算法的AUV三维路径规划[J]. *系统工程与电子技术*, 2023, 45(7):2170-2182.
- [12] HAN Zhenhua, LIU Shugui, YU Fei, et al. A 3D measuring path planning strategy for intelligent CMMs based on an improved ant colony algorithm[J]. *The International Journal of Advanced Manufacturing Technology*, 2017, 93(1): 1487-1497.
- [13] 姜辰凯, 李智, 盘书宝, 等. 基于改进Dijkstra算法的AGVs无碰撞路径规划[J]. *计算机科学*, 2020, 47(8): 272-277.
- [14] BUI D T, ABDULLAHI M M, GHAREH S, et al. Fine-tuning of neural computing using whale optimization algorithm for predicting compressive strength of concrete[J]. *Engineering with Computers*, 2021, 37(1): 701-712.
- [15] 薛瑾. 锂电池健康状态估计及剩余使用寿命预测研究[D]. 长沙:湖南大学, 2022.