

DOI:10.13476/j.cnki.nsbtdqk.2020.0119

刘洁,祝榕婕,姜德迅,等.基于遗传-神经网络的实时水质预测模型[J].南水北调与水利科技(中英文),2020,18(6):93-100.
LIU J,ZHU R J,JIANG D X,et al.Real-time water quality prediction model based on IGA-BPNN method[J].South-to-North
Water Transfers and Water Science & Technology,2020,18(6):93-100.(in Chinese)

基于遗传-神经网络的实时水质预测模型

刘洁^{1,2},祝榕婕¹,姜德迅³,王大蔚⁴,许崇品³,南军¹,王鹏¹

(1.哈尔滨工业大学 环境学院,哈尔滨 150090;2.东北农业大学 水利与土木工程学院,哈尔滨 150030;
3.哈尔滨学院 信息工程学院,哈尔滨 150086;4.黑龙江省农业科学院 农村能源与环保研究所,哈尔滨 150086)

摘要:基于高频水质在线监测数据,结合遗传算法和神经网络模型,建立基于遗传-神经网络(improved genetic algorithm-back propagation neural network,IGA-BPNN)的河流水质预测模型,实现对河流水质的实时预测预警。将该方法应用于美国波托马克河流中,对其水质参数浊度(TURB)和电导率(SC)进行实时预测,并对预测结果进行性能分析,以验证基于IGA-BPNN的河流水质预测模型的准确性与可靠性。与BPNN模型的水质预测结果进行对比分析,结果表明:IGA-BPNN模型对水质参数TURB和SC有更准确的预测效果。同时,IGA-BPNN模型对正常平稳条件下的水质参数TURB和SC预测结果的区间覆盖率PICP分别为99.81%和100%,预测结果具有一定的可靠性。IGA-BPNN水质预测模型可以有效地识别长时间的水质异常或瞬时显著的水质变化情况,可实现对河流水质的风险预警,最终可为河流突发水污染的应急处置措施的制定提供科学依据。

关键词:地表河流;水质预测;改进遗传算法;神经网络模型;预测区间

中图分类号:X323 文献标志码:A 开放科学(资源服务)标志码(OSID):



Real-time water quality prediction model based on IGA-BPNN method

LIU Jie^{1,2}, ZHU Rongjie¹, JIANG Dexun³, WANG Dawei⁴, XU Chongpin³, NAN Jun¹, WANG Peng¹

(1. School of Environment, Harbin Institute of Technology, Harbin 150090, China; 2. School of Conservancy & Civil Engineering, Northeast Agricultural University, Harbin 150030, China; 3. School of Information Engineering, Harbin University, Harbin 150086, China; 4. Rural Energy & Environmental Protection Institute, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China)

Abstract: A surface water quality prediction model is developed to analyze the inherent water quality variation tendencies and provides real-time early warnings according to the high-frequency water quality historical observation data. The developed water quality prediction model is integrated by an improved genetic algorithm (IGA) and a Back Propagation Neural Network (BPNN). To verify the accuracy and reliability of the river water quality prediction model based on IGA-BPNN, the method is applied to the Potomac River in the United States to predict its water quality parameters, turbidity (TURB) and conductivity (SC), and to analyze the performance of the prediction results. The prediction results demonstrate that the developed IGA-BPNN model can provide a more accurate prediction result than the BPNN model. Since the PICP values of TURB and EC prediction can reach 99.81% and 100% under normal stable conditions, IGA-BPNN model also can provide a reliability prediction result. Meanwhile, the developed IGA-BPNN models can reflect the long period and isolated sharp peaks of the water quality variations, and effectively provide real-time early warning for emergency response.

收稿日期:2019-12-25 修回日期:2020-04-16 网络出版时间:2020-05-06

网络出版地址: <http://kns.cnki.net/kcms/detail/13.1430.TV.20200506.1116.002.html>

基金项目:国家自然科学基金(51779066);国家重点研发计划课题(2018YFC0408001);中国博士后科学基金面上项目(2018M631935);哈尔滨学院大学生科技创新项目(HXS20171506)

作者简介:刘洁(1988—),女,黑龙江大庆人,博士(后),助理研究员,主要从事水环境数值模拟研究。E-mail:liujie_198805@163.com.cn

通信作者:王鹏(1957—),男,黑龙江哈尔滨人,教授,博士,主要从事水环境数值模拟研究。E-mail:pwang73@hit.edu.cn

Key words: surface water; water quality prediction; improved genetic algorithm; BPNN; prediction interval

近年来,随着我国工业化、城镇化以及农业化进程的加快,自然灾害、生产事故等引发的突发水污染事件频繁发生,对人类健康、生态环境和经济发展造成了严重影响^[1-2]。伴随着河流水体污染、水质恶化等问题,地表水质预测预警技术的研究受到了一定的关注^[3-4]。大多数河流水质预测理论主要是通过以污染源为主体的机理性水质模型和以受纳水体为主体的非机理性水质模型,对河流水质变化规律和污染物迁移扩散情况进行探索和预测。然而,以污染源为主体的河流水质预测预警理论研究受到环境系统和数学机理复杂多变的影响预测精度受限。

随着人工智能技术的发展,以受纳水体为主体的非机理性水质预测方法得到广泛的应用,主要利用自回归模型^[5]、灰色模型^[6]、支持向量机^[7]和人工神经网络^[8]等方法建立数学模型描述并预测水质变化特征。与其他数学模型相比,人工神经网络模型具有较强的非线性映射能力和柔性的网络拓扑结构,对于非线性复杂问题具有很好的适用性^[9]。因此,可以较好地反映水质参数内在的变化规律和波动特征。Kim 等^[10]将聚类算法与人工神经网络相结合,以减少训练数据集不平衡对神经网络拓扑结构训练所产生的影响;Shi 等^[11]将小波分析方法与人工神经网络相结合,通过对水质时间序列进行小波降噪,提高人工神经网络的预测精度;翟伟等^[12]耦合灰色预测理论和人工神经网络方法,以解决水质时间序列变化响应滞后问题,并提高水质预测精度。由于神经网络模型的收敛速度慢,在模型的训练过程中容易产生过拟合的现象^[13],大多数研究采用多种方法耦合的方式以提高水质预测精度。遗传算法作为一种模仿自然环境中的生物遗传和进化现象的优化求解方法,可以通过优化人工神经网络的初始权重参数,指导和确定拓扑结构的搜索方向,有效避免人工神经网络陷入局部极小值化和收敛速度慢等问题^[14-15]。

本文将遗传算法和 BP 神经网络算法相结合,提出遗传-神经网络模型(improved genetic algorithm-back propagation neural network, IGA-BPNN)对河流水质进行实时预测预警。以美国波托马克河为研究对象,针对水质参数浊度(TURB)和电导率(SC)进行实时预测,与 BP 神经网络模型(back propagation neural network, BPNN)的水质

预测结果进行对比分析,验证 IGA-BPNN 模型对河流水质预测的准确性与可靠性,从而实现河流水质实时预测和预警,为河流突发水污染的应急处置措施的制定提供科学依据。

1 研究方法

将遗传算法与 BP 神经网络算法相结合,构建基于 IGA-BPNN 的河流水质预测模型,实现对河流水质的预测预警。该模型主要包括遗传算法进行适应度函数极值寻优和 BP 神经网络进行模型训练拟合两步。

1.1 反向传播神经网络模型 BPNN

1.1.1 神经网络拓扑结构

神经网络模型包含 3 种基本的网络结构:输入层、隐含层和输出层^[16]。本文采用具有多个输入神经元、一个输出神经元的反向传播单隐含层神经网络拓扑结构。由于地表河流中水体的光合/呼吸作用的昼夜动态变化与一个日循环中温度的昼夜变化有关,进而影响地表河流水质参数在基态条件下浓度的昼夜变化^[17-18],因此,输入层神经元个数由一个日周期模式(24 h)和原始水质时间序列监测数据的采样频率决定。同时,根据预测误差最小原则分别测试 2~16 个隐含层神经元个数,选择最优隐含层神经元个数。本文采用一个神经元的输出层进行水质预测,通过提前 1 步预测(one-step-ahead prediction)模式实现河流水质的实时预测。

1.1.2 样本数据

利用训练和验证数据对 IGA-BPNN 模型的 BPNN 部分进行初始权重参数的调整和网络拓扑结构的训练,以保证河流水质预测模型得到最佳的泛化程度。同时,在模型训练过程中,采用交叉验证技术划分训练数据集和验证数据集以防止训练模型出现过拟合的现象^[19]。最后,需要测试数据验证 IGA-BPNN 模型的 BPNN 部分训练后的河流水质预测性能和模型的泛化能力。因此,本文将水质监测数据分为两个子集(训练和验证集 70%、测试集 30%)以训练、验证和测试 IGA-BPNN 模型的 BPNN 部分的神经网络拓扑结构。

1.1.3 模型训练

IGA-BPNN 模型中 BPNN 部分拓扑结构的训

练本质上是一个优化问题,最小化神经网络水质预测误差^[20]。为了满足 IGA-BPNN 模型中 BPNN 部分预测精度的要求,本文利用 Levenberg Marquardt (LM)反向传播算法的近似非线性函数,以观测值与预测值之间的均方误差(mean squared error, MSE)作为神经网络预测性能函数,训练 BPNN 模型的网络结构以实现河流水质参数变化规律的预测。

1.2 改进遗传算法 IGA

在通常情况下,由于 BPNN 模型的初始权重参数集是随机生成的,在输入数据相同的情况下, BPNN 模型得到的预测结果和预测精度是不同的。同时,在 BPNN 模型训练过程中,不恰当的初始权重参数也可能会导致 BPNN 模型训练过拟合的现象。遗传算法(GA)基于自然进化和选择的原则,通过选择、交叉、突变和复制等遗传算子,根据适应度函数进行评估以搜索并优化多个决策变量之间的内在关系。本文将改进的遗传算法(IGA)引入到 BPNN 模型中,以优化 BPNN 模型的初始权重参数,防止训练模型陷入局部最优,而出现 BPNN 模型训练过拟合的现象,进而提高 IGA-BPNN 模型的河流水质预测精度。

1.2.1 适应度函数

以河流水质观测值和预测值的均方误差(mean squared error, MSE)作为适应度函数,以评价 BPNN 模型神经网络训练和学习的程度。MSE 值越小,模型训练效果越好。遗传算法会向着适应度函数值增大的方向进行。因此,遗传算法的适应度函数公式为

$$\text{fitness}=1/E \quad (1)$$

$$E=1/2m \sum_{k=1}^m \sum_{t=1}^q [C_{\text{est}}^k(t)-C_{\text{obs}}^k(t)]^2 \quad (2)$$

式中: $C_{\text{est}}^k(t)$ 为第 k 个输出神经元在 t 时刻的水质预测值; $C_{\text{obs}}^k(t)$ 为第 k 个输出神经元在 t 时刻的水质观测值; m 为 BPNN 模型的输出神经元的数量; q 为模型训练数据的数量。通过最小化水质预测值与观测值的均方误差,得到适应度最好的个体作为 BP 神经网络的初始权重值。

1.2.2 选择操作

在选择操作中,通常采用轮盘法(roulette wheel method, RWM)根据适应度函数值选择出种群中的优势个体。在轮盘法中,个体 c 的选择概率与其适应度值成正比

$$P_{\text{sel}}(c)=f(c)/\sum_{m=1}^{N_{\text{pop}}} f(m) \quad (3)$$

式中: N_{pop} 为当前种群中个体的数量。然而,在随机操作过程中,不可避免地会有一些具有高适应度函数值的优势个体被淘汰。本文根据当前种群中最优个体的适应度函数值,选择一定比例(精英率 Best)的最优个体直接复制到下一代,以防止最优个体在采用轮盘法进行选择操作的过程中丢失^[21]。

1.2.3 交叉和变异操作

为了提高遗传算法性能,需要提高种群多样性和收敛到全局最优的能力。本文根据搜索空间特点自适应调整交叉概率 P_{cro} 和变异概率 P_{mut} ,以增加种群的多样性,从而提高算法的局部搜索能力,避免算法陷入局部最优。具体公式为

$$P_{\text{cro}}(c)=\begin{cases} P_{\text{max}} \frac{(P_{\text{max}}-P_{\text{min}})(f(c)-f_{\text{avg}})}{f_{\text{max}}-f_{\text{avg}}} & f(c) \geq f_{\text{avg}} \\ P_{\text{max}} & f(c) < f_{\text{avg}} \end{cases} \quad (4)$$

$$P_{\text{mut}}(c)=\begin{cases} P'_{\text{max}} \frac{(P'_{\text{max}}-P'_{\text{min}})(f(c)-f_{\text{avg}})}{f_{\text{max}}-f_{\text{avg}}} & f(c) \geq f_{\text{avg}} \\ P'_{\text{max}} & f(c) < f_{\text{avg}} \end{cases} \quad (5)$$

式中: $f(c)$ 为待交叉和变异个体的适应度函数值; f_{max} 为每代种群中最大的适应度函数值; f_{avg} 为每代种群的平均适应度函数值; P_{max} 为最大的交叉概率值; P_{min} 为最小的交叉概率值。 P'_{max} 为最大的变异概率值; P'_{min} 为最小的变异概率值。在通常情况下,交叉概率取值设定在 0.7~0.9,变异概率取值设定在 0.001~0.05。

1.3 性能分析

本文主要从模型预测的准确性和可靠性两方面对 IGA-BPNN 水质预测模型进行性能评价。采用均方根误差(RSME)、平均绝对误差(MAE)、平均相对误差(MRE)和决定系数(r^2)对 IGA-BPNN 水质预测模型的准确性进行评价。

$$\text{RMSE}=\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i-\hat{y}_i)^2} \quad (6)$$

$$\text{MAE}=\frac{1}{n} \sum_{i=1}^n |y_i-\hat{y}_i| \quad (7)$$

$$\text{MRE}=\frac{1}{n} \sum_{i=1}^n |y_i-\hat{y}_i/y_i| \quad (8)$$

$$r^2=\frac{\sum_{i=1}^n (y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i-\bar{y})^2 \sum_{i=1}^n (\hat{y}_i-\bar{\hat{y}})^2}} \quad (9)$$

式中: y_i 为水质时间序列观测值; \bar{y} 为水质时间序列观测值的平均值; \hat{y}_i 为水质时间序列预测值; $\bar{\hat{y}}$ 为水质时间序列预测值的平均值; n 为水质时间序列数

据数量。

采用预测区间覆盖率 (PICP) 对 IGA-BPNN 水质预测模型的可靠性进行评价。采用预测区间 PI, 即下一时刻的观测值可能的取值范围, 作为接收或拒绝下一时刻预测值的阈值, 并计算 PICP 值来表征观测值落在这个范围内的预期概率^[22-23]。

$$s = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n-1}} \quad (10)$$

$$PI_i = \hat{y}_i \pm t_{n-p}^{\alpha/2} \times s \quad (11)$$

$$PICP = \left(\frac{1}{n} \sum_{i=1}^n c_i\right) \times 100 \quad (12)$$

式中: s 为水质时间序列预测值的标准偏差; PI_i 为水质时间序列预测区间; $t_{n-p}^{\alpha/2}$ 为 $(n-p)$ 个自由度下 $\alpha/2$ 百分比的学生分布列。若 $\{t_i\}$ 服从标准正态分布, $t_{n-p}^{\alpha/2} = 1.96$, 即 95% 置信区间的值为 1.96, 表示 IGA-BPNN 水质预测模型的预测结果有 95% 的概率会落入预测区间 PI 范围内。 c_i 为判断系数, 若 y_i 在预测区间 PI 内, 则 $c_i = 1$, 否则 $c_i = 0$ 。在通常情况下, PICP 值越大越好, 当大于 95% 时可认为 IGA-BPNN 水质预测模型的预测是可靠的。

1.4 计算流程

将遗传算法与 BP 神经网络算法相结合, 构建基于 IGA-BPNN 的河流水质预测模型, 实现对河流水质预测预警。输入层中的神经元的个数 q 由原始时间序列中时间序列的变化周期决定, 输出层的神经元个数为 1 个, 表示当前水质预测值, 采用滑动窗口预测模式实现河流水质的实时预测, 见图 1。



图 1 滑动窗口
Fig. 1 Sliding window

通过前 q 个历史观测值 $D^t = \{x_{t-q+1}, \dots, x_t\}$ 构建滑动窗口, 预测下一时刻的河流水质情况, 即采用提前 1 步预测模式实现河流水质的实时预测。初始化后, 遗传-神经网络算法将 q 个历史水质监测数据填充至滑动窗口中, 并预测下一时刻河流水质数据 X_{t+1} 。采用 IGA-BPNN 模型进行河流水质预测的流程见图 2, 具体计算流程如下:

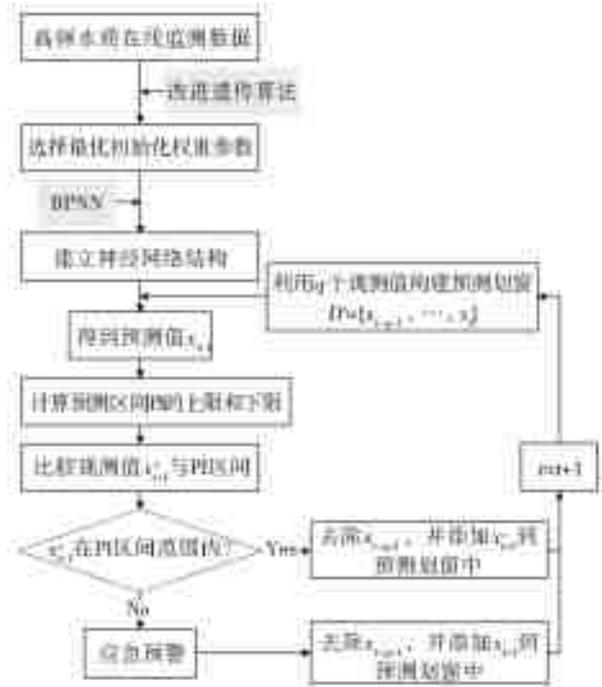


图 2 算法流程示意图
Fig. 2 Flow chart for developed algorithm

(1) 根据由 q 个数据组成的滑动窗口, 采用改进遗传算优化 BP 神经网络的初始权值参数, 并调整算法的神经网络拓扑结构。

(2) 利用 BP 神经网络算法进一步拟合神经网络拓扑结构中的权值参数, 最小化河流水质观测值与预测值之间的误差, 实现水质预测模型的训练, 保证水质预测精度。

(3) 采用提前 1 步预测模式, 以前 q 个历史观测值 $D^t = \{x_{t-q+1}, \dots, x_t\}$ 为输入数据, 得到 $t+1$ 时刻河流水质预测值 x_{t+1} 。

(4) 计算水质预测模型的预测区间 PI。

(5) 与 $t+1$ 时刻河流水质观测值 x'_{t+1} 相比较, 如果 x'_{t+1} 落在区间范围外, 则为异常数据。将滑动窗口数据集 D^t 中最后一个观测值 x_{t-q+1} 去除, 将预测值 x_{t+1} 添加到滑动窗口数据集中, 形成数据集 D^{t+1} 。

(6) 如果 x'_{t+1} 落在区间范围内, 则为正常数据。将滑动窗口数据集 D^t 中最后一个观测值 x_{t-q+1} 去除, 将观测值 x'_{t+1} 添加到滑动窗口数据集中, 形成数据集 D^{t+1} 。

(7) 重复步骤(3)至(6)。

2 结果与讨论

2.1 案例概况

波托马克河(Potomac River), 作为美国中东部最重要的河流, 源于阿巴拉契亚山脉西麓, 由北布朗奇河和南布朗奇河汇合而成, 最终注入大西洋的切

萨皮克湾,全长大约为 652 km,流域面积约为 3.8 万 km²,见图 3。美国地质调查局(USGS)在波托马克河设置了 39 个高频水质监测站,对水质参数流量、水温、pH、溶解氧、电导率和浊度等进行连续在线监测,监测频率为 5/15/60 min。USGS 公开发布主要监测站点的长期水质高频在线监测数据可为本研究提供主要的技术支持,验证 IGA-BPNN 模型对河流水质预测的准确性与可靠性。



图 3 波托马克河位置示意图

Fig. 3 Location and spatial distribution of Potomac River

选取监测站点 No. 01632900 从 2017 年 1 月 27 日 0:00 至 2017 年 3 月 21 日 0:00 之间的水质参数 TURB 和 SC 的高频在线监测数据,监测频率为 15 min,共 5 089 组。由于不同水质参数的量纲不同,本文采用 Z-score 方法对高频水质在线监测数据进行标准化处理,处理后的水质监测数据符合均值为 0 方差为 1 的正态分布,具体计算方式^[24-25]为

$$D_i = \frac{Y_i - \bar{Y}}{\sigma} \quad (13)$$

式中: \bar{Y} 为水质在线监测数据 Y_i 的平均值; σ 为水质在线监测数据 Y_i 的方差。

2.2 水质预测结果分析

建立基于 IGA-BPNN 的河流水质预测模型,主要包括两个阶段:采用改进遗传算法进行适应度函数极值寻优,将得到适应度最好的个体作为 BP 神经网络的初始权重值;采用 BP 神经网络算法对网络拓扑结构进行训练和拟合,以提高 IGA-BPNN 模型对河流水质的预测精度。通过对水质参数 TURB 和 SC 的预测结果的性能分析,验证 IGA-BPNN 的准确性和可靠性。

2.2.1 IGA-BPNN 模型水质预测结果

将标准化处理后的高频水质在线监测数据分为两个子集(训练和验证集 70%,测试集 30%),用于训练、验证和测试 IGA-BPNN 模型中 BPNN 部分的神经网络拓扑结构。以 2017 年 1 月 27 日至

2017 年 2 月 15 日的 1 920 组标准化历史观测值作为训练集,对 BPNN 模型进行初始权重参数的优化和神经网络拓扑结构的训练。以 2017 年 2 月 16 日至 2017 年 2 月 24 日的 864 组标准化历史观测值作为验证集,将得到的预测值与观测值进行对比,验证 IGA-BPNN 模型的河流水质预测性能。同时,选取 2017 年 2 月 25 日至 2017 年 3 月 21 日的 2 305 组标准化历史观测值作为测试集,利用含有水质异常的历史观测值测试 IGA-BPNN 模型对水质预测效果和实时预警性能。

采用提前 1 步预测的滑动窗口预测模式,建立 TURB 和 SC 水质参数的 IGA-BPNN 水质预测模型。首先,输入层的神经元数量,即滑动窗口大小 q ,由 1 个日循环(24 h)和原始水质时间序列监测数据的采样频率(15 min)决定。采样频率为 15 min 的高频水质在线监测数据可以很好地描述 24 h 之内的水质变化特征,因此,本文将输入层神经元个数设置为 96。将水质参数的时间序列观测值 $t-95, \dots, t-3, t-2, t-1$ 和 t 作为滑动窗口模式的输入变量,输出变量为 $t+1$ 。

在 IGA-BPNN 水质预测模型训练阶段,运行 IGA-BPNN 模型 100 次以寻找最佳的 BPNN 初始权重参数和神经网络拓扑结构。IGA-BPNN 调整神经网络拓扑结构的迭代过程,见图 4。与采用单纯的 BPNN 模型进行水质预测相比,IGA 算法对 BPNN 初始权重参数的合理选择,使得 IGA-BPNN 模型在迭代训练的初始阶段的误差比单纯的 BPNN 模型的误差小。同时,IGA-BPNN 模型在迭代次数约为 1 000 次时达到收敛,而单纯的 BPNN 模型需要迭代 1 600 次时达到收敛。结果表明,IGA-BPNN 模型比单纯的 BPNN 模型具有更好的收敛速度和泛化能力。因此,本文建立的 IGA-BPNN 模型可以在水质模型训练过程中得到较合理的神经网络拓扑结构。

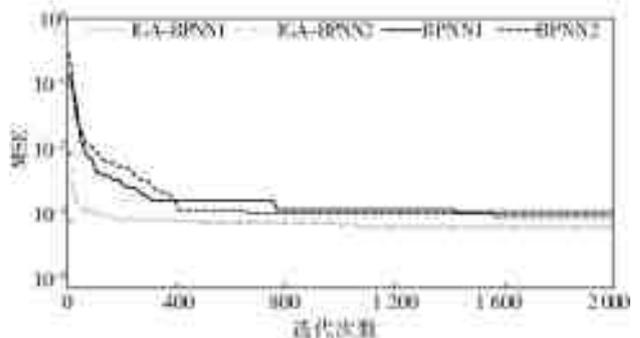


图 4 算法进化曲线(情景 1)

Fig. 4 Evolution curve based on two algorithms under scenario 1

在 IGA-BPNN 水质预测模型测试阶段,从训

练阶段的 100 个神经网络中选取最优的神经网络拓扑结构对河流水质进行预测。水质参数 TURB 和 SC 的预测结果和预测区间见图 5。由水质预测结果可知:自 2017 年 2 月 16 日至 2017 年 3 月 17 日期间,水质变化较稳定,IGA-BPNN 模型可以较好地预测水质参数 TURB 和 SC 的变化和波动情况。自 2017 年 3 月 18 日至 2017 年 3 月 21 日期间,水质参数 TURB 和 SC 产生异常变化,IGA-BPNN 模型对水质参数 TURB 和 SC 的预测结果产生较大的偏差,水质参数 TURB 和 SC 的观测值甚至超出了预测区间 PI。其中,水质参数 TURB 和 SC 的预测残差的最大绝对值可分别达到 10.06

和 2.01。水质监测数据的异常表现为浊度的增加和电导率的下降,且持续时间在数小时甚至一两天时间脱离正常情况基准线。这种水质异常情况可能是由于强降雨事件造成的非点源污染。2017 年 3 月 8 日, TURB 产生了短时或单点的水质异常变化。这种水质异常情况可能是由于在线监测设备故障导致的观测值异常。综上所述,在水质平稳的条件下,本文建立的 IGA-BPNN 水质预测模型可以较好地预测水质参数的变化特征和波动情况。在非平稳条件下,IGA-BPNN 水质预测模型可以有效地识别长时间的水质异常或瞬时显著的水质变化,并对其进行实时风险预警。

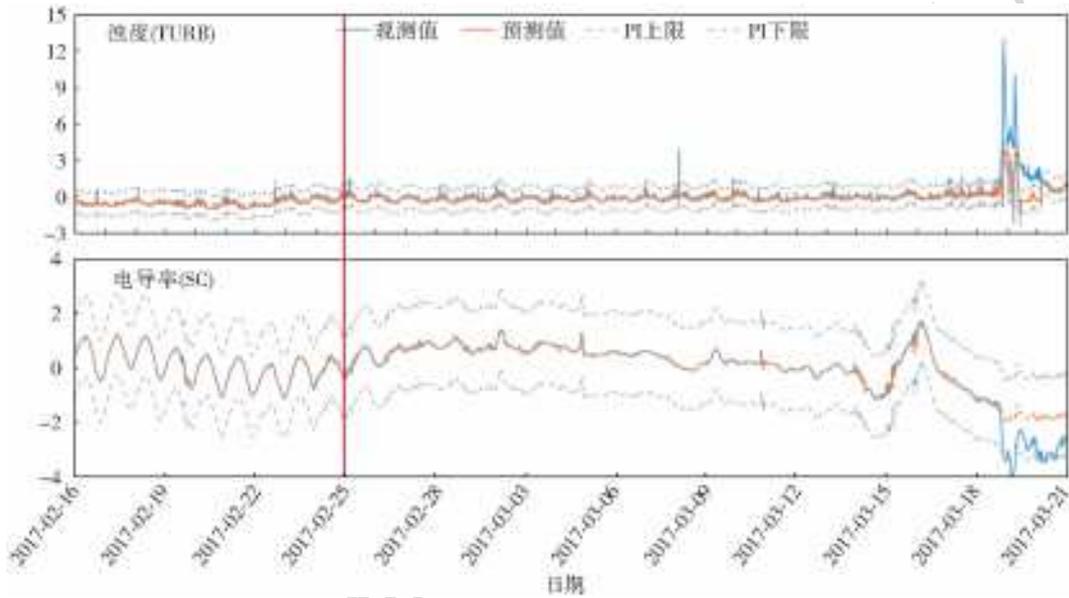


图 5 算法进化曲线(情景 2)

Fig. 5 Evolution curve based on two algorithms under scenario 2

2.2.2 IGA-BPNN 模型性能分析

采用 RSME、MAE、MRE 和 r^2 对 IGA-BPNN 水质预测模型的预测结果进行误差分析,并与单纯 BPNN 模型进行比较,见表 1。根据 2017 年 2 月 16 日至 2017 年 3 月 17 日期间的预测结果,将 IGA-BPNN 与单纯的 BPNN 模型水质参数 TURB 和 SC 的预测结果进行误差分析和比较。可以看出,对于正常平稳条件下的水质参数 TURB 和 SC 的预测,IGA-BPNN 模型的预测结果具有更好的准确性,因此,所建立的 IGA-BPNN 水质预测模型可以有效提高水质预测精度。同时,本文采用 PICP 对 IGA-BPNN 水质预测模型的可靠性进行评价。根据 2017 年 2 月 16 日至 2017 年 3 月 17 日期间的预测结果,计算得到正常平稳条件下的水质参数 TURB 和 SC 预测结果的区间覆盖率分别为 99.81% 和 100%,因此,所建立的 IGA-BPNN 水质预测模型的

预测结果具有一定的可靠性。

表 1 IGA-BPNN 和 BPNN 模型预测结果误差分析

Tab. 1 Comparisons of the performance analysis between IGA-BPNN and BPNN model

水质参数	TURB		SC	
	IGA-BPNN	BPNN	IGA-BPNN	BPNN
RMSE	0.025 6	0.026 4	0.002 0	0.003 2
MAE	0.102 0	0.104 1	0.027 7	0.032 0
MRE	0.812 9	0.875 1	0.138 9	0.170 9
r^2	0.875 7	0.847 2	0.989 3	0.971 1

所提出的水质预测模型 IGA-BPNN 针对一维水质参数进行的水质预测预警,分析识别可能发生的水污染事件。下一步工作重点将围绕多维水质数据的预测预警分析,通过研究多个水质参数之间的变化异常特征实现对水污染事件的识别和预警。IGA-BPNN 的河流水质预测模型具有较高的灵活

性和可移植性。随着我国在线监测技术的进步以及地表水质监测数据的积累,建立大数据支撑下河流水质有效预测预警模型具备了技术可行性和数据可行性。因此,可将该方法应用在我国具有一定监测数据基础的地表河流中。同时,随着我国对河流水质管理的重视和河长制的逐步推行,IGA-BPNN模型可为河流突发水污染的应急处置措施的制定提供科学依据。

3 结 论

本文采用遗传-神经网络模型 IGA-BPNN 对河流水质进行实时预测。将该方法应用于美国波托马克河流中,对其水质参数 TURB 和 SC 进行预测。与 BPNN 模型的水质预测结果进行对比分析,验证基于 IGA-BPNN 的河流水质预测模型的准确性与可靠性,从而实现河流水质预测预警,为河流突发水污染的应急处置措施的制定提供科学依据。

(1) 将遗传算法与 BP 神经网络算法相结合,构建基于 IGA-BPNN 的河流水质预测模型,实现对河流水质实时预测和预警。该模型利用遗传算法优化得到 BP 神经网络的初始权重值,利用 BP 神经网络进行网络拓扑结构的训练拟合,以提高 IGA-BPNN 模型对河流水质预测的准确性和可靠性。

(2) 以美国波托马克河为研究对象,结合 RSME、MAE、MRE、 r^2 以及 PICP,验证基于 IGA-BPNN 的河流水质预测模型的准确性和有效性。结果表明:与 BPNN 模型相比,IGA-BPNN 模型对水质参数 TURB 和 SC 有更准确的预测效果。同时,IGA-BPNN 模型对正常平稳条件下的水质参数 TURB 和 SC 预测结果的区间覆盖率 PICP 分别为 99.81% 和 100%,预测结果具有一定的可靠性。

(3) 根据 IGA-BPNN 模型构建的水质预测区间 PI 可对异常波动条件下的水质参数 TURB 和 SC 进行异常识别和预警。IGA-BPNN 水质预测模型可以有效识别长时间脱离正常基线的水质异常或瞬时显著的水质变化情况,实现河流水质实时风险预警,最终为河流突发水污染的应急处置措施的制定提供科学依据。

参考文献:

[1] SHI S G, CAO J, FENG L, et al. Construction of a technique plan repository and evaluation system based on AHP group decision-making for emergency treatment and disposal in chemical pollution accidents[J].

Journal of Hazardous Materials, 2014(276): 200-206. DOI:10.1016/j.jhazmat.2014.05.034.

- [2] BERTONE E, SAHIN O, RICHARDS R, et al. Extreme events, water quality and health: A participatory Bayesian risk assessment tool for managers of reservoirs[J]. Journal of Cleaner Production, 2016(135): 657-667. DOI:10.1016/j.jclepro.2016.06.158
- [3] JEONG J, PARK E, HAN W S, et al. A subagging regression method for estimating the qualitative and quantitative state of groundwater[J]. Hydrogeology Journal, 2017(25): 1491-1500. DOI:10.1007/s10040-017-1561-9
- [4] 史斌. 水污染动态预警监测模型构建与应急处置工程风险分析[D]. 哈尔滨: 哈尔滨工业大学, 2018.
- [5] 秦文虎, 付亚涛. 基于向量自回归模型的水质异常检测研究[J]. 安全与环境学报, 2018, 18(4): 1560-1563. DOI:10.13637/j.issn.1009-6094.2018.04.056.
- [6] 李娜, 王腊春, 谢刚, 等. 山东省辖淮河流域河流水质趋势的灰色预测[J]. 环境科学与技术, 2012, 35(2): 195-199.
- [7] 孙伯寅, 董国庆, 张荣. 支持向量机在水源水化学耗氧量预测中的应用[J]. 环境与健康杂志, 2016, 33(6): 544-547. DOI:10.16241/j.cnki.1001-5914.2016.06.019.
- [8] 李春华, 胡文, 叶春, 等. 基于 BP 神经网络预测地表水净化装置总氮的去除效果[J]. 环境工程技术学报, 2018, 8(6): 651-655.
- [9] 惠高云, 谈立峰, 褚苏春, 等. 生活饮用水在线监测预警模型的研究进展[J]. 环境与健康杂志, 2018, 35(9): 841-845. DOI:10.16241/j.cnki.1001-5914.2018.09.025.
- [10] KIM S E, SEO I W. Artificial Neural Network ensemble modeling with conjunctive data clustering for water quality prediction in rivers[J]. Journal of Hydro-environment Research, 2015(9): 325-339. DOI:10.1016/j.jher.2014.09.006.
- [11] SHI B, WANG P, JIANG J P, et al. Applying high-frequency surrogate measurements and a wavelet-ANN model to provide early warnings of rapid surface water quality anomalies[J]. Science of the Total Environment, 2018(610-611): 1390-1399. DOI:10.1016/j.scitotenv.2017.08.232
- [12] 翟伟, 毛静, 孟雅丹, 等. 结合灰色理论的人工神经网络方法在水质预测中的应用[J]. 南水北调与水利科技, 2020, 18(1): 138-143.
- [13] 巫莉莉, 黄志宏, 何斌斌, 等. 智能算法在水体氨氮含量预测中的应用研究综述[J]. 中国农机化学报, 2019, 40(6): 191-196. DOI:10.13733/j.jcam.issn.2095-5553.2019.06.36.

- [14] 张旭东,高茂庭.基于 IGA-BP 网络的水质预测方法[J].环境工程学报,2016,10(3):1566-1571. DOI: CNKI;SUN;HJJZ.0.2016-03-090.
- [15] 曹栋华,陈佳袁,刘益志,等.优化神经网络模型在水质预测中的运用[J].三峡大学学报(自然科学版),2016,38(4):12-16. DOI:10.13393/j.cnki.issn.1672-948X.2016.04.003.
- [16] MULIA I E, TAY H, ROOPSEKHAR K, et al. Hybrid ANN-GA model for predicting turbidity and chlorophyll-a concentration[J]. Journal of Hydro-Environment Research, 2013, 7: 279-299. DOI: 10.1016/j.jher.2013.04.003.
- [17] ZHAI X Y, XIA J, ZHANG Y Y. Water quality variation in the highly disturbed Huai River basin, China from 1994 to 2005 by multi-statistical analyses[J]. Science of the Total Environment, 2014, 496: 594-606. DOI: 10.1016/j.scitotenv.2014.06.101.
- [18] HALLIDAY S J, SKEFFINGTON R A, WADE A J, et al. High-frequency water quality monitoring in an urban catchment: hydrochemical dynamics, primary production and implications for the water framework directive[J]. Hydrological processes, 2015, 29: 3388-3407. DOI: 10.1002/hyp.10453.
- [19] WAGH V M, PANASKAR D B, MULEY A A. Estimation of nitrate concentration in groundwater of Kadava River basin-Nashik district, Maharashtra, India by using artificial neural network model[J]. Modeling Earth Systems and Environment, 2017, 3(1): 36. DOI: 10.1007/s40808-017-0290-3.
- [20] MAY R J, DANDY G C, MAIER HR, et al. Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems[J]. Environ Modelling and Software, 2008(23): 1289-1299. DOI: 10.1016/j.envsoft.2008.03.008.
- [21] LIU J, GUO L, JIANG J P, et al. Emergency material allocation and scheduling for the application to chemical contingency spills under multiple scenarios[J]. Environmental Science and Pollution Research, 2017, 24(1): 956-968. DOI: 10.1007/s11356-016-7811-x.
- [22] HILL D J, MINSKER B S. Anomaly detection in streaming environmental sensor data: A data-driven modeling approach[J]. Environmental Modelling & Software, 2010(25): 1014-1022. DOI: 10.1016/j.envsoft.2009.08.010.
- [23] 王祎.基于计算智能方法的河流水质管理数字模拟研究与应用[D].哈尔滨:哈尔滨工业大学,2013.
- [24] 史斌,王鹏,姜继平.基于高频在线水质监测数据异常预警的突发污染预警[J].中国环境科学,2017,37(11):4394-4400.
- [25] 朱炜玉,史斌,姜继平,等.基于水质时间序列异常检测的动态预警方法[J].环境科学与技术,2018,41(12):131-137. DOI: 10.19672/j.cnki.1003-6504.2018.12.019.

(上接第 92 页)

- [27] 钱凤魁,王卫雯,王秋兵.基于耦合协调度模型量化耕地自然质量与立地条件协同关系[J].农业工程学报,2018,34(18):284-291. DOI: 10.11975/j.issn.1002-6819.2018.18.035.
- [28] 邹强,周建中,周超,等.基于可变模糊集理论的洪水灾害风险分析[J].农业工程学报,2012,28(5):126-132. DOI: 10.3969/j.issn.1002-6819.2012.05.021.
- [29] 张慧利,蔡洁,夏显力.水土流失治理效益与生态农业发展的耦合协调性分析[J].农业工程学报,2018,34(8):162-169. DOI: 10.11975/j.issn.1002-6819.2018.08.021.