

文章编号: 1004 - 7271(2002)04 - 0329 - 06

# 渔业水域污染死鱼的 ANN 识别模型的建立

杨 红, 李曰嵩

(上海水产大学海洋学院, 上海 200090)

**摘 要:**本文应用人工神经网络(ANN)方法,对渔业水域污染死鱼特征进行提取,建立了导致污染死鱼的毒物类型的识别模型,其人工神经网络结构选输入层 18 个神经元,中间层 9 个神经元,输出层为 1 个神经元,并对网络参数的优化进行讨论。结果显示通过运用已训练好的渔业水域污染死鱼的 B-P 人工神经网络识别模型,只要通过简单的加法和乘法运算,就可以对任何一例污染死鱼的实际样本进行死鱼原因的判断,该方法具有较好的适用性和较高的实用性。

**关键词:**人工神经网络; B-P 算法; 渔业水域; 污染

**中图分类号:**S912; TP183      **文献标识码:** A

## Identification of the respective causes of toxic deaths in piscatorial waters with model ANN

YANG Hong, LI Yue-song

(Ocean College, Shanghai Fishery University, Shanghai 200090, China)

**Abstract:** Based on artificial neural network (ANN) and extracted characters of dead fish in polluted piscatorial waters, an identifiable model of toxicant type bringing on dead fish is presented in the paper. There are eighteen nerve cells in input-layer, nine nerve cells in middle-layer and one nerve cell in output-layer of the artificial neural network. Network parameters were optimized. Results show that any pollution to deaths can judge death causes by simple addition and multiplication operations using trained B-P artificial neural network, so this method has high applicability and practicability.

**Key words:** artificial neural network; B-P arithmetic; piscatorial waters; pollution

由于未经处理工业废水的排放、船舶航运的溢油、漏油污染,使渔业水域受到不同程度污染,有的直接毒死水生生物,有的影响其生长繁殖并导致品种减少,此外,养殖业的快速发展和高密度养殖也引起了一系列生态负效应,加上农业废水和生活污水的排放,使渔业水域富营养化加剧,海域内湾和沿岸赤潮频频发生,养殖动物病害增多。对由于污染造成的死鱼现象如何识别其污染原因,追踪污染源是进行渔业环境管理的重要环节,具有很强的专业性要求。一般完整的死鱼原因确定的诊断程序需要进行四部分的工作,一是化学物中毒致鱼死亡的确定;第二部分是一种或几种可疑毒物的确定;第三部分是毒物来源调查;第四部分是实验室分析和毒性再现试验。本文试通过对造成死鱼的物理、化学、生物的影响因子分析,辨别污染造成死鱼的类型,结合实际情况,筛选主要影响因子,通过建立 ANN 模型识别死

收稿日期: 2002-06-05

基金项目: 上海水产大学校长基金项目(SFU200003)

作者简介: 杨 红(1962-),女,江苏无锡人,副教授,主要从事海洋环境评价与保护研究, E-mail: hyang@shfu.edu.cn

鱼类型,辅助死鱼原因的快速诊断,为渔业行政管理部门及时处理污染死鱼事件提供科学依据。

## 1 B-P算法的ANN原理

“人工神经网络”(Artificial Neural Network,简称ANN)是在对人脑组织结构和运行机智的认识理解基础之上模拟其结构和智能行为的一种工程系统<sup>[1]</sup>。人工神经网络同现行的计算机不同,是一种非线性的处理单元。只有当神经元对所有的输入信号的综合处理结果超过某一门限值后才输出一个信号。因此神经网络是一种具有高度非线性的超大规模连续时间动力学系统<sup>[2]</sup>。其基本原理为其含有输入层、输出层以及处于输入输出层之间的中间层。中间层有单层或多层,由于它们和外界没有直接的联系,故也称为隐层。隐层虽然和外界不连接,但它们的状态则影响输入输出之间的关系。即改变隐层的权系数,可以改变整个多层神经网络的性能。相邻两层节点之间单向互联,方向由输入层到输出层<sup>[3]</sup>。各神经元节点的输出为  $y_i = f(\sum_j w_{i,j} \times x_j + \theta_i)$ ; 式中:  $y_i$  为第  $i$  个神经节点的输出;  $w_{i,j}$  为上一层  $j$  节点对下一层  $i$  节点的权值;  $x_j$  为输入信号的第  $j$  个分量;  $\theta_i$  为阈值。激发函数一般取:  $f(x) = 1/(1 + e^{-x})$ 。本文通过给定学习样本进行学习训练建模方式,采用误差反向传播的算法(Error propvgantion)反向传播算法分二步进行,即正向传播和反向传播<sup>[4]</sup>。学习结束即可确定网络结构,当进行评价时,只需把数值输入网络进行正向学习,则输出结果中最大的节点就代表了模式的类型。这两个过程的工作简述如下。

### (1) 样本正向传播

输入的样本从输入层经过隐单元一层一层进行处理,通过所有的隐层之后,则传向输出层;在逐层处理的过程中,每一层神经元的状态只对下一层神经元的状态产生影响。在输出层把现行输出和期望输出进行比较,如果现行输出不等于期望输出,则进入反向传播过程。

### (2) 误差反向传播

反向传播时,把误差信号按原来正向传播的通路反向传回,并对每个隐层的各个神经元的权系数进行修改,以望误差信号趋向最小。

## 2 建立污染死鱼类型判定的B-P模型

### 2.1 特征值的提取

根据《淡水鱼类急性中毒死亡诊断方法》<sup>[5]</sup>中列出的18个诊断指标的特征ABC分别进行量化,提取特征值。

对各种死鱼原因造成的鱼的形态、行为反应和环境特征的异同分析,选择以下十八项指标作为区别不同原因死鱼的诊断指标,诊断指标还具有ABC不同的特征值,其含义为:

- A. 易观察测定,即可通过肉眼观察和使用简单的仪器快速测定。
- B. 代表性强,指标可以反应各种原因死鱼的生物和环境特征。
- C. 可以比较。不同原因造成的死鱼,其指标和特征有所差异,可以比较。

表1列出了18个诊断指标及指标的特征,并加以量化。

表2表示不同原因死鱼的诊断指标及其特征,多特征指标取平均值,即  $AB = A \cup B$  (即  $0 \sim 0.7$ ),  $AC = A \cup C$  (即  $0 \sim 0.3$  或  $0.7 \sim 1.0$ ),  $ABC = A \cup B \cup C$  (即  $0 \sim 1.0$ ),其它类似。

### 2.2 建模过程

#### 2.2.1 样本的产生

根据18个诊断指标和四种死鱼原因,建立有18个输入因子、一个输出因子,再由计算机依据已有的四个学习样本再随机产生共200个学习样本的网络模型。样本的选取采用在EXCEL中根据量化结果取随机数。并设输出期望值为:化学物中毒=1、鱼病死鱼=2、缺氧死鱼=3、毒藻死鱼=4,导入到SNN软件中。

表1 急性中毒死鱼诊断指标及指标的特征  
Tab.1 the diagnostic standard and standard character of toxic deaths

序号	诊断指标	指标特征及特征值范围		
		A	B	C
1	死亡速率	突发性强,短期内大批死亡 (2h内死鱼70%尾) 0~0.3	大批死亡之前有一个明显的 少量死亡前期 0.3~0.7	逐渐死亡 0.7~1.0
2	品种选择	品种选择性不明显,死亡的 品种相对较多(多于3种) 0~0.3	耐毒(a)和耐低氧(b)的 品种可能存活 0.3~0.5~0.7	有一定的品种选择性,死亡的 品种相对较少(1~2种) 0.7~1.0
3	死鱼时间	严重的死鱼发生在光照强的白天, 其余时间不发生大批死鱼(6~18时) 0~0.3	严重的死鱼通常发生在午夜过后 至黎明,其余时间不发生大批死鱼 (0~6时,18~24时) 0.3~0.7	
4	死鱼季节		春末、秋初,夏季高温季节,尤其 梅雨气压低时易发生大批死鱼 (5~9月)0.3~0.7	
5	个体大小 选择	小鱼比大鱼先死, 小鱼死亡率比大鱼高 0~0.3	大鱼比小鱼先死, 大鱼死亡率比小鱼高 0.3~0.71	
6	行为反应	初期反应为冲撞、急速游动 0~0.3	在表层吞咽空气,沿池边游动 0.3~0.7	游动缓慢乏力 0.7~1.0
7	形态特征	眼球突出、鳃盖鲜红、 脊柱弯曲等中毒症状 0~0.3	胸鳍向前伸展、 鳍条发白 0.3~0.7	体表及内脏充血溃烂、 体表胞囊等鱼病症状 0.7~1.0
8	鱼体附着物	体表尤其鳃部有附着物 且通常有毒物特异气味 0~0.3		
9	浮游动物 状况	大量死亡,基本消失 0~0.3	品种、数量减少 0.3~0.7	基本正常(80%以上,下同) 0.7~1.0
10	浮游植物 状况	品种数量减少, 1,2种有毒藻类品种占优势 0~0.3	存在大量死藻和濒死的藻细胞 0.3~0.7	基本正常 0.7~1.0
11	其它水生物 状况	水生植物可能变色或死亡, 龟、蛇、蛙、螺等死亡或濒死 0~0.3		基本正常 0.7~1.0
12	病原体			目镜或镜检可在体表、鳃部、内脏观 察到大量细菌、霉菌、寄生虫 0.7~1.0
13	死鱼发生 形式	孤立发生 0~0.3		可能流行发生 0.7~1.0
14	溶解氧	中午强光照射时饱和或过饱和 (9~12mg/L) 0~0.3	低于渔业水质标准 (低于3mg/L) 0.3~0.7	基本正常 (3~9mg/L) 0.7~1.0
15	酸碱度	水体偏碱性中午 强光照射时 pH>9 0~0.3	死鱼发生前几天或 一周强光照射时 pH>9 0.3~0.7	基本正常 0.7~1.0
16	水体气味	毒物特有异味 0~0.3	酸白菜味、霉味、臭味 0.3~0.7	基本正常 0.7~1.0
17	水色	铜绿、黄褐、红棕 0~0.3	灰白、黑色 0.3~0.7	基本正常 0.7~1.0
18	急性致死 试验	中毒或死亡 0~0.3	基本正常 0.3~0.7	短时间内不发生死鱼 0.7~1.0

表2 不同原因死鱼的诊断指标及其特征

Tab.2 The diagnostic standard and character of different deaths reason

序号	诊断指标	死鱼原因			
		化学物中毒	鱼病死鱼	缺氧死鱼	毒藻死鱼
1	死亡速率	1A	1A1B1C	1A1B	
2	品种选择	2A2B(a)	2C	2A2B(b)	2A2B(a)
3	死鱼时间			3B	3A
4	死鱼季节			4B	
5	个体大小选择	5A		5B	5A
6	行为反应	6A	6A6C	6B	
7	形态特点	7A	7C	7B	
8	鱼体附着物	8A			
9	浮游动物状况	9A9B	9C	9B	9A9B
10	浮游植物状况	10B	10C	10B	10A
11	其它水生生物状况	11A	11C	11C	11A
12	病原体		12C		
13	死鱼发生形式	13A	13C		
14	水体溶解氧		14C	14B	14A
15	水体酸碱度		15C	15B	15A
16	水体气味	16A	16C	16B	16B
17	水色		17C	17B	17A
18	急性致死试验	18A	18C	18B	18A

### 2.2.2 选择检验样本

在 200 例样本中随机选取 100 例为训练样本, 25 例为检验样本, 25 例为测试样本。之所以必须选择检验样本, 是因为检验样本的选取能即时对网络训练进行检验。检验样本的数目一般至少为总样本的 10%。

### 2.2.3 选取网络参数

应用 Statistic 软件训练死鱼类型判别模型, 网络结构见图 1, 即 18 个输入因子, 中间层 9 个节点, 1 个输出因子。中间层节点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 12 个, 每个不同的节点数都训练十组, 并且比较训练效果, 取最满意的一组, 该组的节点数即为本模型的节点数。本例取 9 个中间层节点。训练步长间隔为 20, 循环 1000 次, 学习率 0.1, 冲量 0.3, 样本随机并要求验证。设训练停止条件为训练样本与检验样本均方根误差比较接近(相差小于 0.001), 且相邻两次均方根误差比较接近(相差小于 0.001)。

### 2.2.4 训练模型

先记录下初始权值, 然后开始训练, 当满足停止条件时停止训练, 记录下此时网络各参数。再初始化网络, 重新训练, 这样重复 10 次, 取结果最满意的一组。本例采用了网络推荐的 9 个节点, 训练 6000 次后, 样本均方根误差已精确到 0.001。

### 2.2.5 误差分析

随着训练次数的增大, 误差逐步减小且趋于收敛。在 6000 次以后误差减小幅度已很小。这种现象

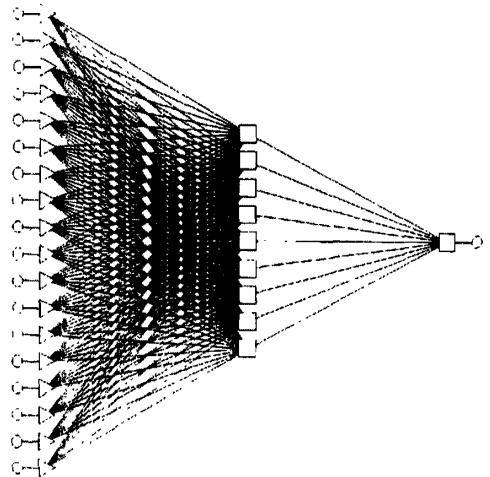


图1 所选的人工神经网络结构图

Fig.1 The structure of artificial neural network

表明,随着训练次数的增大,模型本身的有效性逐渐提高,但其实用性可能因此而降低,即出现评定精度较高的模型导致较差的预测。这是一种“过度训练”现象,实际预测中需加以注意(见表 3)。

表 3 训练样本、检验样本和测试样本在网络中的衰退统计

Tab.3 Regression statistics of training sample, cation and testing sample

	训练样本	检验样本	测试样本
Data Mean(数据平均值)	2.493333	2.6	2.44
Data S.D.(数据方差)	1.127727	1.154701	1.083205
Error Mean(误差平均值)	0.0009341	0.0008436	0.0271
Error S.D.(误差方差)	0.03374	0.05963	0.071862
Abs E. Mean(绝对误差平均值)	0.02619	0.04059	0.0564
S.D. Ratio(方差变异系数)	0.02992	0.05164	0.06634
Correlation(相关系数)	0.999564	0.998667	0.9977977

以上数据误差指标精度均达到 0.01,符合初始设定,网络训练已比较成熟。

### 2.2.6 实例验证

将具体死鱼的表现症状,按照表 4 进行量化。

表 4 死鱼表现症状的量化

Tab.1 numeral transformation of dead symptom

诊断指标	待测样本 1	待测样本 2	待测样本 3	待测样本 4	待测样本 5
死亡速率	0.01		0.11	0.34	0.93
品种选择	0.78	0.11	0.50	0.68	0.86
死鱼时间		0.27		0.39	
死鱼季节				0.54	
个体大小选择		0.10	0.07	0.51	
行为反应	0.76		0.16	0.56	0.03
形态特点	0.93		0.08	0.58	0.89
鱼体附着物			0.02		
浮游动物状况	0.91	0.45	0.70	0.52	0.81
浮游植物状况	0.83	0.07	0.67	0.68	0.94
其它水生物状况	0.87	0.16	0.13	0.76	0.76
病原体	0.81				0.71
死鱼发生形式	0.75		0.10		0.94
水体溶解氧	0.90	0.10		0.32	0.79
水体酸碱度	0.99	0.05		0.43	0.98
水体气味	0.99	0.89	0.22	0.38	0.83
水色	0.86	0.20		0.55	0.94
急性致死试验	0.84	0.19	0.23	0.55	0.99
输出值	1.970123	4.032144	0.9605925	2.958769	1.968577
评判结果	鱼病死鱼	毒藻死鱼	化学物中毒	缺氧死鱼	鱼病死鱼

### 2.2.7 灵敏度分析

此训练模型的灵敏度分析见表 5。

从灵敏度数据中可看出,训练样本与检验样本的排位有 12 个指标相同,5 个指标相差一位次序,一个指标相差两个次序,模型可信度较高。18 个指标中死亡速率排位第一,说明它对模型判别死鱼原因影响最大。18 个指标对模型灵敏度的影响排序依次为:死亡速率与病原体、行为反应、形态特征、死鱼发生形式、水色、酸碱度与死鱼季节、死鱼时间、水体气味和急性致死试验及溶解氧、鱼体附着物、个体大小选择、浮游动物状况、其它水生物状况、品种选择和浮游植物状况。

表 5 训练模型参数的灵敏度分析  
Tab.5 Sensitivity analysis of training model parameter

诊断指标	训练样本			检验样本		
	重要度排序	舍去该参数模型产生的误差	变异系数	重要度排序	舍去该参数模型产生的误差	变异系数
死亡速率	1	0.432	54.45	1	0.479	53.27
品种选择	17	0.008	1.02	17	0.009	0.98
死鱼时间	9	0.019	2.37	9	0.018	1.97
死鱼季节	8	0.030	3.74	7	0.033	3.67
个体大小选择	14	0.011	1.33	14	0.011	1.22
行为反应	3	0.312	39.33	2	0.346	38.55
形态特点	4	0.230	28.97	4	0.255	28.39
鱼体附着物	13	0.011	1.39	13	0.012	1.31
浮游动物状况	15	0.009	1.19	15	0.010	1.14
浮游植物状况	18	0.008	1.00	18	0.008	0.94
其它水生物状况	16	0.009	1.11	16	0.009	1.02
病原体	2	0.312	39.34	3	0.340	37.84
死鱼发生形式	5	0.178	22.39	5	0.196	21.86
水体溶解氧	12	0.013	1.67	11	0.016	1.73
水体酸碱度	7	0.036	4.41	8	0.031	3.45
水体气味	10	0.014	1.82	12	0.013	1.46
水色	6	0.048	6.09	6	0.042	4.71
急性致死试验	11	0.014	1.81	10	0.016	1.74

## 2.3 讨论

### (1) 中间层节点数的选取

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 4~12),发现节点数较少时(4~7个),网络收敛快,训练次数少(3000~4000次),但精度一般;但节点过多(10~12个),虽精度提高,却又使得训练次数明显增多(8000~9000次),收敛放慢。这是因为当要求逼近精度高时<sup>[5]</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。

### (2) 网络参数的选择

学习率和冲量影响到训练的收敛速度,加大学习率和冲量值能加快收敛速度,但网络的稳定性下降,整个模型的可信赖度也随之降低,因而一般取默认值即可。

## 3 结论

应用 B-P 人工神经网络对渔业水域污染死鱼的原因进行判断,能比较客观的反应出污染的原因,并可以进一步判断出各不同种毒物的具体类别。B-P 人工神经网络还可以通过对权重贡献率的分析,提取出关键性的神经元节点,从而实现输入因子的优选。通过运用已训练好的渔业水域污染死鱼的 B-P 人工神经网络识别模型,只要通过简单的加法和乘法运算,就可以对任何一例污染死鱼的实际样本进行死鱼原因的判断,因而该方法具有较好的适用性和较高的实用性。

### 参考文献:

- [1] 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, 1993年, 1-6.
- [2] Ezugwu E O, Arthur S J, Hines E L. Tool wear prediction using artificial neural networks[J]. Journal of Material Processing Technology, 1995, 49(3-4): 255-264.
- [3] 何小荣, 陈丙珍, 赵晓光, 等. 改善 B-P 网络检验效果的研究[J]. 清华大学学报(自然科学版), 1995, 35(3): 31-36.
- [4] 中国水产科学研究院渔业环境保护研究室. 淡水鱼类急性中毒死亡诊断方法[R]. 南京: 江苏省渔政监督管理局, 1998, 81-94.
- [5] 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.