

文章编号:1004-7271(2002)01-0031-06

# 长江口水质人工神经网络模型的建立及现状评价

杨红,李曰嵩

(上海水产大学海洋学院,上海 200090)

**摘要:**应用 B-P 人工神经网络方法建立了长江口的海域水质综合评价模型, B-P 人工神经网络结构中的输出层以单输出代替多输出以保证评价结果的唯一性。以此模型采用溶解氧、化学需氧量(COD)、磷酸盐、无机氮、总汞、镉、铅、砷、油类,对长江口 18 个监测点的海水水质进行综合评价。表明 2 个站点为 I 类水质, 10 个站点为 II 类水质, 5 个站点为 III 类水质, 1 个站点为 IV 类水质。

**关键词:**人工神经网络;水质评价;长江口

中图分类号: S912.8951.2 文献标识码: A

## Comprehensive assessment of water quality of Changjiang Estuary based on artificial neural networks

YANG Hong, LI Yue-song

(Ocean college, Shanghai Fisheries University, Shanghai 200090 China)

**Abstract:** Based on B-P artificial neural network, a model of comprehensive assessment of water quality of Changjiang Estuary is presented. Output-layer of B-P artificial neural network are substituted by multi-unit output for single-unit output. The result derived from the B-P artificial neural network is unique. Using dissolved oxygen, COD, phosphorus, inorganic nitrogen, Hg, Cd, Pb, As and oil, the model is applied to assess quality of the water taken from eighteen sampling site in Changjiang Estuary. It shows that the water at two stations is classified to grade I, ten stations to grade II, five stations attributed to III and one station attributed to grade IV.

**Key words:** artificial neural network; water quality assessment; Changjiang Estuary

水质评价是水资源开发利用的一项重要工作,评价方法是现代环境科学基础理论研究的重要课题。目前用于水质评价的模型较多。主要有综合指数法、灰色聚类法、灰色模式识别法、模糊综合评判法等。由于这些方法中,多数需要确定各评价指标对各级标准的隶属函数及各指标的权重,因此,评价结果受评价者主观因素的影响较大<sup>[1]</sup>。依据水质标准进行水质评价,实际上一个典型的模式识别问题,由于影响水质的因素很多,评价指标与水质标准等级之间是复杂的非线性关系,故至今仍没有一种统一的确定的评价方法<sup>[2]</sup>。人工神经网络(Artificial Neural Network, ANN)是 50~60 年代发展起来的适用性很强的交叉学科,人工神经网络力图模拟人脑的一些基本特征,如自适应性、自组织性和容错性能,已用于模式识别、系统辨识等领域,取得了很好的效果<sup>[3]</sup>。本文试利用误差反向传播 B-P 算法(Back propagation)的人工神经网络方法建立长江口水质评价模型并以此对长江口海区的水质进行现状综合评价。

## 1 B-P 人工神经网络

B-P 算法的人工神经网络,是一种由非线性变换单元组成的前馈型网络。它由一个输入层、一个输出层和一个或若干个隐藏层组成,每一层尽由若干神经元组成。输入层接收外界信息,输出层对输入层的信息进行判别和决策;中间的多个隐层用来表示或存储信息。

B-P 网络中的信息是单向传递的,同一层神经元之间无不相互连接,而层与层之间的神经元采用全连接方式,其连接程度用权值表示,并通过监督学习调整其权值。每层神经网络的状态都将影响下一层的神经元状态,并且每个神经元状态都对应着一个作用函数  $f(x)$ ,从输入层隐层和隐层至输出层之间的作用函数常用非线性的 Sigmoid 函数<sup>[4]</sup>为  $f(x) = \frac{1}{1 + e^{-x}}$  (一般用法)。

B-P 网络的学习过程由正向信息传播和反向误差传播两阶段组成。第一个阶段从输入层向输出层信息传播状态计算,如果网络的结构和初始权值已设定,信息由输入层神经元传递到隐层神经元,再经隐层神经元处理后送到输出层神经元,由输出层神经元处理后产生一个输出模式,这个过程称为正向传播。如果实际输出与给定的期望输出模式有较大的误差而不满足要求,即转到第二阶段,第二阶段是从输出层向输入层进行信息传播误差修正的计算,即将误差沿原来的连接通路由输出层→隐层→输出层返回,通过修改各层神经元的阈值和权值使误差减小,然后再转向正向传播过程,两个过程反复交替,且当网络的误差小于给定的临界值学习结束,此时可得到具有确定的权值和域值的收敛的网络模型,利用这个收敛的网络可以完成实际模式的识别<sup>[5]</sup>。

## 2 水质 B-P 模型的建立

### 2.1 学习样本的产生和神经网络的训练

本文参考 GB3097—1997 海水水质标准,将长江口水质评价标准分为 5 级,具体评价标准见表 1。

表 1 长江口水质评价标准

| Tab.1 Changjiang estuary water quality evaluating standards |         |        |        |        | mg/L   |
|---|---------|--------|--------|--------|--------|
| 水质参数  | I 类     | II 类   | III 类  | IV 类   | 超 IV 类 |
| 溶解氧   | 6       | 5      | 4      | 3      | 2      |
| 化学需氧量   | 2       | 3      | 4      | 5      | 6      |
| 磷酸盐   | 0.015   | 0.030  | 0.030  | 0.045  | 0.060  |
| 无机氮   | 0.2     | 0.3    | 0.4    | 0.5    | 2.5    |
| 总汞  | 0.00005 | 0.0002 | 0.0002 | 0.0005 | 0.001  |
| 镉   | 0.001   | 0.005  | 0.010  | 0.010  | 0.015  |
| 铅   | 0.001   | 0.005  | 0.01   | 0.05   | 0.10   |
| 砷   | 0.02    | 0.03   | 0.05   | 0.05   | 0.10   |
| 油类  | 0.050   | 0.050  | 0.30   | 0.50   | 1.00   |

利用 VB 编程在上述水质评价标准范围内随机产生 300 个学习样本(其中各类水质均为 60 个)。然后将这 300 个样本输入到 STATISTICA Neural Networks(美国 Statsoft 软件公司出品)中<sup>[6]</sup>,选用三层神经网络结构。选取实测的水质参数溶解氧、化学需氧量、磷酸盐、总汞、无机氮、镉、砷、油类等 9 种水质参数作为输入层神经元。隐藏层神经元数的选取采用从少到多原则,即先从 2 个开始训练,然后 3 个、4 个,依次类推,到大约 8 个,每个不同的神经元数都训练十组,并且比较训练效果,取最满意的一组,该组的神经元数即为本模型的节点数。本例隐藏层神经元数取 5 个。输出层神经元数一般可选取 5 个神经元表示五种水质类别,评价时选输出值最大者所在的输出单元作为综合判别类型,但是,如果 5 个输出值其中两个或三个相近或相同时,则很难判断究竟属于哪一类。为克服这一缺点,本文将水质类别用数值表示,即 1 2 3 4 5,输出层仅有一个神经元<sup>[3]</sup>,通过临界值样本来产生水质等级范围,待评水质参数

输入到训练好神经网络,输出的值落在那个范围就属于那个范围的水质等级。程序将自动大致按 2 : 1 : 1 的关系随机产生训练样本( 155 个 ) 检验样本( 71 个 ) 测试样本( 74 个 )。选用 B-P 算法对样本进行训练,学习率和冲量系数采用系统缺省值 0.1 和 0.3。当达到所给的验证误差时,程序会自动停止,网络训练完成。选用的网络结构如图 1 所示。

随着训练次数的增大,误差逐步减小且趋于收敛。在本网络训练 8000 次以后误差减小幅度已很小。这种现象表明,随着训练次数的增大,模型本身的有效性逐渐提高,便其实用性可能因此而降低,即出现评定精度较的模型导致较差的预测。这是一种“过渡训练”现象,实际预测中需加以注意。网络的训练样本、检验样本和测试样本的衰退统计如表 2 所示。

### 2.2 模型灵敏度分析

应用 STATISTICA Neural Networks 软件包中自带的灵敏度分析功能,可以得到各水质参数对水质等级评价的重要性<sup>[6]</sup>结果见表 3。

从表中可以看出,对本网络而言,对水质的等级评判影响最大的是无机氮,其次是溶解氧、化学需氧量( COD )等,对水质评价最不起作用的是磷酸盐和总汞。所以在实测水质中无机氮、溶解氧和 COD 的资料会直接影响到水质的等级。

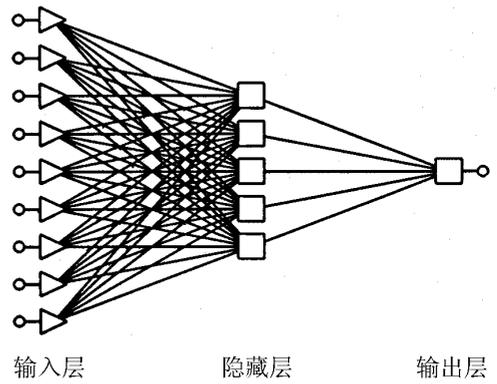


图 1 长江口水质评价的人工神经网络结构图  
Fig.1 The structure of artificial neural network for water quality assessment of Changjiang Estuary

表 2 训练样本、检验样本和测试样本在网络中的衰退统计

Tab.2 Regression statistics of training sample, cation and testing sample

|         | 训练样本      | 检验样本      | 测试样本      |
|---------|-----------|-----------|-----------|
| 数据平均值   | 3.083871  | 2.746479  | 3.067568  |
| 数据方差    | 1.427717  | 1.421451  | 1.378217  |
| 误差平均值   | 0.007266  | 0.004211  | 0.0002736 |
| 误差方差    | 0.08967   | 0.09613   | 0.08423   |
| 绝对误差平均值 | 0.06523   | 0.07211   | 0.06132   |
| 方差变异系数  | 0.06281   | 0.06763   | 0.006112  |
| 相关系数    | 0.9980614 | 0.9977199 | 0.9981327 |

表 3 各水质参数灵敏度分析表

Tab.3 Sensitivity analysis of each parameter

| 参数    | 溶解氧    | 化学需氧量  | 磷酸盐    | 无机氮    | 总汞     | 镉      | 铅      | 砷      | 油类     |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 重要度排序 | 2      | 3      | 8      | 1      | 9      | 7      | 6      | 5      | 4      |
| 舍去后误差 | 0.2029 | 0.1813 | 0.1133 | 0.3157 | 0.1040 | 0.1428 | 0.1463 | 0.1636 | 0.1799 |

### 2.3 临界值的选取

网络训练好以后,将实测值输入到训练好的网络中去,其输出层的输出值即表现相应的评判结果,在大多数文献中采用对输出值四舍五入的方法来判断等级<sup>[1]</sup>,如输出值为 2.8367,经四舍五入取整数为 3,表明此类水质为Ⅲ类。但是用四舍五入的方法来判断水质类型有时会很牵强,所以我们要知道临界值的输出值,作为划分各类水质依据。由表 1 可知,表中所列的各类水质的指标值是区别水质的临界值,因此将表 1 中的水质评价标准作为临界值输入训练好的神经网络,得到结果为表 4 所示。

所以 I ~ V 类水质的输出值范围分别为 0 ~ 1.561, 1.561 ~ 2.466, 2.466 ~ 3.604, 3.604 ~ 4.416 和大于 4.416。

表 4 临界值的输出结果

Tab.4 The output results of critical values

| 级别       | 溶解氧 | COD | 磷酸盐   | 无机氮 | 总汞     | 镉     | 铅     | 砷     | 油类    | 输出值   |
|----------|-----|-----|-------|-----|--------|-------|-------|-------|-------|-------|
| I ~ II   | 6   | 2   | 0.015 | 0.2 | 0.0005 | 0.001 | 0.02  | 0.050 | 1.561 |       |
| II ~ III | 5   | 3   | 0.030 | 0.3 | 0.0002 | 0.005 | 0.005 | 0.03  | 0.050 | 2.466 |
| III ~ IV | 4   | 4   | 0.030 | 0.4 | 0.0002 | 0.010 | 0.01  | 0.05  | 0.30  | 3.604 |
| IV ~ 超IV | 3   | 5   | 0.045 | 0.5 | 0.0005 | 0.010 | 0.05  | 0.05  | 0.05  | 4.416 |

### 3 长江口水质现状评价

#### 3.1 应用人工神经网络进行评价

本文应用 2000 年 8 月份国家海洋局东海分局的调查资料,水质监测项目包括:溶解氧、化学需氧量、磷酸盐、无机氮、总汞、镉、铅、砷、油类,长江口附近的 18 个站点分布如图 2 所示。将各站点的实测数值输入训练好的神经网络,输出的结果和评判类型见表 5。

表 5 长江口水质实际输出值和评价结果

Tab.5 The real output values and evaluated results of Changjiang Estuary at various stations

| 站位   | 输出值      | B-P 评判结果 |
|------|----------|----------|
| ST01 | 2.357944 | II 类     |
| ST02 | 2.804381 | III 类    |
| ST03 | 2.168185 | II 类     |
| ST04 | 1.722373 | II 类     |
| ST05 | 2.165779 | II 类     |
| ST06 | 1.082985 | I 类      |
| Z078 | 1.778288 | II 类     |
| Z079 | 2.135264 | II 类     |
| Z081 | 2.324248 | II 类     |
| Z083 | 2.176397 | II 类     |
| Z084 | 1.989264 | II 类     |
| Z085 | 2.882274 | III 类    |
| Z087 | 3.689239 | IV 类     |
| Z088 | 2.895115 | III 类    |
| Z089 | 1.98597  | II 类     |
| Z090 | 3.455519 | III 类    |
| Z091 | 2.947882 | III 类    |
| Z092 | 1.429545 | I 类      |

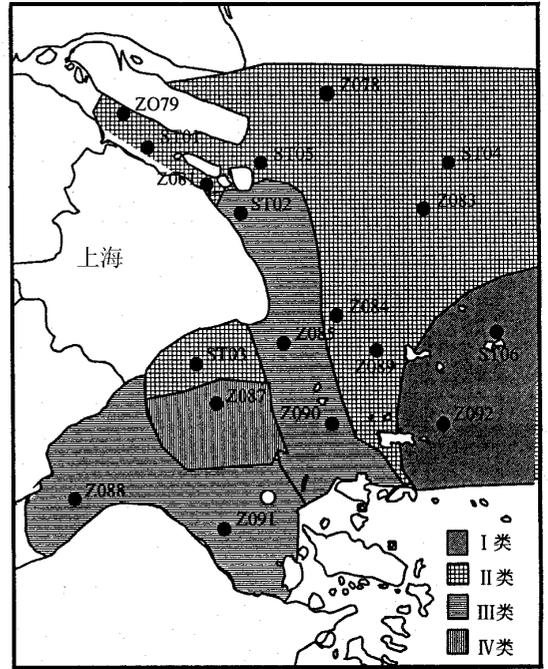


图 2 长江口水质站位分布图

Fig.2 The distribution of sampling site in Changjiang Estuary

#### 3.2 应用单项水质参数评价

在单项水质参数评价中,一般采用标准型指数单元评价<sup>[7]</sup>,其方法为  $X_i = \frac{C_i}{S_i}$

式中  $X_i$  —— i 点的标准型指数单元;

$C_i$  —— 某质量参数的监测统计浓度(或强度)

$S_i$  —— 某一质量参数的评价标准,通常采用国家环境质量标准。

由于溶解氧(DO)与其他水质参数的性质不同,需要用不同的指数单元形式,DO 的标准型指数单元为:

$$X_{DO_j} = \frac{|DO_f - DO_j|}{DO_f - DO_s} \quad (DO_j \geq DO_s), \quad X_{DO_j} = 10 - 9 \frac{DO_j}{DO_s} \quad (DO_j < DO_s), \quad DO_f = \frac{468}{31.6 + T}$$

式中  $X_{DO_j}$  —— j 点的 DO 标准指数单元；

$DO_f$  —— 饱和 DO 浓度；

T —— 水温(℃)

$DO_j$  —— j 点的 DO 浓度；

$DO_s$  —— DO 的评价标准。

长江口的单项水质评价结果见表 6。

表 6 长江口的单项水质评价结果

Tab.6 The evaluating results of single water parameter in Changjiang Estuary

| 水质项目   | 溶解氧 | COD | 磷酸盐  | 无机氮   | 总汞    | 镉     | 铅 | 砷     | 油类 |
|--------|-----|-----|------|-------|-------|-------|---|-------|----|
| 站数超标   | 超Ⅰ类 | 0   | 5.56 | 88.89 | 94.44 | 11.11 | 0 | 94.44 | 0  |
| 百分率(%) | 超Ⅱ类 | 0   | 0    | 33.33 | 94.44 | 0     | 0 | 5.56  | 0  |
|        | 超Ⅲ类 | 0   | 0    | 33.33 | 94.44 | 0     | 0 | 0     | 0  |
|        | 超Ⅳ类 | 0   | 0    | 16.67 | 88.89 | 0     | 0 | 0     | 0  |

无机氮浓度的范围在 0.1827 ~ 2.084mg/L 之间,其中硝酸盐浓度范围为 0.17 ~ 2.066mg/L,亚硝酸盐为 0.0014 ~ 0.0295mg/L,铵盐为 0.007 ~ 0.071mg/L,可以看出在该海区中无机氮主要以亚硝酸盐形式存在,硝酸盐是氮在浅海水中最稳定形式,具有不被悬浮颗粒物吸附或包裹的保守行为,主要来源于生命代谢物质分解、氧化以及陆源径流的输入<sup>[7]</sup>。由图 3(a)可知,无机氮是呈现由西向东逐渐降低的趋势,并且在上海南海区(杭州湾口)出现峰值。由图 3(b)可知磷酸盐浓度在 0.0032 ~ 0.0565mg/L 之间,平均值为 0.0311mg/L 的分布也是从沿岸向外海逐渐降低的,越靠近杭州湾南岸其值越大。这主要是由于沿岸的排污和径流中含有大量的无机氮和磷酸盐导致的。由图 3(c)COD 的分布是从南向北逐渐增加的,图 3(d)海水中的铅主要来自大气的沉降,汽车排出的尾气是铅的主要来源。

### 3.3 评价结果分析

从表 5 中可看出,经人工神经网络评价的长江口附近海域水质大致维持在Ⅱ类水平,一些离上海市较远的监测点(Z092、ST06)可达到Ⅰ类水质标准,但一些离上海市较近的,尤其是靠近排污口的监测点的水质达到Ⅲ类或Ⅳ类水质标准,这些海域应当引起有关部门的注意。

从表 6 的单项水质评价结果来看,长江口的无机氮、磷酸盐及铅的污染程度相对比较严重,18 个站中有 16 个站位的无机氮超过Ⅳ类水质标准,超标率为 88.89%,平均含量达到 1.1340mg/L 最大测出值为 1.9996mg/L 已经是四类水质的 3.9992 倍。长江口的活性磷酸盐超标现象也比较严重,其平均浓度为 0.0311mg/L,已超过三类水质标准,部门海区的活性磷酸盐已经超过四类海水水质标准。长江口的 18 个监测站的溶解氧、镉、砷、油类和 COD 无明显超标现象,只有极个别海区超过国家一类海水标准。

## 4 结论

(1) 用人工神经网络进行水质评价,在水质评价标准基础上生成学习样本,以相应的水质等级为基础生成期望输出。应用 STATISTICA Neural Networks 软件包对网络进行学习训练,当网络收敛后即可用来对所需评价的水质进行水质评价。

(2) 使用人工神经网络可以很方便地对输入参数进和灵敏度分析,了解哪些参数对网络的影响最大,这对数据不全的区域进行水质评价尤为重要。

(3) 将各类水质参数标准临界值输入训练好的神经网络,其输出结果作为划分水质等级的临界值,可使评价结果更准确。

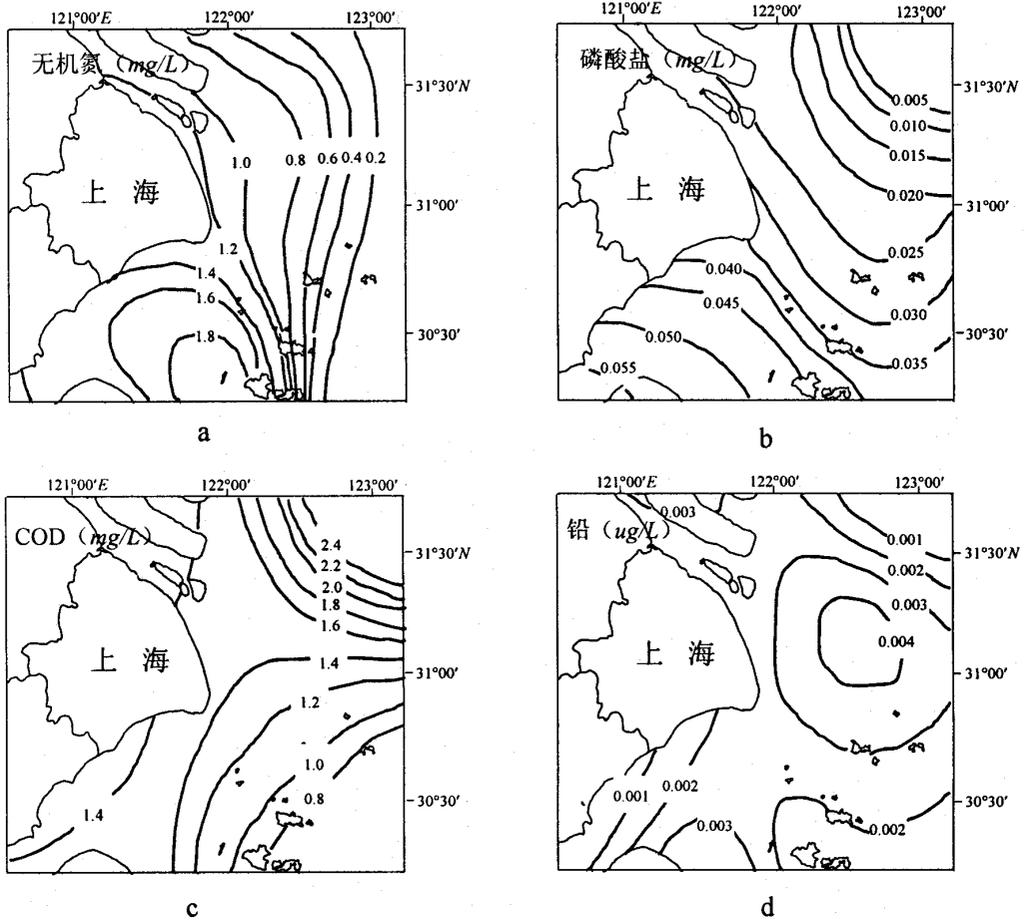


图3 无机氮、磷酸盐、COD、铅平面分布图

Fig.3 the distribution of inorganic nitrogen phosphate COD and *plumbum*

(4) 用一个输出神经元的 B-P 网络进行水质综合评价,可保证评价(输出)结果的唯一性。

(5) 中间层节点数的选取是一个十分复杂的问题,没有很好的解析式来表达,它的数目对网络的训练次数及收敛速度影响较大。通过对不同中间节点数的训练比较(分别选取 2~8),发现节点数较少时(4~6个),网络收敛快,训练次数少(6000~8000次),但精度一般。但节点过多(7~8个),虽精度提高,却又使得训练次数明显增多(10000~12000次),收敛放慢。这是因为当要求逼近精度高时<sup>[8]</sup>,要求逼近三角函数或多项式的项数要增加,因而中间节点数也高。

(6) 利用人工神经网络模型对长江口附近海域水质综合评价,其水质大致维持在 II 类水质标准,水质状况还是比较乐观的。但从单项水质评价结果来看,无机氮、磷酸盐超标比较严重。

#### 参考文献:

- [1] 刘国东,黄川友,丁晶.水质综合评价的人工神经网络模型[J].中国环境科学,1998,18(6):514-517.
- [2] 蔡煜东,汪列,姚林声.水质富营养化的人工神经网络决策模型[J].中国环境科学,1995,15(2):123-127.
- [3] 张立明.人工神经网络的模型及其应用[M].上海:复旦大学出版社,1-2.
- [4] 纪桂霞,李培红.水环境质量评价的人工神经网络模型及应用[J].华北水利水电学院学报,1999,20(1):60-62.
- [5] 李最正.水质综合评价的 B-P 网络模型[J].甘肃环境与监测,1998,11(2):24-27.
- [6] Statsoft. statistica neural networks[Z].http://www.statsoft.com.
- [7] 施晓来,叶新荣.三门湾蛇蟠海域环境质量现状评价[J].东海海洋,2000,18(4):16-21.
- [8] 胡瑞敏.B-P 网络的全集成运行开发环境[J].微计算机应用,1996,17(2):1-5.