

熵值理论和改进 BP 网络在湖泊富营养化评价中的应用研究

于凤存¹, 方国华^{2,3}, 鞠 琴³

(1.安徽省·水利部淮河水利委员会科学研究所(水利水资源安徽省重点实验室),安徽 蚌埠 233000;

2.河海大学水利水电工程学院,江苏 南京 210098;

3.河海大学水文水资源与水利工程科学国家重点实验室,江苏 南京 210098)

摘要:针对湖泊型饮用水源地水体污染、富营养化加剧的问题,引入熵值理论,建立单指标营养状态指数(TSI)和熵权耦合的湖泊综合营养状态指数模型(STSI),计算得到湖泊综合富营养状态指数判断湖泊富营养综合状态;基于神经网络仿真理论和 Matlab 软件系统,采用附加动量法和自适应学习速率改进 BP 算法,建立 5-3-1 结构型式的 BP 网络模型对湖泊富营养状态进行仿真预测。综合富营养化指数模型及改进 BP 模型应用于评价及预测固城湖富营养状态,并对模型评价结果进行验证。结果表明,改进 BP 网络模型可以有效地综合判断水体状态,为富营养评价及预测提供新的方法。

关键词:湖泊富营养化;营养状态指数;熵权;仿真系统;神经网络;BP 算法

中图分类号:X524

文献标识码:A

文章编号:1000-0852(2014)03-0057-04

我国社会经济的快速发展给生态环境造成了一定的压力,湖泊水体富营养问题日益突出,尤其是作为集中式饮用水源地湖泊的富营养化引起了人们的重视和关注。水体富营养加剧,在削减湖泊功能的同时,也给人们的生产生活带来严重危害^[1-2],不利于社会持续稳定和谐发展,又造成了一定的经济损失。因此,科学评价湖泊水库富营养化,可以及时掌握水体富营养状态,为湖泊生态环境的管理提供理论基础。

作为环境评价内容之一的湖泊富营养化评价受到国内外专家学者的关注,提出了很多预测评价方法,主要有特征法、参数法、多元回归、生物指标法等,这些方法均有适用的条件和局限性^[3],至今还没有形成统一的、确定的湖泊富营养化评价方法。基于熵值理论和神经网络的发展与广泛应用,引入熵值理论,建立熵权卡尔森湖泊综合营养状态指数模型,可准确判断湖泊富营养化状态。此外,采用改进动量和自适应学习率的 BP 网络仿真预测模型,以固城湖为例进行研究,指出熵权卡尔森指数模型和改进 BP 网络模型的特点与实用性。

1 基于熵值理论的营养状态综合指数模型

卡尔森指数 TSI(trophic state index)^[4]是美国科学家卡尔森于 1977 年提出的,以透明度(SD)为基准的湖泊营养状态评价指数,见式(1)。

$$\begin{cases} TSI(SD)=10(6-\frac{\ln SD}{\ln 2}) \\ TSI(Chl)=10(6-\frac{0.68\ln chl a}{\ln 2}) \\ TSI(TP)=10(6-\frac{\ln(48/TP)}{\ln 2}) \end{cases} \quad (1)$$

基于卡尔森指数模型,学者金相灿通过对我国湖泊监测资料的统计分析,提出了适用于我国湖泊的部分参数的营养状态指数 TSI^[5]:

$$\begin{cases} TSI(TP)=10(9.436+1.488\ln TP/\ln 2.5) \\ TSI(TN)=10(5.453+1.694\ln TN/\ln 2.5) \\ TSI(COD)=10(0.109+2.438\ln COD/\ln 2.5) \\ TSI(BOD_5)=10(2.118+2.363\ln BOD_5) \\ TSI(NH_3-N)=10(7.77+1.511\ln HN_3-N/\ln 2.5) \end{cases} \quad (2)$$

为了综合评判湖泊富营养化状况,构建湖泊富营

收稿日期:2013-08-10

基金项目:国家自然科学基金(41105077)

作者简介:于凤存(1979-),女,山东聊城人,博士,主要从事农田水利、水旱灾害及生态环境方面的研究工作。E-mail:fcyhu@126.com

养化综合指数状态公式,见式(3):

$$STSI = TSI_M(\Sigma) = \sum_{j=1}^m (W_j \times TSI_M(j)) \quad (3)$$

式中:STSI为湖泊营养综合状态指数;TSI_M(j)代表第j种参数的营养状态指数;W_j为第j种参数的营养状态指数的相关权重。

鉴于监测指标TP、TN、COD和BOD₅具有实测监测数据,可通过监测数据之间的相关关系求得权重,如熵值法、主成分分析法、因子分析法、复相关系数法、变异系数法、模糊聚类分析法、等级相关法、秩和比法、双级值距离法等^[6-8],其中熵值法利用指标信息效用值(即差异度)来计算各指标的客观权重,计算结果可信度大,自适应功能强,现已在工程技术、社会经济等领域得到了更多的应用,故选用熵值法求营养状态指数的权重。关于熵权计算见参考文献[9-10]。

2 改进BP网络仿真预测模型

目前已有学者对BP提出了改进方案,如在修改权值中加入“动量项”^[11],采用共轭梯度法,采用Cauchy误差估计器代替传统的LMS误差估计器^[12]等。采用附加动量法^[13]和自适应学习速率改进BP算法^[14],从而构建改进BP湖泊富营养预测模型。

2.1 附加动量法

在逆传播法的基础上在每一个权值变化值上加上一项正比于前次权值变化的值(动量项),并根据逆传播法来产生新的权值和阈值变化。带有附加动量因子的权值调节公式为:

$$\Delta w_{ij}(t+1) = (1-mc)\eta\delta_i p_j + mc\Delta w_{ij}(t) \quad (4)$$

$$\Delta b_i(t+1) = (1-mc)\eta\delta_i + mc\Delta b_i(t) \quad (5)$$

式中:t为训练次数;Δw_{ij}为权值变化量;Δb_i为阈值变化量;δ_i为误差项;p_j为第j个输入变量;η为学习率;mc为动量因子。

2.2 自适应学习速率

在循环训练中根据权值的变化调整学习速率,它检查权值的修正值是否根据权值的变化作调整,它检查权值的修正值是否真正降低了误差函数,如果确实如此,则说明所选取的学习速率值小了,可以对其增加一个量;否则就产生了过调,应减小学习速率。自适应学习速率的调整公式为:

$$\eta(k+1) = \begin{cases} \alpha\eta(k) & E(k+1) < E(k) \\ \beta\eta(k) & E(k+1) > 0.4E(k) \\ \eta(k) & \text{other} \end{cases} \quad (6)$$

式中:η为学习率;t为训练次数;E(t)为累计误差平方和;α和β分别为递乘因子和递减因子。

湖泊富营养化改进BP预测模型,网络计算模式分为训练学习和仿真测试两个过程。传递函数取用log sig(n),训练函数取用trainbpx。训练学习就是把训练样本输入matlab工具箱模块中,自动适应的建立输入层与隐含层、隐含层与输出层之间的权值和阈值矩阵,当权值和阈值达到稳定时,即获得了BP网络从输入到输出的映射关系。仿真测试过程就是利用已训练好的网络对评价样本进行仿真测试。

3 湖泊富营养化预测仿真

卡尔森指数法结合熵权理论,通过式(2)计算各指标营养状态指数,然后计算熵权,由式(3)计算富营养化状态综合指数。在此不再具体详述。

3.1 数据资料

固城湖位于高淳县境南部,高淳区西南部境内,属青戈江、水阳江水系的一个洪水滞留型浅水湖泊,是高淳区最重要的集中式饮用水水源地,其维持着全区66%人口饮用、75%的工业和建筑业产值、83%的农田灌溉用水,在保证全区生活用水、社会经济发展和生态环境平衡等方面起到了不可替代的作用。近10年来,随着固城湖水域资源的利用及区域经济的迅猛发展,工业点源污染、农业面源污染、居民生活污染和湖内网围养殖的二次污染等,导致大量污染物及氮、磷等营养物质汇入湖体,进而造成固城湖水环境质量的明显下降,尤其是富营养化严重,且呈加剧发展趋势。

固城湖富营养预测仿真采用总磷(TP)、总氮(TN)、化学需氧量(COD)、生化需氧量(BOD)和氨氮(NH₃-N)五项,监测指标统计资料来源于高淳区水务局和环保局提供的资料。2005年选取大湖区、港口河口、胥河口、漆桥河口、湖心区、拦河网、小湖区作为监测断面进行了逐月监测,从中选取25个实测样本数据资料及其营养状态综合指数归一化后,作为学习样本对神经网络进行训练,归一化公式(7)^[15]。按照丰水期、平水期、枯水期和综合值通过加权平均得到各个指标加权平均值,归一化后见表1。

$$x_i = \frac{0.1*(x_i - x_{\max}) + 0.9*(x_{\min} - x_i)}{x_{\min} - x_{\max}} \quad (7)$$

3.2 BP仿真预测过程

选择结构简单实用的输入层、隐含层和输出层的3层神经网络模型,选取TP、TN、COD、BOD和NH₃-N

表 1 固城湖水质参数平均值与相应的归一化数值
Table 1 The means of water quality for Guchenghu lake and normalization

数据类型	评价期	TP	TN	COD	BOD	NH ₃ -N	STSI
原始监测数据	丰水期	0.12	1.00	15.64	1.63	0.40	57.81
	平水期	0.04	1.14	12.11	1.60	0.23	52.05
	枯水期	0.02	2.28	15.57	1.43	0.14	52.31
	综合值	0.08	1.32	14.44	1.56	0.29	55.42
归一化数据	丰水期	0.74	0.34	0.69	0.53	0.52	0.74
	平水期	0.28	0.38	0.38	0.50	0.34	0.47
	枯水期	0.16	0.72	0.68	0.36	0.24	0.48
	综合值	0.50	0.43	0.58	0.47	0.41	0.63

注:原始监测数据各个指标的单位均为 mg/l,归一化数据各个指标为无量纲数据。

作为 5 个输入节点;经过调试,选定隐含层包含 3 个神经元;输出层有 1 个神经元,即湖泊富营养化所对应的数值,从而建立 5-3-1 型式的神经网络结构。

始学习速率取 0.01,学习速率增比为 1.05,学习速率降比为 0.7,动量因子为 0.9,计算期望误差采用 0.03,利用上述训练样本对 BP 网络进行训练,网络在训练了 930 次后,网络的总均方误差小于 0.03,停止运算,训练学习阶段结束,25 个学习样本训练值与计算值的拟合曲线见图 1。学习结束时的权值是 3*5 和 1*3,阈值 θ_1 和 θ_2 分别为 3*1 和 1*1。

$$w_1 = \begin{bmatrix} 4.1022 & -0.6774 & 4.4042 & 0.7244 & 6.1935 \\ -3.7929 & 4.6762 & -3.5271 & 2.1474 & 1.3877 \\ 2.2369 & 4.3090 & 5.3837 & 5.4928 & -3.6091 \end{bmatrix}$$

$$w_2 = [3.6971 \quad 1.8583 \quad -3.1999]$$

$$\theta_1 = [-4.7732 \quad 1.5374 \quad 1.9333]^T$$

$$\theta_2 = -0.5275$$

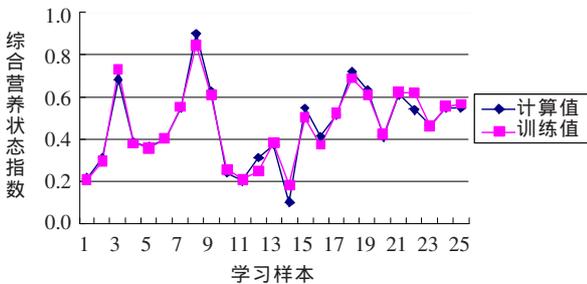


图 1 学习样本计算值和训练值结果比较

Fig. 1 Comparison between the computation and training of learning samples

3.3 结果分析

湖泊富营养化状态与评分值对应表 2,由以上两种计算方法计算结果见表 3。

(1)集成应用熵值理论及改进卡尔森单指标模型,

表 2 湖泊水体富营养化综合状态指数 STSI 与水质关系
Table 2 The correlation between STSI and lake water quality

营养状态分级	基于熵权的卡尔森模型	水质评价
贫营养	$0 < STSI \leq 30$	优
中营养	$30 < STSI \leq 50$	良好
轻度富营养	$50 < STSI \leq 60$	轻度污染
中度富营养	$60 < STSI \leq 70$	中度污染
重度富营养	$70 < STSI \leq 100$	重度污染

表 3 固城湖水体富营养化评价结果
Table 3 Comparison between STSI and BP

营养等级 (综合营养状态指数)	丰水期	平水期	枯水期	综合
STSI 综合状态指数	57.81	52.05	52.31	55.42
归一化综合状态指数	0.74	0.47	0.48	0.63
改进 BP 仿真值	0.67	0.51	0.52	0.75
相对误差/%	-8.5	9.7	7.2	19.2
评价等级	轻度富营养	轻度富营养	轻度富营养	轻度富营养

构建富营养综合指数模型 STSI,为综合判断湖泊富营养状态提供了依据,采用熵值理论计算各个评价因子的权重,避免了权重确定中的主观因素,使 STSI 计算结果更加符合实际。

(2)从结果来看,改进 BP 湖泊富营养预测模型对富营养综合状态指数的预测相对误差在 0~19.2%,基本满足了预测需求。该模型的成功与否与样本数据的真实性有很大关系,模型的确定很大程度上建立在试错的基础上,涉及到最佳输入模式的识别、隐含层个数和隐含节点的确定,还有权值和阈值的判定,由于试错法一般难以得到最优值,在一定程度上影响了模型的精度,但可以作为判断湖泊富营养化状态的一种方法。

(3)固城湖富营养化程度在不同的时段具有一定的差异。丰水期、平水期、枯水期和综合状态下均进入轻度富营养状态,且丰水期和综合值略偏向于中度富营养状态,可见,固城湖整体状况已经进入轻度营养~中度富营养状态,水质受到轻度污染。应结合工程和非工程措施改善水质,如控制水产养殖规模、降低面源污染程度、建设生态修复和保护工程、实施湖底清淤等工程措施,完善水源地保护管理制度、提高环境管理能力等。

4 结论

(1)单指标营养状态指数(TSI)和熵权耦合的湖泊营养综合指数模型(STSI),利用熵值理论计算各个指标的熵权,从数据本身反映的信息无序化效用值得到权重系数,规避了主观定权重,使评价结果更加客观,

计算简单方便,无需长系列监测样本。

(2) 改进附加动量因子和自学习速率湖泊富营养化 BP 网络仿真预测模型,具有较高的学习精度和准确率。可以针对评价的重点选取水质参数进行学习,改进的 BP 神经网络预测模型摒弃了较多主观的因素,具有客观性和唯一性,具有广泛的适应性,其中一系列学习样本的获取是重要的基础工作。

参考文献:

- [1] 金相灿. 湖泊富营养化研究中的主要科学问题 [J]. 环境科学学报, 2008,28(1):21-23. (JIN Xiangcan. The key scientific problems in lake eutrophication studies[J]. Acta Scientiae Circumstantiae,2008,28(1):21-23. (in Chinese))
- [2] 王东丽,翟胜,王巨媛,等. 东昌湖水质动态变化与富营养化评价[J]. 生态科学, 2013,32(1):63-67. (WANG Dongli, ZHAI Sheng, WANG Juyuan, et al. Analysis of water quality and eutrophication in the Dongchang Lake[J]. Ecological Science, 2013,32(1):63-67. (in Chinese))
- [3] 张宝,刘静玲. 湖泊富营养化影响与公众满意度评价方法[J]. 水科学进展, 2009,20(5): 695-700. (ZHANG Bao, LIU Jingling. Evaluation method for lake eutrophication influence and public satisfaction [J]. Advances in Water Science, 2009,20(5):695-700. (in Chinese))
- [4] Robert E. Carlson. A trophic state index for lakes[J]. Limnology and Ceanography, 1997,22 (2):361-369.
- [5] 金相灿. 中国湖泊富营养化[M]. 北京: 中国环境科学出版社, 1990: 134. (JIN Xiangcan. Lake Eutrophication in China[M]. Beijing: China Environmental Science Press, 1990:134. (in Chinese))
- [6] Carrizosa E, Conde E, Fernandez R, Puerto J. Multi-criteria analysis with pertain information about the weighting coefficients [J]. European Journal of Operation Research, 1995,81(2): 291-301.
- [7] Cook W, John D, Rodney G, Kress M. Multi-criteria modeling and ordinal date: evaluation in terms of subsets of criteria [J]. European Journal of Operation Research, 1997, 98(3): 602-607.
- [8] Trainphyllou E. A sensitivity analysis approach for some deterministic multi-criteria decision making methods[J]. Decision Sciences, 1997, 28 (1):151-194.
- [9] Yu Fengcun, Fang Guohua, Ru Xiangwen. Eutrophication, health risk assessment and spatial analysis of water quality in Gucheng Lake, China [J]. Environ. Earth Sci., 2010, 9(8):1741-1748.
- [10] 周惠成, 张改红, 王国利. 基于熵权的水库防洪调度多目标决策方法及应用 [J]. 水利学报, 2007,38(1):391-397. (ZHOU Huicheng, ZHANG Gaihong, WANG Guoli. Multi-objective decision making approach based on entropy weights for reservoir flood control operation[J]. Journal of Hydraulic Engineering, 2007,38(1):391-397. (in Chinese))
- [11] Rumelhart D E, Hilnton G E, Williams R J. Learning internal representations by error propagation in Parallel[A]. Distributed Processing [C]. MIT Press, 1986.
- [12] 彭松, 方祖群. BP 神经网络学习算法的联合优化 [J]. 电路与系统学报, 2000,5(3):26-30. (PENG Song, FANG Zuqun. The joint optimization of BP learning algorithm[J]. Journal of Circuits and Systems, 2000,5(3):26-30. (in Chinese))
- [13] 吴建华, 魏茹生, 赵海生, 等. 神经网络模型下的土石坝安全监测仿真研究 [J]. 系统仿真学报, 2008,20(4):1052-1055. (WU Jianhua, WEI Rusheng, ZHAO Haisheng, et al. Computer simulation research on embankment dam safety monitoring used operational nerval network model [J]. Journal of System Simulation, 2008,20(4):1052-1055. (in Chinese))
- [14] 丛爽. 面向 MATLAB 工具箱的神经网络理论与应用[M]. 合肥: 中国科学技术大学出版社, 2003. (CONG Shuang. Neural Network Theory and Applications with Matlab Toolboxes[M]. Hefei: Press of University of Science and Technology of China, 2003. (in Chinses))
- [15] 鞠琴, 郝振纯, 余钟波, 等. 基于人工神经网络的降雨径流模拟研究[J]. 辽宁工程技术大学学报, 2007,26(6):940-943. (JU Qin, HAO Zhenchun, YU Zhongba, et al. Study on rainfall-runoff simulation based on artificial neural networks[J]. Journal of Liaoning Technical University,2007,26(6):940-943. (in Chinese))

Study on Lake Eutrophication Based on Entropy Theory and Improved BP Neural Network

YU Fengcun¹, FANG Guohua^{2,3}, JU Qin³

(1. Anhui Province Key Laboratory of Water Conservancy and Water Resources, Anhui and Huaihe River Institute of Hydraulic Research, Bengbu 233000, China; 2. College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China; 3. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China)

Abstract: Aiming at the water quality pollution and eutrophication problems for drinking water source of lake type, the synthesized trophic state index (STSI) model was established based on trophic state index (TSI) and entropy weights to evaluating lake eutrophication status through calculating STSI. Based on neural network simulation theory and Matlab software, BP algorithm model was improved through additional momentum method and the learning rate self-adjustment. Improved BP model of 5-3-1 type was established to simulate and predict lake eutrophication. The synthesized trophic state index (STSI) model and improved BP model were applied to Gucheng Lake to evaluate and predict trophic state, and the model results were verified. The results show that the improved BP model can determine the water body's state, which provides a new method to evaluate and predict eutrophication.

Key words: lake eutrophication; trophic state index; entropy weight; simulation system; neural network; back-propagation algorithm