

DOI: 10.13718/j.cnki.xdzk.2019.12.015

GA-BP 优化 TS 模糊神经网络水质监测 与评价系统预测模型的应用

——以太湖为例^①

肖金球^{1,3}, 周翔^{1,3}, 潘杨^{2,4}, 冯威¹, 陈多观^{1,3}

1. 苏州科技大学 电子与信息工程学院, 江苏 苏州 215009; 2. 江苏省环境科学与工程重点实验室, 江苏 苏州 215009;
3. 江苏省工程技术中心, 江苏 苏州 215009; 4. 苏州科技大学 环境生物科技研究所, 江苏 苏州 215009

摘要: 针对水质监测与评价系统在太湖应用过程中水质数据和水质等级评价不准确的问题, 建立了一种多隐含层改进型 GA-BP 神经网络来辨识复杂的水质模型, 以均方误差 MSE 作为个体适应度, 并在权值调整过程中加入动量因子来加快收敛速度, 获取最优权阈值, 提高其拟合程度和泛化能力. 根据校准后水质的 pH、溶解氧、浊度和氨氮数据, 利用 TS 模糊神经网络建立了适用于当地水质评价的模型. 仿真测试结果充分说明改进型 GA-BP 优化 TS 模糊神经网络对复杂水质模型的拟合程度更高, 水质数据的均方误差、绝对误差更小, 绝对误差保持在 1.5% 以内, 水质等级预测精度提高 14.28%.

关键词: 水质评价; 多参数; 遗传算法; BP 神经网络; TS 模糊神经网络

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-9868(2019)12-0110-10

太湖作为中国五大淡水湖之一, 位处江浙两省之间. 从上个世纪 80 年代开始, 太湖的水污染问题日益严峻, 水质污染严重. 现有的现场水质监测技术与遥感技术对于水体监测缺乏连续性和准确性, 而利用电子信息技术和互联网技术, 以传感器为节点的水质监测系统则实现对太湖不同区域地表水的实时监测. 长期使用中水质传感器易受外界影响导致测量不准确, 传统评价方法容易受主观因素影响, 使得传统水质监测系统在对水质等级综合评价中表现不佳. 随着人工智能的发展, 越来越多的人将神经网络运用到水质评价系统中, 但是由于非线性水质模型具有复杂性、多变性以及不确定性, 简单的 BP 神经网络已经无法获得准确的水质数据, 因此无法对水质情况进行有效的监测和评价.

本文提出一种改进型 GA-BP 算法优化的 TS 模糊神经网络(TSFNN), 并应用到多参数水质监测与评价系统中. 利用人工神经网络对多输入多输出模型有着很好的辨识效果, 本文通过改进型 GA-BP 算法先对不同传感器采集到的多维水质数据进行校准处理, 消除温漂等诸多干扰因素带来的误差, 系统再根据网络校准后得到的近似真实值的数据, 通过 TS 模糊神经网络来综合判断传感器所在水域的水质等级.

1 系统设计

1.1 硬件总体结构

多参数水质监测系统需要对水质的 pH 值、DO(溶解氧)、浊度和氨氮 4 个参数实时监测, 将传感器测

^① 收稿日期: 2019-01-10

基金项目: 国家科技部重大水专项(2017ZX07205003); 江苏省产学研前瞻性联合基金项目(BY2011132); 江苏省研究生创新与教改项目(091580001); 苏州科技大学研究生创新工程基金项目(SKCX17_025).

作者简介: 肖金球(1963—), 男, 硕士, 教授, 主要从事智能仪器与自动检测系统研究.

通信作者: 周翔.

量的数据通过数据控制端及时存储,并通过 GPRS 模块发送至监测端,监测中心的人员能够根据获得的数据定位到异常水域位置.

系统的硬件设计是基于 ARM 嵌入式技术,分别通过玻璃电极型 pH 传感器、极谱式薄膜探体溶解氧传感器、散射光式浊度传感器、分光光度式氨氮传感器和防水型 DS18B20 温度传感器来测量水质的 pH 值、DO 值、浊度、氨氮值和水温,将获得的数据首先发送到低通滤波电路和 50 Hz 工频陷波电路来消除信号中存在的高频毛刺及外界信号的干扰和温漂误差,然后分别通过各类参数的调理电路对传感信号进行优化处理,接着经过 A/D 模块对数据进行高精度的模数转换发送至主处理器 920TS3C2410,最后通过 GPRS 模块将测量的水质参数信息发送回监测端进行分析和存储.系统硬件结构图如图 1 所示.

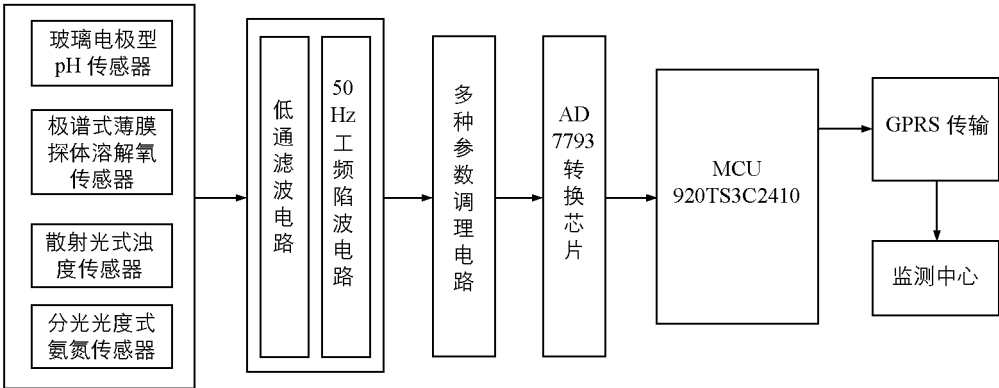


图 1 硬件结构图

2 数据获取与预处理

系统内包含多个采样点,每个采样点拥有多个传感器测量水质的 pH 值、DO 值、浊度值、氨氮值.系统每 4 h 测量一次数据,通过从 WEB 监测平台随机导出 2 200 组测量值,从中选取 2 000 组数据用于 GA-BP 神经网络的训练,200 组数据作为测试数据,部分数据如表 1. 根据青悦开放环境数据中心提供的水质数据作为输出的期望数据,用于检验 GA-BP 网络部分对非线性系统的拟合能力.

表 1 水质传感器数据

时间	2016/08/01 4: 00	2016/08/01 8: 00	2016/08/01 12: 00	2016/08/01 16: 00	...	2017/03/01 4: 00	2017/03/01 8: 00	2017/03/01 12: 00	2017/03/01 16: 00	...	2017/08/03 4: 00	2017/08/03 8: 00	2017/08/03 12: 00	2017/08/03 16: 00
pH	7.21	7.19	7.03	7.04	...	8.41	8.05	8.05	8.07	...	6.49	7.39	7.34	7.41
DO	8.82	8.84	8.45	8.61	...	10.17	10.23	10.2	10.24	...	6.11	7.61	8.03	8.18
浊度	2.88	2.88	2.88	2.88	...	2.1	2	2.1	2.2	...	2.36	2.54	2.51	2.46
氨氮	0.08	0.11	0.23	0.11	...	0.08	0.1	0.09	0.11	...	0.18	0.25	0.26	0.2

地表水环境质量标准(GB3838-2002)的分类标准见表 2,通过等隔均匀分布方式内插水质指标标准数据生成 400 组数据,用于建立 TS 模糊神经网络^[1],其中 350 组用于训练,其余 50 组用于检测网络是否有良好的泛化能力和判断准确性.

表 2 地表水分类标准

项 目	分 类				
	I 类	Ⅱ类	Ⅲ类	Ⅳ类	V 类
溶解氧(≥)	7.5	6	5	3	2
氨氮(≤)	0.15	0.5	1	1.5	2
浊度(≤)	3	10	20	100	500

注: pH 值在 I 类到 V 类均为 6~9.

其中, I 到 II 类, II 到 III 类, III 到 IV 类及 IV 到 V 类数据各 100 组, 利用网络输出值来判断该区域水质等级, 若该值在 B-1 类和 B 类之间, 则判定该区域水质等级为 B 类^[2-5]. 此外为了防止各维数据之间的数量差过大, 导致神经网络在预测过程中误差过大, 无法有效地拟合有效的非线性函数, 本文采用最大最小法对输入输出数据进行预处理, 将样本的数据进行归一化处理^[6].

3 神经网络构造与改进

本文采用神经网络来实现数据的多输入多输出并行处理, 根据传感器实际数据预测出接近真实数据的预测数据, 将预测值作为水质评价的输入, 通过 TS 模糊神经网络的模糊化层和模糊化规则计算层得到由校准后数据评价的水质等级, 以此来评定该水域的水质状况^[7-9]. GA-BP 优化 TS 模糊神经网络算法流程如图 2 所示.

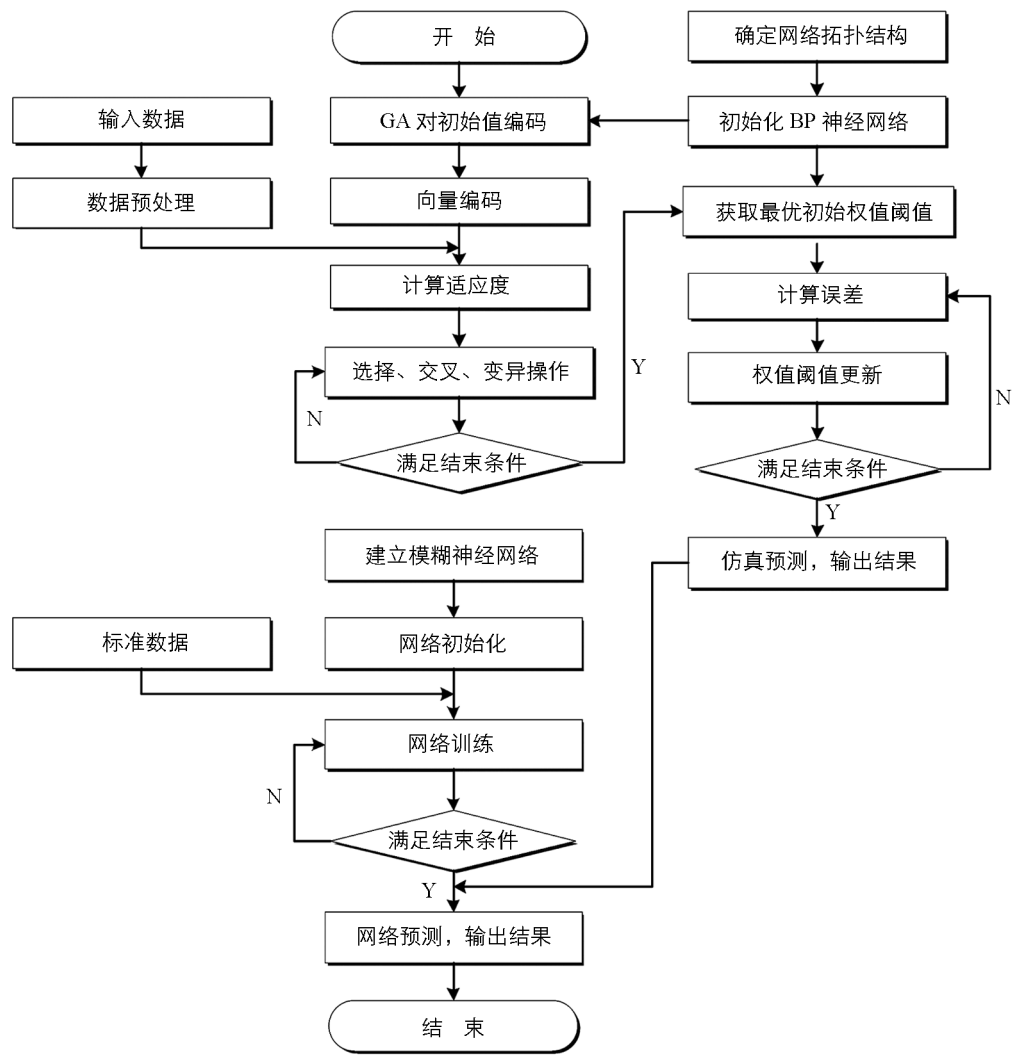


图 2 算法流程图

3.1 网络结构

改进型 GA-BP 算法优化的 TS 模糊神经网络在多参数水质评价中分为两个部分: 改进型 GA-BP 神经网络与 TS 模糊神经网络, GA-BP 神经网络结构分为输入层, 隐含层和输出层. 由于监测系统测量 4 种参数的数据, 输入层为 4 个节点. 隐含层的节点数影响着网络的预测精度和训练时间, 其选取先利用以下公式确定隐含层节点数的大致范围, 再用试凑法来确定最佳数目^[10].

$$l=\sqrt{(m+n)}+a$$

(1)

$$l=2n+1$$

(2)

式中 n 为输入层节点数, m 为输出层节点数, a 为 0~10 之间的常数. 通过 MATLAB 仿真测试, 可以得到 BP 神经网络预测误差与隐含层节点数的关系如图 3 所示.

由图 3 可知, 本文中神经网络的隐含层节点数定为 7, 此时神经网络中 4 个水质参数的预测误差最小. 由于水质模型存在着较为复杂的映射关系, 普通的单层隐含层结构无法满足预测精度, 所以神经网络拟采用双隐含层. 本文根据预测误差百分比、均方差、运行时间 3 个指标来衡量网络的性能, 此处的预测误差百分比取 4 个节点误差百分比平均值. 在 MATLAB 中可以通过配置 S 向量来实现双隐含层.

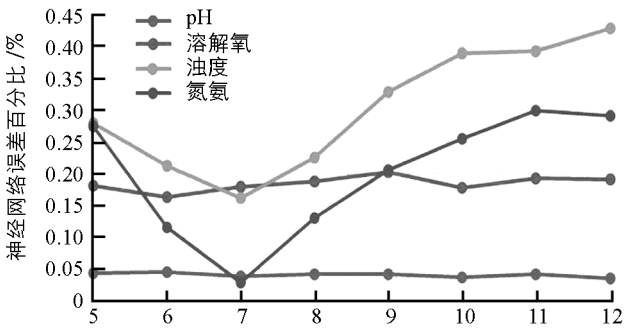


图 3 预测误差与隐含层节点数的关系图

根据仿真测试, 当 GA-BP 神经网络的结构为 4-7-6-4 时, 预测精度最高. 两者性能比较如表 3 所示:

表 3 预测结果比较

网络类型	预测误差百分比/%	均方误差	运行时间/s
单隐含层 BP 神经网络	2.28	0.056 2	2.944
双隐含层 BP 神经网络	1.45%	0.023 6	3.272

TS 模糊神经网络(TSFNN)结构分为输入层、模糊化层、模糊规则计算层和输出层. 输入层与输入向量相连接, 它的各个节点与输入分量 X_i 连接, 则输入节点数为 4, 但因为 pH 在水质等级评价标准中没有明确界限, 则设置网络输入节点数为 3. 模糊化层的目的是通过隶属度生成函数实现对输入值各个分量的模糊化, 得到每个输入分量的模糊隶属度. 计算层则通过连乘该层的节点数计算得到 $\omega^{[12]}$. 因为输出层为水质等级, 其节点个数为 1, 根据输入输出节点数的个数, 确定隶属度函数个数为 7, 因此确定网络结构为 3-7-1, 隶属度函数的中心与宽度以及系数则通过初始化的方式求得.

3.2 网络参数优化

BP 神经网络的参数包括权阈值、学习率、迭代次数、训练目标最小误差等, 这些参数的选取直接影响神经网络整体性能. 学习率越大, 权值的变化率越大, 其取值范围一般在[0.01, 0.1]之间. 本文学习率取 0.05, 迭代次数取 200, 训练目标的允许误差为 0.000 01.

BP 神经网络算法的权值阈值经过随机初始化产生, 在训练过程中容易遇到局部最小陷阱, 导致训练效果不佳. GA 算法将网络中所有的权值和阈值组成种群个体, 通过计算个体的适应度并选取在群体中的最优个体, 最终将最优个体赋值给网络的初始权值和阈值, 提高网络的整体拟合程度^[13]. 经典的 GA-BP 网络通常以误差绝对值和作为个体适应度值, 而均方误差 MSE 作为估计量与被估计量的差异程度, 更能反映出个体的最优程度. 本文以均方误差 MSE 作为个体适应度值, 当进化次数为 14 次时, 个体适应度取到最小值为 0.054 22.

GA 算法的种群规模越大, 越容易求解最优个体, 也增大了计算量. 种群的分布要保持相对均匀, 如果过于集中, 则需要增大变异概率, 减少交叉概率. 如果过于分散, 则反之. 本文根据现有水质模型算法参数, 并结合太湖水质数据的实际情况不断调整, 最终确定 GA 算法的群体规模取 50, 种群中个体长度为 135, 迭代次数取 100, 交叉概率 0.4 和变异概率为 0.25.

传统的 BP 神经网络通过最速下降法将权值向着负梯度方向改变, 在网络训练过程中容易出现收敛缓慢、振荡等现象. 通过在权值和阈值的调整中增加附加动量的学习公式(3), 使得网络在修改权值时考虑以前积累的学习经验. 网络动量因子取值为 0.95.

$$w(k) = w(k - 1) + \Delta w(k) + a[w(k - 1) - w(k - 2)]$$

(3)

式中, a 为动量学习率. 最后将节点的输入数据、改进的 GA-BP 网络预测的数据、期望数据分别进行水质等级评价, 比较前两者的水质等级精确度^[14].

4 应用测试与分析

为了更加直观地分析改进后监测与评价系统在实际应用中的效果, 通过实验来获取传感器数据, 并进行 Matlab 仿真与分析. 在苏州西山太湖的不同位置放置检测系统的传感器节点(图 4)及室外控制单元(图 5), 设定系统数据采样间隔为 4 h, 分别采集河流的 pH 值、DO 值、浊度、氨氮 4 个数据, 通过 3G 网关节点(图 6)将传感器数据传输到监测终端(图 7), 以 WEB 的形式显示出各节点采集的数据情况,

选择某个节点, 导出其在 2016/08/01 到 2017/08/03 的 2 200 组数据, 并选择同时间段内青悦开放环境数据中心提供的水质数据作为真实值来进行网络训练.



图 4 水质传感器



图 5 室外控制单元

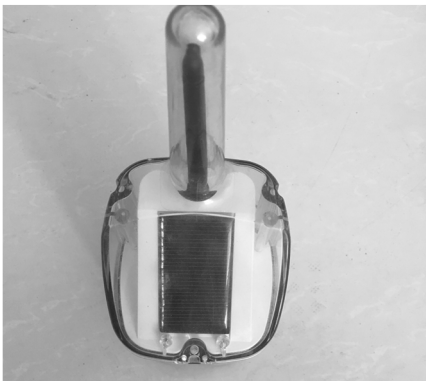


图 6 3G 网关节点

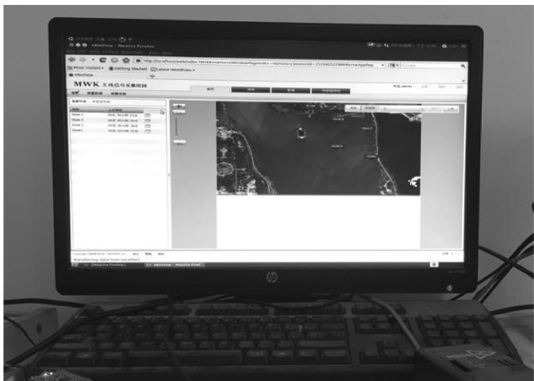


图 7 监测终端

4.1 算法性能分析

在 Matlab R2015a 软件上分别使用 BP 神经网络算法、GA-BP 神经网络算法与改进型 GA-BP 神经网络算法对水质模型进行拟合, 通过比较 BP、GA-BP 和改进 GA-BP 神经网络的绝对误差与绝对误差百分比来分析两者对复杂的水质模型的拟合状况. 图 8 和图 9 分别是 GA-BP 神经网络与改进 GA-BP 神经网络在训练过程中的均方误差变化图.

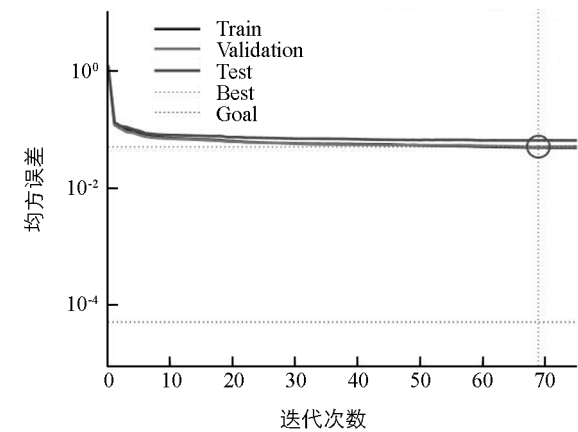


图 8 GA-BP 均方误差图

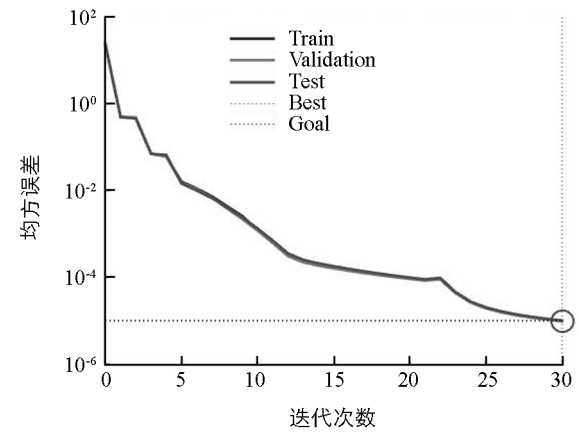


图 9 改进 GA-BP 均方误差图

由图 8 和图 9 可知，当 BP 神经网络在迭代次数为 69 次时，其均方误差达到最小值 0.0505。当改进型 GA-BP 神经网络在迭代次数为 30 次时，其均方误差达到最小值 0.0235，相比前者，迭代次数减少了 56%，均方误差减少了 53%，大大提高了网络的预测精度。

如图 10 所示，普通 BP 神经网络在对水质模型进行拟合和 4 个参数预测时，pH 和溶解氧的误差基本维持在 0.2 以内，氨氮的误差较小，浊度的误差波动较大，四者的误差百分比均大于 2%。根据图 11 可知，GA 优化的 BP 网络在复杂水质模型拟合上，误差基本在 0.1 和 0.2 之间波动，浊度预测误差仍然波动较大，但整体明显小于未优化前。4 个水质参数在 GA-BP 网络中的预测误差百分比均控制在 1.5% 左右，说明该网络对水质模型具有良好的训练能力与预测能力，GA 优化 BP 网络相比于 BP 网络，在预测精度上分别提高了 21.9%，30.2%，38.4%，23.9%。

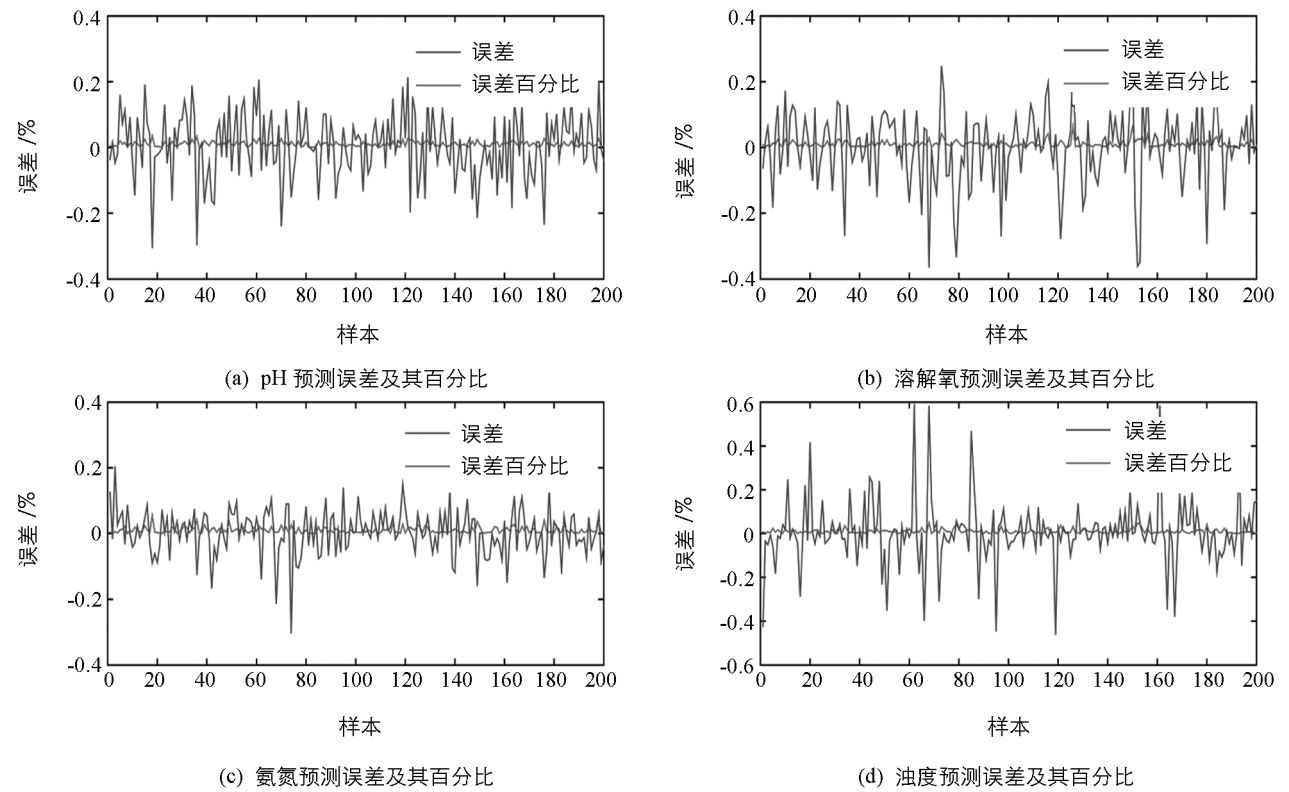


图 10 BP 网络预测图

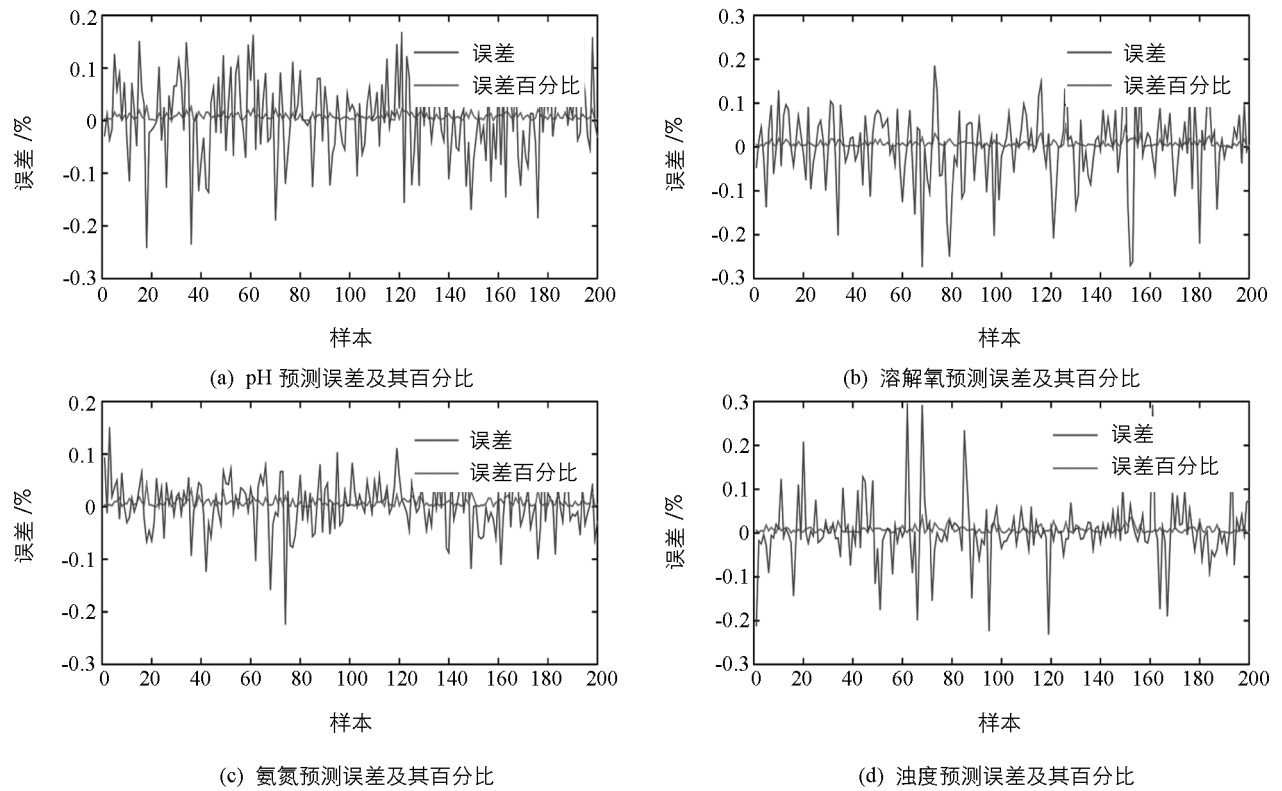


图 11 GA-BP 网络预测图

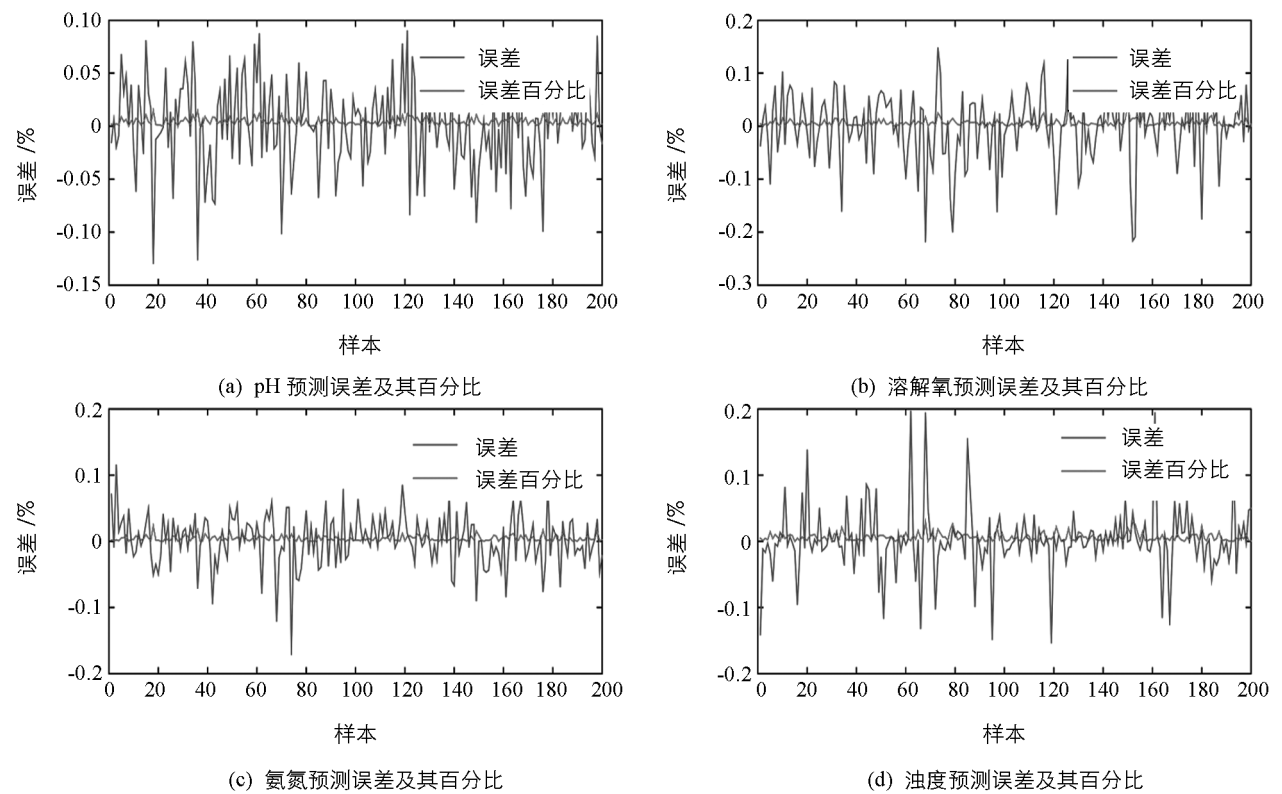


图 12 改进 GA-BP 网络预测

由图 12 可知，以均方误差 MSE 作为个体适应度值的改进型 GA-BP 网络对水质模型进行拟合后，网络预测绝对误差波动振幅减小，整体呈现更加平稳态势。其中水质参数 pH 值、溶解氧和氨氮的预测误差

基本保持在 ± 0.1 以内,而浊度的预测误差大部分在 ± 0.1 以内,但波动较大,个别误差接近 ± 0.2 .改进型 GA-BP 网络预测误差百分比与改进前的 GA-BP 网络相比分别减少了 25.7%,29.2%,23.2%,25.6%.三者误差百分比如表 4 所示:

表 4 预测绝对误差百分比对比图

预测误差百分比	pH 值/%	溶解氧/%	氨氮/%	浊度/%
BP 神经网络	2.14	2.35	2.03	2.46
GA-BP 神经网络	1.67	1.64	1.25	1.87
改进 GA-BP 神经网络	1.24	1.16	0.96	1.39

4.2 优化后模糊神经网络水质评价分析

利用 Matlab 软件和等间隔插入法得到的 400 组标准水质数据对 TS 模糊神经网络进行训练与测试,网络训练次数为 200 次. TS 模糊神经网络在标准数据训练过程中,根据计算网络训练数据预测的均方误差为 0.002 312,绝对误差百分比为 1.23%,结果可知预测输出对实际输出拟合程度较高,水质等级预测误差较小.训练后的网络再使用 50 组标准数据进行测试,测试结果的误差曲线平稳且在 0 附近小范围上下波动,根据计算可得均方误差为 0.002 022,误差百分比为 1.12%. TS 模糊神经网络训练完成后,将系统中的 200 组测试数据、优化后的 200 组测试数据分别用于网络输入进行水质等级评价,结果如图 13 所示.

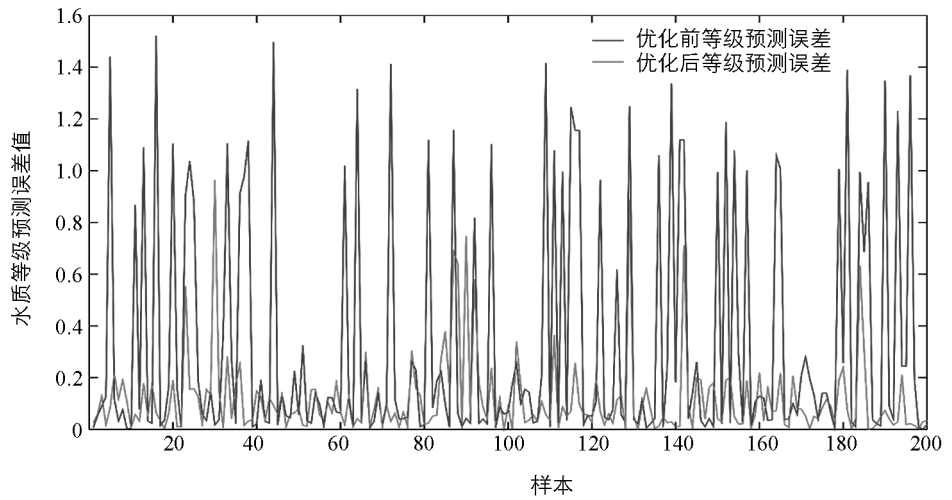


图 13 优化前后预测误差图

因为水质等级为 1~5 的正整数,所以当 TS 模糊神经网络预测误差值大于 1 时,会影响系统评价结果.由图 13 可知原始数据得到的水质等级预测误差波动较大且影响结果的样本数为 25,其精确度 87.5%,而经过改进型 GA-BP 网络优化后的水质等级预测误差整体上小于优化前,误差波动减小且影响结果的样本数为 0,其精确度达到 100%,较前者提高了 14.28%.

5 结 论

本文所述的多参数水质监测与评价系统^[15-16]已经取得实用新型专利,针对其在长期使用过程中出现的数据不准确和水质等级评价不正确的状况,本文采用改进型 GA-BP 神经网络优化 TS 模糊神经网络对多输入多输出的多参数水质监测系统进行校准和对测量区域的水质等级进行评估,通过 MATLAB 仿真与实

际情况相结合,可以得到以下结论:

- 1) 均方误差 MSE 作为一种描述估计量与被估计量之间的差异程度的度量,相比于误差绝对值和,当 MSE 作为个体适应度值去求最优个体时,网络得到的水质参数预测值更加接近真实值.
- 2) 水质评价系统的准确度不仅仅取决于网络的性能,更由网络输入数据的精确性决定.通过改进型 GA-BP 神经网络对传感器数据进行校准预处理,可以大大提高系统水质等级评价的准确性.
- 3) 对于多输入多输出的复杂水质模型,具有不同节点的多隐含层网络比传统的单层网络泛化能力更强,预测精度更高,有效地拟合了非线性水质模型.
- 4) 用于评价水质等级的参数越多,等级评价就越精确.而多参数水质监测系统传感器节点仅能测量水质的 4 种参数且仅有 3 种数据可以用于水质评价,所以一定程度上影响系统对水质评价的精确性.

参考文献:

[1] 王 英,陈聪伟,肖金球. 水质多参数监测系统设计 [J]. 苏州科技学院学报(自然科学版), 2015, 32(1): 54-58, 63.

[2] SANDHYA SAMARASINGHE. Neural Networks for Applied Sciences and Engineering-From Fundamentals to Complex Pattern Recognition [M]. Beijing: China Machine Press, 2010.

[3] ALVISI S, FRANCHINI M. Fuzzy Neural Networks for Water Level and Discharge Forecasting with Uncertainty [J]. Environmental Modelling & Software, 2011, 26(4): 523-537.

[4] 李 福,郭 健. 改进的 BP 神经网络算法在水质监测中的应用 [J]. 计算机系统应用, 2015, 24(10): 243-247.

[5] 马正华,王 腾,杨 彦,等. BP 神经网络模型在太湖出入湖河流水质预测中的应用 [J]. 计算机应用与软件, 2013, 30(11): 172-175.

[6] 王泽平. 基于 GA-BP 与多隐层 BP 网络模型的水质预测及比较分析 [J]. 水资源与水工程学报, 2013, 24(3): 154-160.

[7] 唐源英,王海洋. 汇水区土地利用方式对水库水质的影响 [J]. 西南大学学报(自然科学版), 2017, 39(11): 134-141.

[8] 丁金婷,臧泽林,黄 敏. 模糊方法改进的反向传输神经网络预测南美白对虾养殖的水质 [J]. 浙江大学学报(农业与生命科学版), 2017, 43(1): 128-136.

[9] YANG Y Q, WANG G J, YANG Y. Parameters Optimization of Polygonal Fuzzy Neural Networks Based on GA-BP Hybrid Algorithm [J]. International Journal of Machine Learning & Cybernetics, 2014, 5(5): 815-822.

[10] SEDKI A, OUAZAR D, EL MAZOUZI E. Evolving Neural Network Using Real Coded Genetic Algorithm for Daily Rainfall-Runoff Forecasting [J]. Expert Systems with Applications, 2009, 36(3): 4523-4527.

[11] XU L Q, LIU S Y. Study of Short-Term Water Quality Prediction Model Based on Wavelet Neural Network [J]. Mathematical & Computer Modelling, 2013, 58(3-4): 807-813.

[12] LOHANI A K, KUMAR R, SINGH R D. Hydrological Time Series Modeling: A Comparison Between Adaptive Neuro-Fuzzy, Neural Network and Autoregressive Techniques [J]. Journal of Hydrology, 2012, 442-443: 23-35.

[13] HUANG M Z, MA Y W, WAN J Q, et al. Simulation of a Paper Mill Wastewater Treatment Using a Fuzzy Neural Network [J]. Expert Systems with Applications, 2009, 36(3): 5064-5070.

[14] HAN H, LI Y, QIAO J. A Fuzzy Neural Network Approach for Online Fault Detection in Waste Water Treatment Process [J]. Computers & Electrical Engineering, 2014, 40(7): 2216-2226.

[15] 雷 岩,肖金球,陈多观,等. 蜂窝式无线通讯的水质参数远程监测系统, ZL201420369965.3 [P]. 2014. 11. 05.

[16] 郭鹏飞,温志渝,周 颖,等. 基于树莓派的远程水质监测系统的设计 [J]. 重庆理工大学学报(自然科学版), 2018, 32(4): 186-192.

Application of a GA-BP Optimized TSFNN Water Quality Monitoring and Evaluation System Prediction Model

——A Case Study of Taihu Lake

XIAO Jin-qiu^{1,3}, ZHOU Xiang^{1,3}, PAN Yang^{2,4},
FENG Wei¹, CHEN Duo-guan^{1,3}

1. Department of Electronic and Engineering, Suzhou University of Science and Technology, Suzhou Jiangsu 215009, China;

2. Jiangsu Key Laboratory of Environment Science and Engineering, Suzhou Jiangsu 215009, China;

3. Jiangsu Engineering Technology Center, Suzhou Jiangsu 215009, China;

4. Environment Biotechnology Research Institute, Suzhou University of Science and Technology, Suzhou Jiangsu 215009, China

Abstract: The present Water Quality Monitoring and Evaluation System which is applied in Taihu Lake has the disadvantage of inaccurate water quality data and water quality rating. To solve this problem, a multi-hidden layer improved GA-BP neural network is proposed in this paper for identifying the complex water quality model. MSE (mean square error) is used as the individual fitness, and a momentum factor is added to the weight adjustment process to speed up the convergence for obtaining the optimal weight and threshold and improving its goodness of fit and generalization ability. Based on the calibrated data of pH, dissolved oxygen, turbidity and ammonia nitrogen of the water, and using the TS fuzzy neural network, a model suitable for evaluating the local water quality is established. The results of a simulation test show that the improved GA-BP optimized TS fuzzy neural network has a higher degree of fitting for the complex water quality model, and the mean square error and the absolute error of the water quality data are smaller. In addition, the absolute error percentage is kept within 1.5%, and the water quality level prediction accuracy is improved by 14.28%.

Key words: water quality evaluation; multi-parameter; genetic algorithm; BP neural network; TS fuzzy neural network

责任编辑 汤振金