

文章编号: 2095-2163(2020)08-0261-05

中图分类号: X52; TP18

文献标志码: A

基于智能算法的水系化学污染物预测模型研究

陈 岑, 田晓丹, 武文星

(华北科技学院, 河北 廊坊 101601)

摘要: 水系污染不仅会严重影响居民的日常用水,还会使得许多水域附近农田受到不同程度的有机物或者重金属污染。对水系化学污染物进行监测是对水系污染治理和水环境管理的基础。本研究在人工神经网络模型的基础上,构建了水系化学污染的BP神经网络智能预测模型。在此基础上,利用淮河水系某监测站点的数据,验证了该预测水系化学污染的BP神经网络模型的有效性,得到相应的预测结果。研究表明,对于衡量水系化学污染的pH值、DO、NH₃-N以及COD等四个指标,BP神经网络预测模型的预测精度较高,在一定程度上反映出水质变化的规律。验证了采用BP神经网络智能预测模型,预测水系化学污染物的有效性和可行性。

关键词: BP神经网络; 水系; 污染物; 预测

Study on the prediction model of water chemical pollutants based on Intelligent Algorithm

CHEN Cen, TIAN Xiaodan, WU Wenxing

(North China Institute of Science and Technology, Langfang 101601, Hebei, China)

【Abstract】 Water pollution will not only seriously affect the daily water use of residents, but also make farmland near many water areas polluted by different degrees of organic matter or heavy metals. Monitoring chemical pollutants in water system is the basis of water pollution control and water environment management. In this study, on the basis of artificial neural network model, the BP neural network intelligent prediction model of water system chemical pollution is established. On this basis of using the data of a monitoring station in Huaihe River system, the validity of the BP neural network model is verified, and the corresponding prediction results are obtained. The results show that the BP neural network prediction model has high prediction accuracy and can reflect the rules of water quality change to a certain extent. This study verifies the validity and feasibility of BP neural network intelligent prediction model for water chemical pollutants. It is hoped that this study can provide some reference for the study of intelligent prediction model of water chemical pollutants based on BP neural network.

【Key words】 BP neural network; water system; pollutant prediction

0 引言

随着社会经济的快速发展,众多水流域的污水排放量不断增加,许多河流湖泊受到严重污染,水质急剧恶化^[1-3]。水系污染中最主要的是化学成分污染,常见的水系化学污染物包括DO、NH₃-N以及COD等化合物。水污染不仅会影响居民的正常生活用水,还会使得许多水域附近农田受到不同程度的有机物或者重金属污染,最终致使有害物质残留在农作物中,影响居民的身体健康和生命安全^[4-6]。进行水系污染治理和水环境管理的前提是水质评价,因而需要对水质进行监测和预测^[7-8]。BP神经网络作为一种应用广泛的计算机网络系统,可以应用于水系污染的监测和评价。因此,本研究在BP神经网络的基础上,构建水系化学污染物智能预测

模型。

1 基于BP神经网络的水系化学污染物预测模型构建

1.1 BP神经网络

人工神经网络是一种能够进行并行计算的神经网络系统,包含许多神经元,并且神经元之间按照一定的拓扑结构相互连接,形成众多的处理单元^[9-12]。正是由于人工神经网络具有复杂的结构,使得这种神经网络系统形成了非线性的复杂动态特征。人工神经网络的基本单元是神经元,主要元素包括一个加法器、一组连接以及一个激励函数。其中一组连接主要是指将一个神经元的输出值与另一个神经元相连接,并将其作为下一个神经元的输入信号。一个加法器主要是用于求解通过神经元输入层信号与神经元对

基金项目: 国家重点研发计划(2018YFC0808306);河北省重点研发计划(19270318D);河北省物联网监控工程技术研究中心(3142018055);

青海省物联网重点实验室(2017-ZJ-Y21);廊坊市科技计划自筹项目(2020011019)。

作者简介: 陈 岑(1996-),男,硕士研究生,主要研究方向:神经网络。

收稿日期: 2019-05-25

应突触的加权和。激励函数在一定范围内可以压制输入信号,其形式多样,包括阶梯函数、分段线性函数以及非线性函数。阶梯函数可以表示如下:

$$f(x) = \begin{cases} 1, & x \geq 0; \\ 0, & x < 0. \end{cases} \quad (1)$$

在阶梯函数中,若函数自变量 $x \geq 0$, 则其输出值为 1;函数的自变量 $x < 0$, 则其输出函数值为 0。分段线性函数可以表示如下:

$$f(x) = \begin{cases} 1, & x \geq 1; \\ x, & -1 < x < 1; \\ -1, & x \leq -1. \end{cases} \quad (2)$$

在线性分段函数中,如果函数的自变量 $x \geq 1$, 则其输出值为 1。如果函数的自变量 $-1 < x < 1$, 则其输出值为自变量本身;如果函数的自变量 $x \leq -1$, 则其函数输出值为-1。BP 神经网络是人工智能的一个分支,是当前应用最广的神经网络之一。作为一种多层反馈型网络,在进行正向信号传播的同时,还能反向传播所得误差,进而调整网络权系数。BP 神经网络通常具有三层网络结构,主要包括输入层、中间层和输出层,其中中间层也称为隐含层,由多个节点组成。输入层和输出层的节点一般可以是一个,也可以是多个,模型结构如图 1 所示。

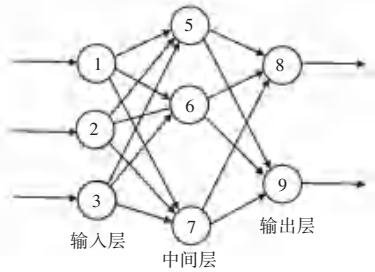


图 1 BP 神经网络模型结构

Fig. 1 Structure of BP neural network model

BP 神经网络学习过程主要包括信息的正向传播以及误差的反向传播。假定输入层 i 节点的输出为 I_i , i 节点与中间层 j 节点之间的权值为 w_{ji} , 节点 j 的阈值为 θ_j , 节点 j 与输出层节点 k 之间的权值为 V_{kj} , 节点 k 的阈值为 r_k 。在信息的正向传播过程中,都要经过中间隐含层与输出层的非线性激活函数,而后将结果输出,相关的非线性激活函数如下式(3)所示。

$$f(x) = \frac{2}{1 + e^{-2x}} - 1. \quad (3)$$

BP 神经网络信息的正向传播流程如图 2 所示。由图 2 可见:首先赋予连接权值和阈值为 $(-1, 1)$ 之

间的随机小数,初始化网络参数;其次将第一个样本从输入层输入;随之计算中间各节点的输出,计算公式为 $H_i = f(\sum_{i=1}^M w_{ji} I_i + \theta_j)$;最后计算输出层各节点的输出,计算公式为 $O_k = f(\sum_{j=1}^H V_{kj} H_j + r_k)$ 。

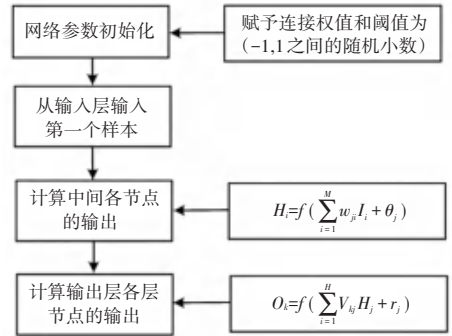


图 2 BP 神经网络信息正向传播流程

Fig. 2 Information forward propagation process of BP neural network

BP 神经网络信息的反向传播流程如图 3 所示。从图中可以看出,该流程主要包括利用计算公式,计算输出层节点的输出误差、中间层节点误差、修正输出层节点的阈值与连接权值、修正隐含节点的阈值和连接权值。

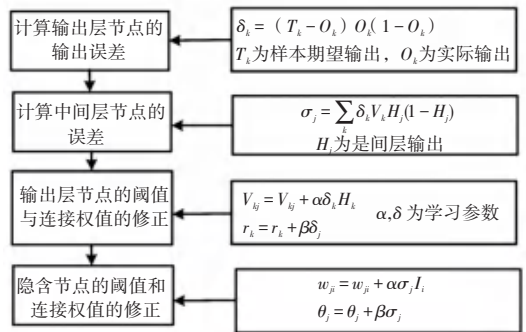


图 3 BP 神经网络误差反向传播流程

Fig. 3 Error back propagation process of BP neural network

1.2 基于 BP 神经网络的水污染监测模型构建

本研究所建立的 BP 神经网络为标准的 3 层结构。其中,模型的输入输出层主要采用线性变换,隐含层采用 tansig 非线性变换。在利用 BP 神经网络建立水系污染物预测模型时,需要数量足够多且具有典型性和高精度的水系监测数据样本。一般情况下,将数据样本分为训练样本、测试样本和检验样本。本研究以淮河某监测站点的数据为例,数据主要包括监测站点 2013 年 1 月~2018 年 2 月五年的 pH 值、COD、NH3-N 以及 DO 等 4 种水质监测项目的月均含量。为了避免训练过程中发生过拟合现

象, 需要保证测试样本与检验样本的数量不少于总数的 10%。通常来说, 训练过程中要求输入的数据位于 0~1 之间, 以此提高灵敏度和训练速度, 因而需要对输入数据进行规范化处理, 而后再将处理完成的数据输送至输出层。预处理公式如式(4)所示。

$$(x_i^{P_1})' = \frac{x_i^{P_1}}{x_{i,\max} + x_{i,\min}} \quad (4)$$

其中, $(x_i^{P_1})'$ 表示经过规范化处理的数据; $x_i^{P_1}$ 表示输入样本 P_1 第 i 个指标的污染物质量浓度; $x_{i,\max}, x_{i,\min}$ 分别表示第 i 个指标污染物质量浓度的最大值和最小值。本研究所建立的水质污染预测模型的输入层对应 4 个节点, 输出层有 6 个节点。利用试错法确定中间隐含层的节点数量, 训练 50 次的最小和最大次数结果如表 1 所示。

由表 1 可以看出, 中间层节点数为 4 和 6 时, 最大训练次数均为 ¥, 最小训练次数分别为 8 700 和 6 860, 产生了局部最小的情况, 进而造成结果的不收敛。当中间层节点数为 8 个~12 个时, 训练次数相较而言明显减小, 并且继续增加节点数目, 训练次

数变化不大, 因而中间节点数选取 10 个。

表 1 BP 神经网络水质污染预测模型训练次数与中间层的关系
Tab. 1 Relationship between training times of BP neural network water pollution prediction model and middle layer

中间层节点数	最小次数	最大次数
4	8 700	¥
6	6 860	¥
8	3 600	26 500
10	2 680	21 000
12	2 200	16 200

根据相应的模型参数, 需要对模型进行训练, 以判断水质污染情况, 并进一步得到水污染预测效果。由于数据样本较多, 因而分批次对样本进行训练。其中以 2013 年 1 月~2017 年 12 月的站点观测数据作为训练样本, 以 2018 年 1 月~2018 年 2 月的观测数据作为测试样本和检验样本。输入站点 pH 值的月均含量样本数据, 部分输入样本数据如式(5)~(8)所示。

$$P_1 = [7.188, 7.528, 7.308, 7.355, 7.405, 7.708, 7.713, 7.870, 7.785, 7.688], \quad (5)$$

$$P_2 = [7.188, 7.528, 7.308, 7.355, 7.405, 7.708, 7.713, 7.870, 7.785, 7.820], \quad (6)$$

$$P_3 = [7.405, 7.708, 7.713, 7.870, 7.785, 7.820, 7.688, 7.308, 7.443, 7.298]. \quad (7)$$

⋮

$$P_9 = [7.667, 7.639, 7.888, 7.740, 8.058, 8.038, 7.520, 7.570, 7.620, 7.463]. \quad (8)$$

在样本数据输入完成后, 需要确定输出变量。输出变量是需要分析的系统外生变量, 既可以是一个数, 也可以是多个数。研究中将某观测站点 2018 年 1 月~2018 年 2 月的 4 个水质项目监测项目分别作为各自 BP 神经网络模型的输出变量, 同时也是

检测样本, 将检测样本与实际值进行比较。检验水质污染预测效果主要是判断 BP 神经网络对于水质预测的可行性, 以及该模型是否能反映出水质监测项目的变化情况。同样地, 部分输入样本经过训练后得到的输出样本数据如式(9)~(12)所示。

$$T_1 = [7.188, 7.528, 7.308, 7.355, 7.405, 7.708, 7.713, 7.870, 7.785, 7.820], \quad (9)$$

$$T_2 = [7.405, 7.708, 7.713, 7.870, 7.785, 7.820, 7.688, 7.308, 7.443, 7.298], \quad (10)$$

$$T_3 = [7.785, 7.820, 7.688, 7.308, 7.443, 7.298, 7.243, 7.365, 7.460, 7.563], \quad (11)$$

⋮

$$T_9 = [7.740, 8.058, 8.038, 7.520, 7.570, 7.620, 7.463, 7.913, 7.770, 8.488]. \quad (12)$$

同样, 其它 3 种水质监测项目也做与 pH 值相同的处理, 从而建立 BP 神经网络监测模型实现水质污染的预测。由于研究所选取的样本数据为水质监测项目指标, 则需要按照式(4)对样本数据进行预处理。

2 基于 BP 神经网络的水系污染预测结果

数据完成预处理, 并对样本数据完成训练后, 则建立水质监测和水质污染 BP 神经网络模型, 得到部分 2013 年 1 月~2018 年 2 月的 pH 预测值如表 2 所示。

表2 基于BP神经网络的水质 pH 值预测结果

Tab. 2 Prediction results of water pH value based on BP neural network

日期	0901	0902	0903	0904	1001	1002	1003	1004
真实值	7.847 6	8.135	7.771 7	7.965	8.192	7.47	7.592 5	7.797 5
预测值	7.366 8	7.427 4	7.386 6	7.777 4	7.665 1	7.666 6	7.792 1	7.684 9
绝对误差	0.480 8	0.707 6	0.385 1	0.187 6	0.526 9	-0.196 6	-0.199 6	0.112 6
相对误差%	6.126 7	8.698 2	4.955 2	2.355 3	6.431 9	2.631 9	2.628 9	1.444 1

从表中可以看出,结果中分别包括 pH 真实值、预测值、绝对误差以及相对误差。其中,大部分日期的水质 pH 值均在 7.5 左右,仅 9 月 2 日和 10 月 1 日的水质 pH 值在 8 以上。并且多数 pH 值的绝对误差均在 1 以下,多数 pH 值的预测值低于真实值,绝对误差为正。仅 10 月 2 日和 10 月 3 日的 pH 预测值高于真实值,绝对误差为负。此外,BP 神经网络模

型的 pH 值预测结果的最大相对误差为 8.698 2%,相对误差最小值为 1.444 1%,这在一定程度上说明 BP 神经网络模型在预测水质 pH 值时的精度较高,在一定程度上反映出水质 pH 值的内在变化规律。利用同样的方法可得到 DO、NH₃-N 以及 COD 三项水质监测项目预测结果。

表3 基于BP神经网络模型的水质 DO 预测结果

Tab. 3 Do prediction results of water quality based on BP neural network model

日期	0901	0902	0903	0904	1001	1002	1003	1004
真实值	10.825	9.74	8.472 5	7.467 5	6.165	6.497 5	5.992 5	6.435
预测值	10.459 2	9.719 7	8.042 4	7.269 6	6.393 5	6.400 1	6.351 2	6.358 7
绝对误差	0.365 8	0.020 3	0.430 1	0.197 9	-0.228 5	0.097 4	-0.358 7	0.076 3
相对误差%	3.379 2	0.208 4	5.076 4	2.650 2	3.706 4	1.499	5.985 8	1.185 7

表2为建立神经网络预测模型后得到的水质 DO 值预测结果。从表中可见,该水质项目指标分别包括 DO 真实值、预测值、绝对误差和相对误差。其中,DO 真实值变化范围在 5~10 之间,且大部分 DO 值的预测值和真实值之间的绝对误差为正,数值小于 1,仅 10 月 1 日和 10 月 3 日的绝对误差为

负,分别为-0.228 5和-0.358 7。此外,BP 神经网络模型的 DO 值预测结果的最大相对误差为 5.985 8%,相对误差,最小值为 0.208 4%,并且多数 DO 值的相对误差仅为 3.836 5%。可以表明 BP 神经网络模型用于预测水质 DO 是可行的,其预测精度较高。

表4 基于BP神经网络的水质 NH₃-N 预测结果Tab. 4 Prediction results of water quality NH₃-N based on BP neural network

日期	0901	0902	0903	0904	1001	1002	1003	1004
真实值	0.662 5	1.277 5	1.057 5	1.212 5	1.562 5	1.327 5	1.155	0.615
预测值	0.684 6	1.279 6	1.064	1.207 7	1.353 7	1.418 4	1.168 3	0.675 7
绝对误差	-0.022 1	-0.002 1	-0.006 5	0.004 8	0.208 8	-0.090 9	-0.013 3	-0.060 7
相对误差%	3.335 9	0.164 4	0.614 7	0.395 9	13.363 2	6.847 5	1.151 5	9.869 9

表3为建立神经网络预测模型后得到的水质 NH₃-N 值预测结果。该水质项目指标分别包括

NH₃-N 真实值、预测值、绝对误差和相对误差。该水质项目的预测值与真实值之间的相对误差最大值

为13.363 2%,相对误差最小值为0.164 4%。计算可得相对误差的平均值为 4.53%,表明 BP 神经网络模型用于预测水质 NH₃-N 也是可行的,其预测精度较高,能在一定程度上反映出该项目值的内在规律。

水质中的 COD 预测结果如表 4 所示。预测值与真实值之间的相对误差最大为7.352 8%,相对误差最小为0.075 6%,表明 BP 神经网络模型用于预测水质 COD,其预测精度较高,能够很好地反映该值的内在变化规律,具有可行性的。

表 5 基于 BP 神经网络的水质 COD 预测结果

Tab. 5 Prediction results of water quality cod based on BP neural network

日期	0901	0902	0903	0904	1001	1002	1003	1004
真实值	4.45	5.6	4.975	4.5	3.675	5.3	5.275	4.625
预测值	4.122 8	5.570 4	4.901 1	4.496 6	3.703	5.314 1	5.290 9	4.613 1
绝对误差	0.327 2	0.029 6	0.073 9	0.003 4	-0.028	-0.014 1	-0.015 9	0.011 9
相对误差%	7.352 8	0.528 6	1.485 4	0.075 6	0.761 9	0.266	0.301 4	0.257 3

3 结束语

监测和预测水系化学污染物有利于管理水环境和治理水污染。本研究在人工神经网络模型的基础上,构建了水系化学污染的 BP 神经网络智能预测模型。在此基础上,以淮河水系某监测站点的水系 pH 值、DO、NH₃-N、以及 COD 等监测项目为例,验证了该预测水系化学污染的 BP 神经网络模型的有效性,得到相应的预测结果。研究表明,对于衡量水系化学污染的四个指标,BP 神经网络的预测精度均较高。本次研究验证了 BP 神经网络模型应用于水系化学污染物预测的有效性和可行性,但其广泛适用性还有待进一步研究。

参考文献

- [1] 李国平,王奕淇. 地方政府跨界水污染治理的“公地悲剧”理论与中国的实证[J]. 软科学,2016,30(11):24-28.
- [2] 刘仁涛,姜继平,史斌,等. 突发水污染应急处置技术方案动态生成模型及决策支持软件系统[J]. 环境科学学报,2017,37(2):763-770.

- [3] 袁圃,毛剑琳,向凤红,等. 改进的基于遗传优化 BP 神经网络的电网故障诊断[J]. 电力系统及其自动化学报,2017,29(1):118-122.
- [4] 丁凡,黄立勇,王锐,等. 中国 2004—2015 年突发水污染事件监测数据分析[J]. 中国公共卫生,2017,33(1):59-62.
- [5] 孙宝磊,孙嵩,张朝能,等. 基于 BP 神经网络的大气污染物浓度预测[J]. 环境科学学报,2017,37(5):1864-1871.
- [6] 王东,赵越,姚瑞华. 论河长制与流域污染防治规划的互动关系[J]. 环境保护,2017,45(9):17-19.
- [7] 刘明喆,孔凡青,张浩,等. 基于层次分析法和模糊综合评价的突发水污染风险等级评估[J]. 水电能源科学,2019,37(1):53-56.
- [8] 赵领娣,徐乐. 基于长三角扩容准自然实验的区域一体化水污染效应研究[J]. 中国人口·资源与环境,2019,29(3):50-61.
- [9] 牛桂敏,郭珉媛,杨志. 建立水污染联防联控机制促进京津冀水环境协同治理[J]. 环境保护,2019,47(2):64-67.
- [10] 张青,王学雷,张婷,等. 基于 BP 神经网络的洪湖水水质指标预测研究[J]. 湿地科学,2016,14(2):212-218.
- [11] 徐晓梅,吴雪,何佳,等. 滇池流域水污染特征(1988-2014 年)及防治对策[J]. 湖泊科学,2016,28(3):476-484.
- [12] 游丹丹,陈福集. 基于改进粒子群和 BP 神经网络的网络舆情预测研究[J]. 情报杂志,2016,35(8):156-161.