

BP神经网络与多元线性回归的净水装置脱氮预测比较

李金泽^{1,2}, 叶 春³, 李春华³, 罗艳丽¹, 魏伟伟³

(1. 新疆农业大学草业与环境科学学院, 新疆 乌鲁木齐 830052;
2. 燕京理工学院建筑学院, 河北 廊坊 065201; 3. 中国环境科学研究院, 北京 100012)

摘要: 地表水净化装置在实际运行过程中, 其最终出水水质会受多种因素影响制约, 且该过程具有非线性、时变性与随机性, 无法进一步分析该装置水质净化效果。对此, 文章分别提出基于 BP 神经网络和多元线性回归对地表水净化装置水净化效果进行预测。结果表明, BP 神经网络与多元回归预测 TN 准确度都较为良好, 通过将拟合曲线得到的 TN 预测值与 TN 实际值进行绝对误差和相对误差分析。说明 BP 神经网络在 TN 预测上优于多元线性回归, 更适用于该地表水净化装置脱氮过程预测的拟合计算。

关键词: BP 神经网络; 多元线性回归; 绝对误差; 相对误差

中图分类号: X524

文献标志码: A

DOI: 10.16803/j.cnki.issn.1004-6216.2018.05.010

Comparison of BP Neural Network and Multiple Linear Regression in the Prediction of Nitrogen Removal Efficiency in a Surface Water Purification Device

Li Jinze^{1,2}, Ye Chun³, Li Chunhua³, Luo Yanli¹, Wei Weiwei³

(1. College of Grassland and Environmental Science, Xinjiang Agricultural University, Urumqi 830052, China;
2. School of Architecture, Yanching Institute of Technology, Langfang 065201, China;
3. Chinese Research Academy of Environmental Sciences, Beijing 100012, China)

Abstract: In the actual operation process, water quality of a surface water purification device will be restricted by many factors, and the process is nonlinear, time varying and stochastic, so it is impossible to analyze the effects of water purification device. Therefore, in this paper, a prediction model of water purification effects based on BP neural network and multiple linear regression is presented. The results show that the predicted results of BP neural network and multiple linear regression are good. Finally, fitted curve is used to obtain the TN predicted value and TN actual value for analysis of the absolute error and the relative error, indicating that the BP neural network is superior to the multiple linear regression in TN prediction and more suitable for fitting calculation of predicting the nitrogen removal process of the surface water purification device.

Keywords: BP Neural Network; Multiple Linear Regression; Absolute Error; Relative Error

CLC number: X524

在水环境领域中, 针对地表水处理的水净化装置测试的研究报道屡见不鲜^[1], 但对于装置运行中水质指标预测研究却不多见, 而大多数水质指标预测研究都主要集中在污水处理系统上, 最常见的预测方法就是神经网络。21 世纪以来, 张文艺等^[2]通过 BP 神经网络模型对污水处理厂的运行进行模拟, 使活性污泥法工艺系统在线智能控制由猜想变为可能。辛刚等^[3]研发一种人工神经

网络模型的化学强化一级处理系统, 使 CEPT 系统可以在线实时控制。杨维维等^[4]利用递归高阶神经网络对污水处理过程的 COD、BOD、氨氮以及悬浮固体等参数一一建模, 充分反映了污水处理系统的动态性能, 大大提高了参数预测结果的精度。白桦等^[5]针对混凝投药过程的特点, 利用神经网络的学习能力和记忆功能, 建立了净水厂混凝投药量的预测模型, 使预测结果的准确性明显提高。

收稿日期: 2018-04-27

基金项目: 国家水体污染控制与治理科技重大专项(2017ZX07203-005) 基金资助

作者简介: 李金泽(1994-), 男, 硕士、助教。研究方向: 市政工程和装置相关模型理论研究与应用。E-mail: LJZ_zky@163.com

通信作者: 叶 春(1970-), 男, 博士、研究员。研究方向: 湖泊富营养化治理和生态恢复理论与技术。E-mail: yechbj@163.com

文章采用实验装置由中国环境科学研究院,湖泊水污染治理与生态修复技术国家工程实验室研制,装置分为进水区、布水区、净水区、排水区4部分。桶内的主要部分是微生物营养补充仓,主要作用是培养土著微生物对水体进行净化,同时有泵体对水进行循环可以增加水中溶解氧浓度。实验地点选在江苏省宜兴市的市政府池塘内,池塘水质总体呈浅绿,净化前的TN在1.8~2.1 mg/L之间。由于地表水净化装置的水净化过程复杂,无法直接预测其出水TN情况,故选取多元线性回归和BP神经网络两种方法进行预测,相比于直接使用BP神经网络预测,使用两种方法进行比较,更能说明其中一种方法在水净化预测上的优越性。通过其预测效果选取合适的方法作为该装置的水净化预测方法。

1 两种脱氮预测模型

1.1 多元线性回归模型

多元线性回归的基本原理和基本计算过程与一元线性回归相同,但由于自变量个数多,计算相当麻烦。事实上,一种现象常常是与多个因素相联系的,由多个自变量的最优组合共同来预测或估计因变量,比只用一个自变量进行预测或估计更有效,更符合实际,自变量主要由实际影响因素决定数量^[6],方程如下:

$$y_{hat} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + E$$

多元线性回归模型包含多个解释变量^[7],多个解释变量同时对被解释变量 y_{hat} 发生作用,若要考察其中一个解释变量对 y_{hat} 的影响就必须假设其它解释变量保持不变来进行分析。因此多元线性回归模型中的回归系数为偏回归系数,即反映了当模型中的其它变量不变时,其中一个解释变量对因变量 y_{hat} 均值的影响。

1.2 BP神经网络模型

BP(Back Propagation)神经网络^[8]是一种按误差逆传播算法训练的多层前馈网络,是应用最广泛的神经网络之一。BP算法是一种有监督式的学习算法,其主要原理是输入学习样本,通过反向传播算法对网络的权值和偏差进行反复的调整训练,最终使输出的结果与期望的结果尽可能接近^[9-10],具体步骤,见图1。

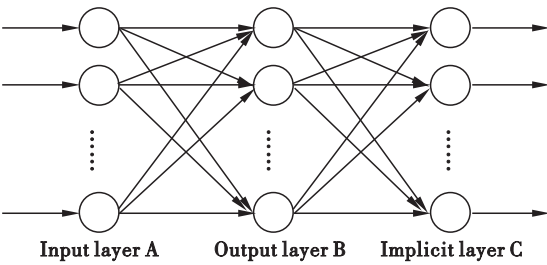


图1 三层BP神经网络结构

假设输入层、隐含层和输出层的单元数分别为 I, J 和 K ,输入为 $(x_0, x_1, x_2, \cdots, x_{I-1})$,隐含层输出为 $(h_0, h_1, h_2, \cdots, h_{J-1})$,网络实际输出为 $(y_0, y_1, y_2, \cdots, y_{K-1})$, $(d_0, d_1, d_2, \cdots, d_{K-1})$ 表示训练样本的期望输出。输入层单元 i 到隐含层单元 j 的权值为 v_{ij} ,隐含层单元 j 到输出层单元 k 的权值为 w_{jk} ,用 Θ_j 和 Θ_k 来分别表示隐含层单元和输出层单元的阈值。

2 基于多元线性回归净水装置脱氮预测

通过整理2016年5月23日到7月12日得到市政府池塘TN、氨氮、硝氮、COD_{Mn}等指标的数据,为保证数据训练及检验的准确性,取当天监测指标平均值对应当天各点所有指标,其他一监测指标在某一采样点的监测值均采用这种方法,并以时间节点进行划分。以各指标平均值为例,将数据进行一一对应整理,见表1。

表1 市政府池塘各指标监测平均值

时间段	统计表			mg·L ⁻¹
	TN	氨氮	硝氮	
1	1.840	0.390	0.160	6.00
4	1.923	0.454	0.212	6.15
7	1.700	0.420	0.200	5.94
10	1.608	0.380	0.180	5.54
17	1.550	0.372	0.160	5.51
24	1.450	0.340	0.160	5.36
31	1.430	0.332	0.150	5.14
38	1.413	0.320	0.140	4.63
45	1.380	0.270	0.138	4.41
52	1.250	0.260	0.130	3.97

之后,建立回归方程,建模过程由SPSS软件完成,输入数据即可得到分析结果。回归方程如下:

$$y = -893.065X_1 - 289.349X_2 - 254.164X_3 - 8.267X_4 + 8145.648$$

标准化处理:

$$y = -0.874X_1 - 0.517X_2 - 0.504X_3 - 0.129X_4$$

其中, X_1 、 X_2 、 X_3 、 X_4 分别是净化前 TN、氨氮、硝氮、 COD_{Mn} 的数值, y 是净化后的 TN 值。

回归方程的 F 检验结果见表 2, 可知方程的

回归效果有效, 可以作为预测方程, R^2 满足要求, 并且偏大, 说明拟合效果较好。

通过计算得出所建立的回归模型的预测的样本数据中的所有实测值与预测值的拟合曲线见图 2, 绝对误差和相对误差见表 3。

表 2 脱氮阶段回归方程的 F 检验表

自由度	$F_{0.05}$	$F_{0.05}$ 计算	Sig	R	R^2	R^2 预测
(4, 10)	3. 49	22. 667	0. 000	0. 871	0. 829	0. 782

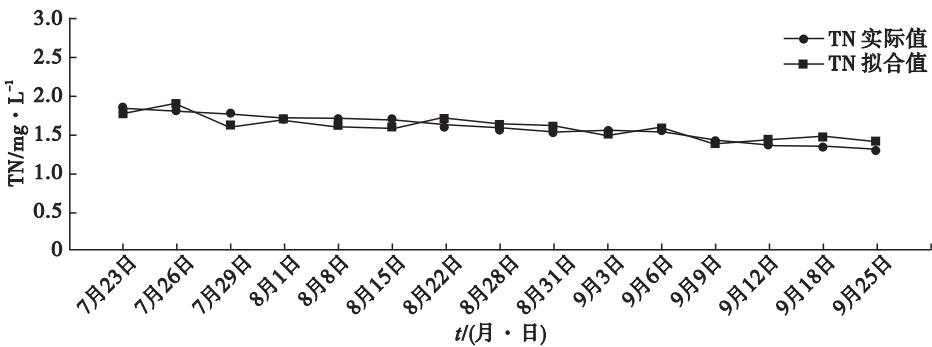


图 2 TN 实际值与拟合值对比

表 3 装置实际 TN 值及多元线性回归拟合值

$t/\text{月} \cdot \text{日}$	TN _{实际} / $\text{mg} \cdot \text{L}^{-1}$	预测值 / $\text{mg} \cdot \text{L}^{-1}$	绝对 误差	相对误 差/%
7 月 23 日	1. 840	1. 719	-0. 121	-7. 04
7 月 26 日	1. 816	1. 872	0. 056	2. 99
7 月 29 日	1. 772	1. 558	-0. 214	-13. 74
8 月 1 日	1. 708	1. 503	-0. 205	-13. 64
8 月 8 日	1. 704	1. 606	-0. 098	-6. 10
8 月 15 日	1. 700	1. 554	-0. 146	-9. 40
8 月 22 日	1. 608	1. 753	0. 145	8. 27
8 月 28 日	1. 562	1. 687	0. 125	7. 41
8 月 31 日	1. 556	1. 693	0. 137	8. 09
9 月 3 日	1. 546	1. 476	-0. 070	-4. 74
9 月 6 日	1. 538	1. 792	0. 254	14. 17
9 月 9 日	1. 428	1. 318	-0. 110	-8. 35
9 月 12 日	1. 376	1. 489	0. 113	7. 59
9 月 18 日	1. 350	1. 517	0. 167	11. 01
9 月 25 日	1. 304	1. 421	0. 117	8. 23

由图 2 可见, 预测值与实测值拟合程度, 总体趋势的拟合是准确的, 但还是有部分预测不准确, 具体表现为预测值与具体值间还是有一定差距的。通过表 3 可以看出, 多元线性回归预测的结果与实际值相比误差较大, 已超出规定的 10%, 最大相对误差为 13. 74%, 相对误差在 10% 以上的占总数的 27%, 且相对误差在 5% 以上的占总数的 87%。

3 基于 BP 神经网络净水装置脱氮效果预测

3. 1 基于 BP 神经网络的装置脱氮数据前处理

同样, 整理 2016 年 5 月 23 日到 7 月 12 日的数据, 将它们平均分成 10 组, 进行 BP 预测分析。先以平均值为例, 进行 BP 神经网络训练和测试, 检验 9 组和 10 组 TN 的预测与实际情况, 通过回归对可决系数进行求解, 观察拟合度是否在误差范围内。见图 3。

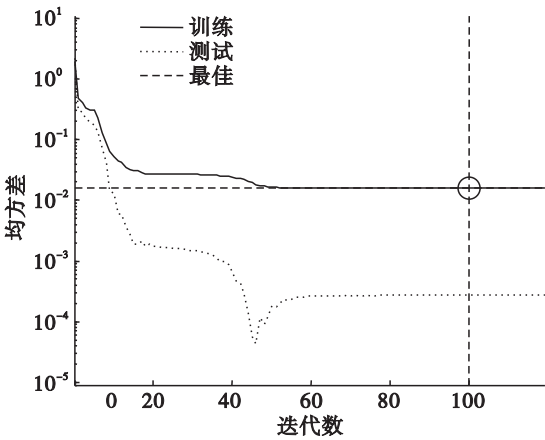


图 3 BP 神经网络运行训练测试迭代图

通过验证可以得到均方差训练结果, BP 网络训练到第 100 代的时候 BP 训练结果最理想, 均方差为 0. 016。

3.2 基于 BP 神经网络的装置脱氮预测效果

通过训练 66 组监测数据,确定各参数,对之后装置监测数据进行模拟预测,最终模拟值与实际值对比情况,见图 4 和表 4。

由图 4 可见,预测值与实测值拟合程度相比于多元线性回归预测的结果更好,总体趋势的拟

合是十分准确的。从表 4 可以看出,数据模拟误差范围为 ±10%,有 7 次模拟结果低于实际值,8 次高于实际值。其中,最小误差为 1.79%,最大误差为 9.52%,相对误差超过 5%的只占 40%,明显优于多元线性回归预测的结果。

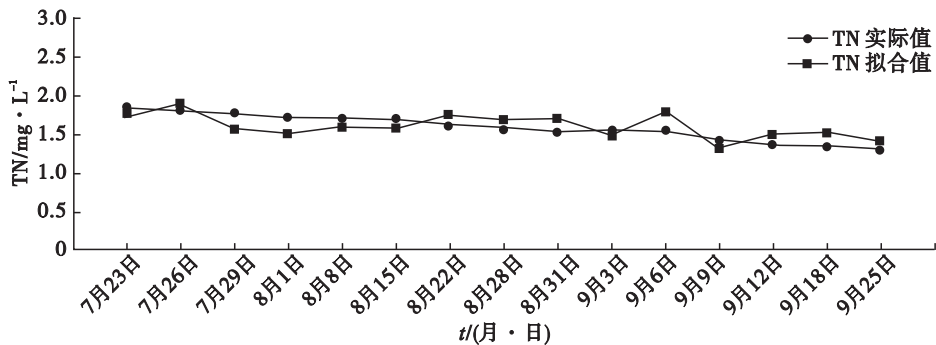


图 4 TN 实际值与拟合值对比

表 4 装置实际 TN 值及 BP 神经网络计算拟合值

t/月·日	TN 实际 /mg·L ⁻¹	BP 拟合值 /mg·L ⁻¹	绝对 误差	相对误 差/%
7 月 23 日	1.762	1.847	-0.077 8	-4.43
7 月 26 日	1.893	1.816	0.077	4.07
7 月 29 日	1.618	1.772	-0.154	-9.52
8 月 1 日	1.678	1.708	-0.03	-1.79
8 月 8 日	1.617	1.704	-0.087	-5.38
8 月 15 日	1.589	1.702	-0.111	-6.99
8 月 22 日	1.721	1.608	0.113	6.57
8 月 28 日	1.626	1.562	0.064	3.94
8 月 31 日	1.618	1.556	0.062	3.83
9 月 3 日	1.497	1.546	-0.049	-3.27
9 月 6 日	1.584	1.538	0.046	2.90
9 月 9 日	1.385	1.428	-0.043	-3.10
9 月 12 日	1.426	1.376	0.05	3.51
9 月 18 日	1.467	1.351	0.117	7.98
9 月 25 日	1.411	1.304	0.107	7.58

4 结论

综上,从地表水净化装置的模拟预测结果可知,出水的实际 TN 与两种方法预测的 TN 误差也均在一定范围内,均符合预测范围值。但是,基于 BP 神经网络的该装置预测 TN 相比于多元线性回归的方法具有较高的精度,可适用于水处理的水质模拟。这是由于多元线性回归主要研究线性问

题,而 BP 神经网络主要解决非线性问题,因此认为 BP 神经网络更适用于该装置脱氮预测,这也为今后其它指标的预测以及多个时间段的预测提供了预测方法上的借鉴。

参 考 文 献

[1]焦李成,杨淑媛,刘芳,等.神经网络七十年:回顾与展望[J].计算机报,2016,39(8):1697-1716.

[2]张文艺,钟梅英,蔡建安.活性污泥法系统人工神经网络模型[J].给水排水,2002,28(6):12-15.

[3]辛刚,蒋文举,金燕,等.用神经网络辨识化学强化一级处理系统[J].环境科学与技术,2002,25(4):10-13.

[4]杨维维,乔俊飞.基于递归高阶神经网络的污水处理系统建模[J].信息控制,2011,40(5):710-714+720.

[5]白桦,李硅白.混凝投药智能控制系统实现方法的探讨[J].给水排水,2003,29(8):81-84.

[6]陈威,艾婵.基于多元线性回归模型的武汉市水资源承载力研究[J].河南理工大学学报(自然科学版),2017,36(1):75-79.

[7]Malik Samander Ali,Farooq Assad,Gereke Thomas,et al. Prediction of blended yarn evenness and tensile properties by using artificial neural network and multiple linear Rregression[J]. Autex Research Journal, 2016,16(2):43-50.

[8]章毅,郭泉,王建勇.大数据分析的神经网络方法[J].工程科学与技术,2017,49(1):9-18.

[9]张景阳,潘光友.多元线性回归与 BP 神经网络预测模型对比与运用研究[J].昆明理工大学学报(自然科学版),2013,38(6):61-67.

[10]He G,Huang C,Guo L,et al. Identification and adjustment of guide rail geometric errors based on BP neural network[J]. Measurement Science Review,2017,17(3):135-144.