

文章编号: 2095—0411 (2013) 03 - 0062 - 04

# 基于 BP 神经网络的太湖富营养化时空变化预测<sup>\*</sup>

马正华, 王 腾, 周炯如

(常州大学 信息科学与工程学院, 江苏 常州 213164)

**摘要:** 建立拓扑结构为 5-9-1 的 3 层 BP 神经网络模型, 预测太湖的富营养化水平, 为水环境监测平台提供决策支持。该模型采用动量梯度下降法对太湖的湖体的富营养化水平进行预测并评价。验证结果表明, 训练后的网络得出的预测结果与实际值很接近。因此, 采用 BP (Back Propagation) 神经网络对太湖湖体富营养化水平进行预测并评价是一种有效的方法。

**关键词:** 太湖; 富营养化; BP 神经网络; 预测

中图分类号: TP 399

文献标识码: A

doi: 10.3969/j.issn.2095—0411.2013.03.016

## Spatiotemporal Changes Prediction of Taihu Lake Based on BP Neural Network

MA Zheng-hua, WANG Teng, ZHOU Jiong-ru

(School of Information Science and Engineering, Changzhou University, Changzhou 213164, China)

**Abstract:** BP (Back Propagation) artificial neural network has excellent nonlinear approximation ability and can be used for prediction. In this paper, a BP neural network model with the topology of 5-9-1 is established for prediction and evaluation. Test results showed that this model is an effective way in predicting the eutrophication level of Taihu Lake, and the error between network output value and actual value is very little, so this model can be used in routine work for prediction and give necessary support to the water environment platform.

**Key words:** Taihu Lake; eutrophication; BP neural network; prediction

太湖是我国平原地区五大淡水湖之一。太湖流域是国家对外开发前沿和经济增长最具活力的地区, 在我国国民经济发展中具有举足轻重的作用。太湖流域的社会经济发展与优越的水资源条件密不可分<sup>[1]</sup>。多年来, 特别是近 30 多年来, 以城市化、工业化进程加速和土地的快速非农化流转为标志的社会经济高速发展, 导致流域的水质污染日益严重、洪涝灾害损失不断增加等诸多问题。随着社会的不断发展, 人口的不断增加, 工农业水平不

断提高, 大量的污染物排入太湖, 导致太湖以及周边河流水质不断恶化, 太湖水体营养化程度不断加重, 危害极大。2007 年 5 月 29 日, 太湖提前发生蓝藻, 使无锡城区及周围地区居民引用水困难, 多家工厂停工, 损失巨大。太湖的水环境问题已经严重影响到社会经济的发展, 成为各级政府部门的工作重点, 也成为老百姓最关心的环境问题之一。

湖泊富营养化现象的发生, 主要是湖水中氮、磷等营养元素增加, 致使水体中自氧型生物 (主要

<sup>\*</sup> 收稿日期: 2012 - 12 - 12

基金项目: 江苏省科技支持计划工业部分 (BE2011061)

作者简介: 马正华 (1962—), 男, 江苏昆山人, 教授。

是藻类)大量繁殖,湖水透明度下降,耗氧量增加,底层水氧消耗过程加剧。因此目前国内外富营养化的评价中,大多选择与上述富营养化过程关系密切的因子作为评价参数,通常包括物理的、化学的和生物的因子。

BP (Back Propagation) 网络是 1986 年由 Rumelhart 和 McClelland 为首的科学家小组提出,是一种按误差逆传播算法训练的多层前馈网络,是目前应用最广泛的神经网络模型之一。BP 网络能学习和存贮大量的输入输出模式映射关系,而无需事前揭示描述这种映射关系的数学方程。它的学习规则是使用最速下降法,通过反向传播来不断调整网络的权值和阈值,使网络的误差平方和最小<sup>[2]</sup>。BP 神经网络模型拓扑结构包括输入层 (input layer)、隐含层 (hide layer) 和输出层 (output layer)<sup>[3]</sup>。传统评价方法工作量大、主观性较强、需要复杂的预处理工作<sup>[4]</sup>。BP 神经网络具有自学习、自组织、较好的容错性和优良的非线性逼近能力,在实际应用中显示出较强的优势。BP 神经网络可进行高维非线性的精确映射,以在许多领域得到了广泛的应用,例如地震灾害人员伤亡预测模拟、年径流预报、人口预测等,并获得了理想的效果。

为此,通过 BP 神经网络为模型,以 1996 年到 2004 年太湖例行环境监测数据为基础,进行太湖水质时空变化预测以及研究,探讨水质变化的规律<sup>[5]</sup>,为太湖水环境的合理开发利用提供依据,为建立太湖水质时空变异模型及太湖流域管理规划平台提供决策支持。

## 1 BP 神经网络模型预测模型

### 1.1 数据的归一化处理

由于 BP 神经网络输入的 1 维代表 1 个特征,当神经网络的输入是多维时,要识别的模式有多个特征,当这多个特征的数据相差很大时,就需要归一化,变成相同数量级,以防某些数值低的特征被淹没<sup>[6-8]</sup>。除了为了方便数据的处理,另外为了保证程序运行时收敛加快,在网络训练前首先要对数据进行归一化处理,将数据限制在 (0.2, 0.8) 区间内。本例中的数据处理 (1) 及反预处理公式 (2) 如下:

$$y = 0.6 * \frac{x - x_{\min}}{x_{\max} - x_{\min}} + 0.2 \quad (1)$$

$$x = (y - 0.2) * \frac{x_{\max} - x_{\min}}{0.6} + x_{\min} \quad (2)$$

式中:  $y$  为归一化后输入或输出的值,  $x$  为输入或输出样本,  $x_{\max}$  为样本中变化范围内的最大值,  $x_{\min}$  为样本中变化范围内的最小值。

### 1.2 样本集选择及分配

本文选择 1997 年到 2004 年江苏省例行监测数据作为研究对象,将数据进行分组,共分为 6 组。第 1 组:取 1996~1998 年的数据作为输入,1999 年的数据作为对应的输出;第 2 组:取 1997~1999 年的数据作为输入,2000 年的数据作为对应的输出;第 3 组:取 1998~2000 年的数据作为输入,2001 年的数据作为对应的输出;第 4 组:取 1999~2001 年的数据作为输入,2002 年的数据作为对应的输出;第 5 组:取 2000~2002 年的数据作为输入,2003 年的数据作为对应的输出;第 6 组:取 2001~2003 年的数据作为输入,2004 年的数据作为对应的输出。

其中 5 组数据用于训练样本,1 组数据作为验证样本<sup>[9-11]</sup>。为了更好的验证 BP 神经网络,本文将每组数据都作为 1 次验证。训练好样本后,采用 2002~2004 年的数据作为输入,预测 2005 年的数据。然后采用 2003、2004 及预测到的 2005 年的数据作为输入,预测 2006 年的数据。以此类推,预测 2013 年的数据。

### 1.3 网络的建立

#### 1.3.1 确定网络层数

本文优先选用 3 层 BP 网络,因为 3 层的 BP 网络具有良好的泛化能力,能以任意精度逼近任意有理函数。若增加隐含层的层数,则会增加权值的训练时间,同时使网络复杂化<sup>[12]</sup>。本文中隐含层使用 sigmoid 单元、输出层使用线性单元。

#### 1.3.2 确定网络输入层节点数 $i$

根据每组输入样本的数目确定输入层的节点数  $i$ ,在本文中  $i$  为 3。

#### 1.3.3 确定那个隐含层节点数 $j$

本文采用 3 层的 BP 网络,隐含层节点数的选择参照下式:

$$n = \sqrt{a + b} + c \quad (3)$$

式中:  $n$  代表隐含层节点数目,  $a$  代表输入节点数目,  $b$  代表输出节点数目,  $c$  为 1~10 之间的常数。隐含层节点数过少,则无法产生足够的连接权组合数来满足若干样本的学习;隐含层节点数过多,则学习以后网络的泛化能力变差。本文通过神经网络

训练来确定隐含层的个数，首先根据经验公式确定隐含层中节点数目的范围，设计一个隐含层神经元数目可变的 BP 网络，通过误差对比，确定最佳的隐含层神经元的个数。由经验公式知隐含层节点数为 4~12 之间。根据本文设计的神经网络，试验结果表明在经过 2 023 次训练后，隐含层的神经元个数为 9 的 BP 网络对函数的逼近效果最好，因为它的误差最小。

1.3.4 确定输出层节点数  $k$

在湖体富营养化研究中，采用湖体富营养化指数来表示富营养化水平，本文采用湖体富营养化指数作为网络的输出，输出节点为 1 个。

1.3.5 确定转换函数

中间层采用正切激活函数，输出层采用线性函数。传递函数为 sigmoid 型函数，公式如下：

$$f(x) = \frac{1}{1 + \exp(-x)}$$
 (4)

1.3.6 设置网络的训练参数

net.trainParam.epochs—最大训练次数（缺省为 10）；net.trainParam.goal—训练要求精度（缺省为 0）；net.trainParam.lr—学习率（缺省为 0.01）；net.trainParam.max\_fail—最大失败次数（缺省为 5）；net.trainParam.min\_grad—最小梯度要求（缺省为 1e-10）；net.trainParam.show—显示训练迭代过程（NaN 表示不显示，缺省为 25）；net.trainParam.time—最大训练时间（缺省为 inf）；net.trainParam.mc—动量因子（缺省为 0.9）。

本文中采用的是动量梯度下降法，上述 net.trainParam.mc 是动量梯度下降法特有的训练参数。动量梯度函数采用了有动量的梯度下降法。

1.4 网络预测精度检验

首先采用 1996~1998 年的数据作为验证样本，其余为训练样本，然后采用 1997~1999 年作为验证样本，其余为训练样本，以此类推，训练网络并检验精度。对比结果如表 1 所示。

表 1 1996~2004 年预测值与真实值对比

年份	太湖湖体富营养化指数输出		相对误差/%
	期望	实际	
1999	59.0	59.074 3	0.12
2000	64.5	63.716 0	-1.22
2001	60.9	60.919 2	0.03
2002	59.3	59.356 7	0.09
2003	57.6	58.104 7	0.78

2004	62.7	62.916 1	0.34
------	------	----------	------

验证结果表明，训练样本平均相对误差为 0.43%。从预测结果看，BP 神经网络对太湖 1996~2002 年的富营养化历史样本拟合精度较高，预测结果的相对误差较小，能较好地预测太湖湖体富营养化变化趋势，此模型可以满足日常预测工作的需要。

2 数据来源及分析方法

2.1 网络预测精度检验数据来源

本文的研究数据为以上点位 1997~2004 年例行监测数据，监测单位为无锡、苏州和常州环境监测站，监测时间为每月的 1~10 日。湖体监测项目为水文、pH、溶解氧、高锰酸盐指数、化学需氧量、5 日生化需氧量、氨氮、总磷、总氮、铜、锌、氟化物、硒、砷、汞、镉、6 价铬、铅、氰化物、挥发酚、石油类、阴离子表面活性剂、硫化物、粪大肠菌群、透明度和叶绿素 a。

2.2 富营养化指数计算方法

本文根据江苏省环境监测中心的评价标准，结合太湖的实际情况，选择总磷（TP）、总氮（TN）、高锰酸盐指数（COD<sub>mn</sub>）、叶绿素 a（Chl-a）和透明度（Transp）5 项指标作为评价参数，采用综合营养状态指数法评价富营养化情况。综合营养状态指数计算公式（5）和营养状态指数计算公式（6）为：

$$T_{Li}(\Sigma) = \sum_{j=1}^m W_j \cdot T_{Li}(j)$$
 (5)

$$W_j = \frac{r_{ij}^2}{\sum_{j=1}^m r_{ij}^2}$$
 (6)

2.3 评价标准

本文的水质评价均以《地表水环境质量标准》（GB3838—2002）为标准。对综合营养状态指数进行分级，可得到湖区富营养状况，见表 2。营养状态指数值越高，营养程度越重。

表 2 湖库富营养化状况分级标准

Table 2 Grading standard of lake eutrophication status	
富营养化状况	综合营养状态指数
贫营养	$T_{Li} < 30$
中营养	$30 \leq T_{Li} \leq 50$
富营养	$T_{Li} > 50$
轻度富营养	$50 < T_{Li} \leq 60$
中度富营养	$60 < T_{Li} \leq 70$

重度富营养

$T_{LI}>70$

### 3 结果与分析

#### 3.1 网络预测

采用已经训练好的神经网络模型对太湖 2005 年湖体富营养化指数进行预测, 采用循环的方法, 预测 2005~2013 年太湖湖体富营养化指数。由 2002~2004 年的数据作为输入, 由网络计算输出 2005 年的太湖湖体富营养化指数, 以此类推。根据表 2 湖泊富营养化分级标准的 6 个等级, 将预测结果进行营养化等级分类, 太湖湖体 2005~2013 年的富营养化指数成果预测成果及营养状态分类如表 3 所示。

表 3 2005~2013 年太湖湖体富营养化指数预测成果  
Table 3 Prediction of Taihu lake's eutrophication indexes from the years 2005 to 2013

年份	太湖湖体富营养化指数	营养状态
2005	61.874 6	中度富营养
2006	58.100 9	轻度富营养
2007	58.415 4	轻度富营养
2008	61.868 6	中度富营养
2009	60.681 2	中度富营养
2010	58.645 4	轻度富营养
2011	60.860 0	中度富营养
2012	62.382 5	中度富营养
2013	63.589 9	中度富营养

#### 3.2 太湖富营养化预测结果评价

由历年的数据, 可以知道 1996~2004 年太湖湖体出现了 4 种富营养类型: 中营养、轻度富营养、中度富营养和重度富营养。将各种富营养类型汇总, 中度富营养和轻度富营养费别占了 55%和 42.5%, 重度富营养都出现在五里湖, 分别是 1997、2000、2001 年和 2002 年, 在 1999 年和 2003 年的东部沿岸区。从以上的分析结果可以看到, 太湖全湖的富营养化正在由轻度富营养向中度富营养过渡, 符合太湖湖体富营养化的发展趋势。特别的, 2007 年太湖爆发蓝藻, 富营养化水平从中度富营养向重度富营养转变, 全湖富营养化水体面积比例从 39%增加到 63%, 研究表明, 主要原因是污染源急剧增加, 导致“太湖病了”, 太湖的主要污染源是工矿企业、农业、围网养殖和城市污染。针对这种人为因素, BP 神经网络模型还有待进一步研究, 为水污染预警预报提供决策支持。

### 4 结 论

应用 BP 神经网络的非线性动力学特征及自学

习特性建立了太湖湖体富营养化的 3 层 BP 神经网络预测模型。利用 1996~2004 年的 6 组数据作为训练样本和验证样本, 训练的过程中调整、修正神经网络中各权值和阈值, 结果表明预测精度较高, 对湖泊水质预测的适应性较好。对太湖 2005 年的湖体富营养化指数进行了预测, 结果表明 2005 年没有较大的改善, 营养状态为中度富营养, 符合太湖污染情况发展态势。对 2006~2013 年的太湖湖体富营养化水平预测情况较符合太湖的发展趋势。在 2007 年太湖蓝藻暴发, 太湖的外部污染源主要有工业污染、农业面源污染、围网养殖和城市生活污染等。若考虑到人为因素和突发因素, 本文提出的 BP 神经网络模型有待改进。BP 神经网络适应性较好、预测精度较高, 能较好地反应水质指标的内在变化规律, 使得预测精度更高, 为控制水环境监测与决策支持平台提供科学依据。

#### 参考文献:

[1] 裴洪平, 罗妮娜, 蒋勇, 等. 利用 BP 神经网络方法预测西湖叶绿素 a 的浓度 [J]. 生态学报, 2004, 24 (2): 246 - 251.

[2] 刘成, 何可人, 周天彤, 等. 左右手运动想象脑电模式识别研究 [J]. 常州大学学报: 自然科学版, 2013, 25 (1): 25 - 30.

[3] 韦德志. 城市区域空气质量的 BP 神经网络预测研究 [D]. 武汉: 华中科技大学, 2009.

[4] 肖永辉, 王志刚, 刘曙照. 水体富营养化及蓝藻水华预警模型研究进展 [J]. 环球科学与技术, 2011, 34 (11): 152 - 158.

[5] 张亦含. 基于 GIS 的太湖水环境质量时空变异规律研究 [D]. 南京: 南京师范大学, 2005.

[6] 钟伟红, 马修水, 关宏伟, 等. 基于 RBF 神经网络的三坐标测量机动态测量误差预测 [J]. 中国科技论文, 2012, 7 (7): 560 - 562.

[7] 董长虹. Matlab 神经网络与应用 [M]. 北京: 国防工业出版社, 2005.

[8] 郭庆春, 何振芳, 李力, 等. BP 神经网络模型在太湖水污染指标预测中的应用 [J]. 计算机应用研究, 2011, 42 (10): 1303 - 1306.

[9] 李贺, 刘春光, 樊娟, 等. BP 神经网络在河流叶绿素 a 浓度预测中的应用 [J]. 中国给水排水, 2009, 25 (5): 75 - 79.

[10] Wu Q T, Yang X T, Hua B. Evaluation of network connection credibility based on neural network [J]. Journal of Computer, 2011, 6 (12): 2567 - 2573.

[11] Long C S, Wang X, Wu D H. A recognition model of hand odor based on BP artificial network [J]. Computer Science and Application, 2012, 2 (2): 57 - 60.

[12] Dragovic S, Onjia A. Artificial neural network data analysis for classification of soils based on their radionuclide content [J]. Russian Journal of Physical Chemistry A, 2007, 81 (9): 1477 - 1481.