

BP 网络模型和 LOGIT 模型在森林害虫测报上的应用初报 ——以安徽省潜山县马尾松毛虫为例

张爱兵¹, 陈 建², 王正军¹, 李典谟^{1*}, 田 洁¹

(1. 农业虫鼠害综合治理研究国家重点实验室, 中国科学院动物研究所, 北京 100080; 2. 湖北大学生态科学学院, 武汉 130062)

摘要:以安徽省潜山县马尾松毛虫测报为例, 在潜山县全县设置 102 个诱捕点, 诱集越冬代成虫, 并同时记录环境因子数据, 例如树高、树龄、坡向、植被等。以这些环境因子为测报因子, 以性外激素诱集到的雄蛾数代表田间实际发生情况, 构建了森林重要害虫——马尾松毛虫的分类测报模型——BP 人工神经网络模型和概率测报模型。BP 网络模型的拟合率为 100%, 预留样本检验报准率为 100%。LOGIT 模型的一致对率为 74%。两类模型的拟合结果和预测结果基本吻合。BP 网络的测报能力优于 LOGIT 模型。

关键词:BP 网络模型; LOGIT 模型; 森林害虫; 测报; 马尾松毛虫; 性外激素

The application of BP model and logit model to prediction of forest insect pests

ZHANG Ai-Bing¹, CHEN Jian², WANG Zheng-Jun¹, LI Dian-Mo^{1*}, TIAN Jie¹ (1. The State Key Laboratory of Integrated Management of Insect Pests and Rodents, Institute of Zoology, Chinese Academy of Sciences Beijing 100080; 2. Faculty of Life Science, Hubei University, Wuhan 430062, China). *Acta Ecologica Sinica*, 2001, 21 (12): 2159~2165.

Abstract: The prediction of masson pine moth, *Dendrolimus punctatus*, in Qianshan County of Anhui Province was studied and 102 trapping locations was chosen to place traps-baited with sex pheromone throughout the county. The number of adult males trapped of the overwintering generation was counted and recorded. Some environmental factors such as the height of the pine tree, slope direction, plant type were recorded at the same time. The independent variables consist of these environmental factors and the number of adult males trapped stands for the occurrence of *D. punctatus* in the field. Two models are developed for the classification and the probability prediction of forest pests using BP artificial neural network and LOGIT model with factors related to forest pests. The results of prediction to masson pine moth show that the qualified rates of fitting and prediction using BP model acquire 100%. The rate of concordant in LOGIT model is 74%. The prediction result of BP model agrees with that of LOGIT model. BP model is better than LOGIT model in prediction.

Key words: pest prediction model; forest; sex pheromone; *Dendrolimus punctatus* (Walker)

文章编号: 1000-0933(2001)12-2159-07 中图分类号: Q958.11 文献标识码: A

我国 27 种松毛虫中, 危害较严重的有 6 种, 其中列于第一位的为马尾松毛虫(*Dendrolimus punctatus* Walker)^[1]。马尾松毛虫是典型的东洋区类型, 其危害区主要在长江流域及其以南地区。大量猖獗时, 针叶全部食尽, 似火烧一般, 有“不冒烟的森林火灾”之称^[2]。在江南丘陵及东南沿海丘陵区, 广布马尾松纯林,

基金项目: “国家重点基础研究发展规划”项目(编号: G-2000016210)和中国科学院创新方向项目(KSCX2-1-02)资助。

收稿日期: 2001-03-06; 修订日期: 2001-08-10

作者简介: 张爱兵(1974~), 男, 江苏宝应人, 博士, 主要从事系统生态学、分子生态学研究。

马尾松毛虫危害严重,呈此起彼伏状,马尾松毛虫的发生受诸多环境因子的影响,这种影响往往是非线性的,用常规的数量统计方法找出这些环境因子与其发生的某种函数关系有一定难度。人工神经网络(Artificial neural network,简称 ANN)由于其模拟人脑思维结构,具有很强的自学习、自组织、自适应和容错性等特点,特别适用于非线性问题的处理^[3]。本文采用具 Levenberg-Marquardt 算法的误差反向传播(Back Propagation)的人工神经网络(以下简称 BP 网络)建立马尾松毛虫的分类测报模型,同时作为比较,构建了马尾松毛虫发生的概率测报模型——LOGIT 模型。

1 材料与方法

1.1 研究区域自然概况

安徽省潜山县位于皖西南,长江北岸,大别山东南部,介于东经 116°15′~116°30′,北纬 30°27′~31°04′之间,南北长 62km,东西宽 35km,总面积 1696.8km²。全县人口 56 万,属亚热带季风型气候区,四季分明,雨量充沛,气候温和,光照充足,无霜期较长,全县历年平均气温 16.3℃,年平均降水量 1307.1mm,年无霜期 237d。全县林业用地面积 8.7 万 hm²,森林覆盖率 52%,其中纯松林面积 7.1 万 hm²,占森林总面积的 82.4%。

1.2 诱捕点的分布

诱捕点设在潜山县全县,以常发区为重点,兼顾偶灾区和不发生区,共计 102 个诱捕点,具体分布在黄铺镇、痘母乡、牌楼乡、天柱山乡、占井乡 5 个乡镇,每个乡(镇)10 个点,青楼乡(3 个),岭头乡(16 个),余井乡(17 个),杜埠乡(5 个),源潭乡(2 个),三妙乡(7 个),水吼乡和黄柏乡各 1 个。诱捕器设在地势较高的高大树上。挂置时间为 2000 年 5 月 17 日~6 月 4 日(蛹末期成虫初现到成虫期结束),诱集越冬代成虫,诱捕器的挂置和收集都在同一天进行。

1.3 数据分析

2000 年从潜山县 102 个诱捕点共诱集马尾松毛虫越冬代成虫 1157 头,综合树高、树龄、坡向、植被、郁闭度等数据共 700 多个数据。所有数据均录入 SAS 数据集,并在 SAS6.12 上进行统计分析和 LOGIT 建模。BP 网络模型在 MatLab5.3 上运行和调试。数据按如下公式进行标准化^[4]:

$$x_{ij} = \frac{y_{ij} - \bar{y}_j}{s_j} \quad (i = 1, 2, \dots, t; j = 1, 2, \dots, n) \quad (1)$$

其中

$$\bar{y}_j = \frac{1}{t} \sum_{i=1}^t y_{ij} \quad (j = 1, 2, \dots, n) \quad (2)$$

$$s_j = \left[\frac{1}{t-1} \sum_{i=1}^t (y_{ij} - \bar{y}_j)^2 \right]^{\frac{1}{2}} \quad (j = 1, 2, \dots, n) \quad (3)$$

其中, i 表示因子序号, j 表示样本序号。

2 结果与分析

2.1 BP 网络模型的构建

2.1.1 BP 网络模型的一般原理

人工神经网络(Artificial neural network,简称 ANN)方法作为一种智能算法,由于其非线性、自适应性以及处理复杂系统的能力,已经广泛应用于信号处理、模式识别、最优化问题计算、复杂控制等领域^[5-6],D. E. Rumelhart 和 J. L. McClelland 等于 1986 年首次提出了多层前馈网络的反向传播算法(Back Propagation),简称 BP 网络或 BP 算法^[7]。在神经网络中,前馈式的 BP 网络是一种最简单而用途广泛的人工神经网络^[8]。典型的 BP 网络是三层结构,包括输入层、隐含层和输出层,各层之间实行全连接,如图 1 所示,每个神经元具有 R 个输入,每个输入都通过一个适当的权值 w 与下一层相连,网络输出可表示成

$$a = f(w \times p, b)$$

其中 b 为阈值。BP 网络的传递函数均是可微的单调递增函数,通常采用 sigmoid 型函数——logsig 函数或 tansig 函数。BP 算法由信息正向传播和误差的反向传播两部分组成,算法步骤为:1)以小的随机数对每一

层的权值 W 和偏差 B 初始化;2)计算网络各层输出矢量和误差 E ;3)计算各层反传的误差变化并计算各层权值的修正值和新权值;4)再次计算权值修正后的误差平方和 SSE;5)检查 SSE 是否小于期望误差最小值,若是,训练结束;否则转至 2)继续。算法的详细说明和公式推导参见文献^[11]。

2.1.2 模型的训练和检验

选取与马尾松毛虫林间发生较为有关的 6 个生境因子,树高(x_1)、树龄(x_2)、坡向(x_3)、坡度(x_4)、植被(x_5)和郁闭度(x_6)。已有的 102 个诱捕点样本分为 A、B 两类。BP 网络用于 K 类预报时,其各类样本的期望输出通常设定为 K 维向量 $D_1(1,0,\dots,0), D_2(0,1,0,\dots,0), \dots, D_K(0,0,\dots,1)$ ^[10]。本例每类样本相应的期望输出如下:

样本类别	A 类(偏轻发生)	B 类(偏重发生)
期望输出 D_j	(1,0)	(0,1)

首先样本的各因子数据按公式(1)进行标准化。全部 102 个样本标准化后数值及其类别见表 1。用前 96 个样本作 BP 网络的训练样本集,后 6 个样本留作对模型进行检验。建立 6 个输入节点、10 个隐节点和 2 个输出节点的 3 层 BP 网络,传递函数采用 logsig 函数。以小的随机数赋予网络初始权值和阈值后,将表 1 中 96 个样本因子数值阵和样本期望输出(96×2 阵)作为网络的输入向量和目标向量,采用 Levenberg-Marquardt 规则对网络进行训练,经过不同训练参数组合的比较,发现学习率 $lr=0.01$,学习速率的初始值 $tp(5)=0.001$,增加系数 $tp(6)=10$,减少系数 $tp(7)=0.01$ 时,网络的训练效果比较理想,经过 29 次训练,误差平方和 $SSE=0.0002732 \ll 0.001$,得到终止学习后的权值和阈值(见表 2)。对训练好的网络,重新给以带有 6 个检验样本的输入向量阵,可得前 96 个样本的拟合和后 6 个样本的预测检验结果(见表 1)。其拟合率为 100%,预测准确率为 100%。

表 1 样本因子标准化值及分类、拟合和预测检验结果

Table 1 Normalled values of factors and classification, fitting and predicting results of samples

样品序号 Number code of sample	x_1	x_2	x_3	x_4	x_5	x_6	D_1	D_2	BP 网络预 测结果 Prediction classes	实际结果 Practical classes	LOGIT 模 型的概率 预测结果
1	-1.1916	0.3608	-1.2400	-0.5363	-0.6346	-1.5474	0.0002	0.9997	B	B	0.5392
2	-0.9012	1.7573	-1.2400	-0.5363	-0.6346	-1.5474	0.0004	0.9995	B	B	0.7692
3	-0.9012	0.3608	0.8731	-0.3860	0.9687	-1.5474	1.0000	0.0000	A	A	0.7890
4	-1.4819	-1.0357	0.8731	-0.3860	0.9687	-0.7698	0.9998	0.0002	A	A	0.7089
5	-0.9012	-0.1978	-0.5356	-0.5363	-0.9687	-0.7698	0.9998	0.0001	A	A	0.6629
6	-1.4819	-1.0357	0.8731	-0.5363	-0.6346	-0.7698	1.0000	0.0000	A	A	0.6660
7	-0.9012	-0.1978	-0.5356	-0.5363	0.9687	-0.7698	0.9998	0.0001	A	A	0.6629
8	-1.1916	-1.0357	-0.5356	-0.5363	-0.6346	0.0078	0.0018	0.9981	B	B	0.6071
9	-0.3206	-0.1978	-1.2400	-0.5363	-0.6346	0.7854	0.0006	0.9993	B	B	0.7382
10	-0.6109	-0.7564	-0.5356	-0.3860	0.9687	0.0078	0.0003	0.9996	B	B	0.6556
11	1.4214	1.1987	0.1688	-0.3860	-0.6346	1.5630	0.9999	0.0001	A	A	0.9479
12	0.2601	0.9194	0.1688	0.8162	-0.6346	-0.7698	1.0000	0.0000	A	A	0.8318
13	1.4214	1.4780	1.5775	-0.6866	-0.6346	0.0078	1.0000	0.0000	A	A	0.9383
14	1.4214	2.0366	0.8731	-0.5363	-0.6346	-0.7698	1.0000	0.0000	A	A	0.9102
15	0.2601	-1.3150	-0.5356	-0.3860	0.9687	0.7854	0.9996	0.0005	A	A	0.6238
16	-0.3206	-0.1978	0.1688	-0.5964	-0.6346	1.5630	0.9996	0.0004	A	A	0.9054

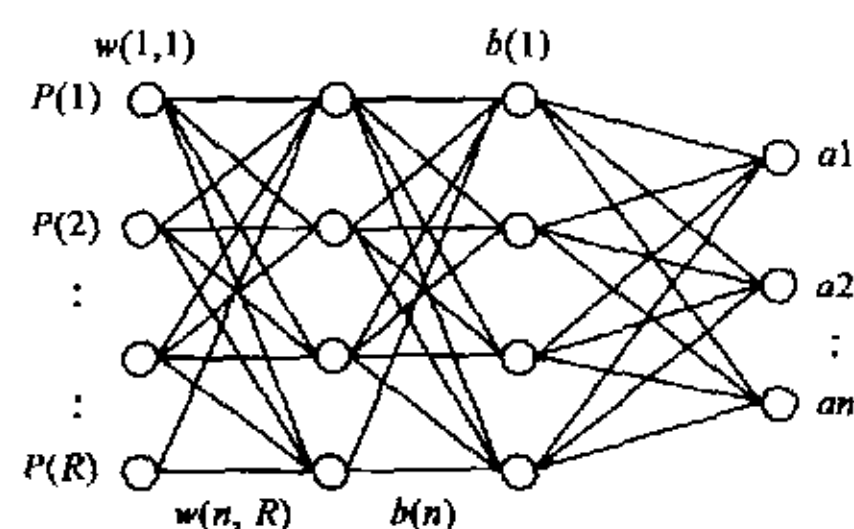


图 1 BP 网络结构

Fig. 1 BP artificial neural network

$p(1), p(2), \dots, p(R)$: 网络输入; $w(1,1), \dots, w(n, R)$: 网络权值; $b(1), b(2), \dots, b(n)$: 阈值; $a1, a2, \dots, an$: 网络输出

$p(1), p(2), \dots, p(R)$: input of network; $w(1,1), \dots, w(n, R)$: weight of network; $b(1), b(2), \dots, b(n)$: threshold of network; $a1, a2, \dots, an$: output of network

续表

样品序号 Number code of sample	x_1	x_2	x_3	x_4	x_5	x_6	D_1	D_2	BP 网络预 测结果 Prediction classes	实际结果 Practical classes	LOGIT 模 型的概率 预测结果
17	0.2601	0.3608	-0.5356	1.5676	0.9687	0.0078	1.0000	0.0000	A	A	0.8336
18	0.2601	0.3608	0.1688	1.5676	0.9687	-0.7698	1.0000	0.0000	A	A	0.8369
19	0.2601	-0.1978	1.5775	-0.6866	-0.6346	-0.7698	0.9996	0.0003	A	A	0.7646
20	0.8407	1.7573	0.1688	-0.3860	-0.6346	0.7854	1.0000	0.0000	A	A	0.9524
21	1.1310	1.4780	0.1688	2.3190	0.9687	2.3406	1.0000	0.0000	A	A	0.9867
22	2.0020	1.4780	0.1688	-0.3860	-0.6346	1.5630	1.0000	0.0000	A	A	0.9505
23	-0.6109	-1.0357	0.1688	2.0185	0.9687	-0.7698	0.0003	0.9995	B	B	0.6503
24	-0.6109	-0.1978	0.1688	2.0185	2.5721	0.7854	1.0000	0.0000	A	A	0.9270
25	-0.9012	-1.3150	1.5775	-0.6866	-0.6346	-0.7698	0.0008	0.9991	B	B	0.6498
26	-1.4819	-1.5943	1.5775	-0.6866	-0.6346	-0.7698	1.0000	0.0000	A	A	0.6376
27	-0.9012	-1.0357	1.5775	-0.6866	-0.6346	0.0078	0.9994	0.0008	A	A	0.7934
28	-0.0303	-0.1978	1.5775	-0.6866	-0.6346	0.0078	0.9984	0.0017	A	A	0.8577
29	0.2601	-0.1978	1.5775	-0.6866	-0.6346	0.0078	0.0027	0.9975	B	B	0.8457
30	-0.9012	-1.0357	-1.2400	-0.5363	-0.6346	-0.7698	0.0000	1.0000	B	B	0.3700
31	1.1310	0.3608	-1.2400	-0.3860	0.6346	0.7854	0.0013	0.9988	B	B	0.7460
32	0.8407	0.9194	-1.2400	-0.2358	-0.6346	2.3406	0.9999	0.0001	A	A	0.9371
33	1.4214	0.6401	0.1688	-0.2358	-0.6346	1.5630	0.0023	0.9976	B	B	0.9219
34	-0.6109	0.3608	-0.5356	1.5676	0.9687	0.7854	1.0000	0.0000	A	A	0.9170
35	1.4214	0.3608	0.1688	1.5676	-0.6346	0.0078	1.0000	0.0000	A	A	0.8077
36	-0.3206	-1.3150	-1.2400	-0.2358	-0.6346	1.5630	0.9969	0.0027	A	A	0.6689
37	0.8407	0.9194	-1.2400	-0.5363	-0.6346	0.7854	0.9999	0.0001	A	A	0.8315
38	-0.6109	-0.1978	-0.5356	2.3190	2.5721	0.0078	0.0002	0.9998	B	B	0.8460
39	0.2601	0.3608	0.8731	-0.5363	-0.6346	0.7854	0.9997	0.0003	A	A	0.9141
40	0.8407	-0.1978	0.1688	0.5363	-0.6346	1.5630	0.0013	0.9986	B	B	0.8721
41	-0.3206	-1.0357	0.1688	2.3190	0.9687	2.3406	0.9998	0.0002	A	A	0.9371
42	2.0020	1.1987	1.5775	-0.6866	-0.6346	1.5630	0.0100	0.9903	B	B	0.9672
43	1.1310	0.9194	1.2400	-0.6866	-0.6346	0.7854	1.0000	0.0000	A	A	0.8118
44	0.2601	0.9194	-1.2400	-0.5363	0.6346	-1.5474	0.0000	1.0000	B	B	0.5440
45	0.8407	0.3608	-1.2400	-0.6866	-0.6346	0.0078	0.9996	0.0004	A	A	0.5435
46	0.5504	0.3608	-1.2400	-0.5363	-0.6346	0.7854	0.9991	0.0009	A	A	0.7737
47	-1.1916	-0.1978	-0.5356	-0.5363	0.9687	-0.7698	1.0000	0.0000	A	A	0.6820
48	-1.4819	-1.0357	1.5775	-0.6866	0.9687	-1.5474	0.9998	0.0002	A	A	0.6606
49	-0.9012	2.0366	0.1688	-0.3860	-0.6346	-1.5474	1.0000	0.0000	A	A	0.8958
50	1.4819	-1.0357	0.1688	-0.5363	-0.6346	-1.5474	0.0001	0.9998	B	B	0.4525
51	1.4214	3.1538	0.5356	-0.5363	-0.6346	0.0078	1.0000	0.0000	A	A	0.9550
52	0.2601	-0.1978	1.5775	-0.6866	-0.6346	0.0078	1.0000	0.0001	A	A	0.8467
53	-0.9012	-0.1978	-0.5356	-0.5363	0.9687	-1.5474	0.0002	0.9997	B	B	0.5362
54	-0.9012	-0.1978	1.2400	-0.5363	2.5721	0.0078	0.9988	0.0002	A	A	0.7382
55	-1.1916	-0.7564	1.5775	-0.6866	2.5721	-0.7698	0.9996	0.0005	A	A	0.8199
56	-0.9012	-0.7564	1.5775	-0.6866	-0.6346	-1.5474	0.9999	0.0001	A	A	0.6319
57	-0.9012	-0.4771	1.5775	-0.6866	-0.6346	-1.5474	0.0013	0.9987	B	B	0.6829
58	-1.4819	-1.0357	1.5775	-0.5363	-0.6346	-0.7698	0.9999	0.0001	A	A	0.7383
59	-1.1916	-0.7564	-0.5356	-0.6866	-0.6346	-0.7698	0.9992	0.0004	A	A	0.5275
60	-0.9012	-0.1978	1.5775	-0.6866	-0.6346	0.0078	1.0000	0.0000	A	A	0.8865

续表

样品序号 Number code of sample	x_1	x_2	x_3	x_4	x_5	x_6	D_1	D_2	BP 网络预 测结果 Prediction classes	实际结果 Practical classes	LOGIT 模 型的概率 预测结果
61	-1.1916	-1.0357	-1.2400	-0.5363	-0.6346	0.0078	0.9999	0.0001	A	A	0.5211
62	-1.4819	-1.5943	1.5775	-0.6866	-0.6346	-0.7698	1.0000	0.0000	A	A	0.6376
63	-0.9012	-1.0357	-1.2400	-0.5363	2.5721	-0.7698	0.9996	0.0004	A	A	0.4564
64	-0.3206	-0.1978	1.5775	-0.6866	0.9687	0.0078	0.9999	0.0002	A	A	0.8871
65	0.2601	0.3608	-0.5356	-0.5363	-0.6346	1.5630	0.9999	0.0001	A	A	0.9900
66	-0.9012	-1.0357	-0.5356	0.5363	-0.6346	-0.7698	0.0000	1.0000	B	B	0.1544
67	0.2601	-0.4771	-1.2400	-0.2358	-0.6346	0.0078	0.0002	0.9998	B	B	0.5368
68	-0.6109	1.0357	-0.5356	-0.5363	0.9687	0.0078	0.0001	0.9999	B	B	0.6083
69	-0.6109	-0.7564	0.8731	0.0648	0.9687	-0.7698	0.0004	0.9995	B	B	0.7151
70	-0.6109	0.3608	0.8731	0.8162	0.9687	1.5630	1.0000	0.0000	A	A	0.9714
71	-0.6109	-0.7564	0.8731	2.3190	0.9687	0.0078	0.9999	0.0000	A	A	0.8545
72	-0.6109	-1.0357	-0.5356	2.0185	0.9687	-0.7698	0.0000	1.0000	B	B	0.5673
73	-0.6109	-0.1978	-1.2400	2.3190	2.5721	0.7854	0.9997	0.0003	A	A	0.8682
74	-0.6109	0.3608	-0.5356	3.0704	2.5721	0.7854	1.0000	0.0000	A	A	0.9424
75	1.7117	1.7573	0.1688	-0.3860	-0.6346	-0.7698	1.0000	0.0000	A	A	0.8422
76	0.8407	0.3608	0.1688	-0.5363	0.6346	0.7854	0.9998	0.0002	A	A	0.8632
77	1.4214	0.3608	-1.2400	-0.6866	-0.6346	0.7854	0.9998	0.0002	A	A	0.7207
78	0.8407	-0.7564	-1.2400	0.8162	-0.6346	0.7854	0.0001	0.9998	B	B	0.6052
79	2.0020	0.9194	0.1688	-0.5363	-0.6346	1.5630	1.0000	0.0000	A	A	0.9227
80	0.8407	0.9194	-1.2400	-0.5363	-0.6346	0.7854	0.9999	0.0001	A	A	0.8315
81	0.2601	0.3608	-1.2400	-0.3860	0.6346	0.7854	0.9991	0.0008	A	A	0.7921
82	0.2601	0.3608	-1.2400	-0.5363	-0.6346	-0.7698	0.9991	0.0011	A	A	0.5632
83	2.0020	-0.1978	1.5775	-0.6866	-0.6346	0.0078	1.0000	0.0000	A	A	0.7665
84	0.8407	0.9194	-1.2400	-0.5363	0.6346	0.7854	0.9999	0.0001	A	A	0.8315
85	1.4214	2.0366	0.8731	-0.3860	-0.6346	0.0078	1.0000	0.0000	A	A	0.9163
86	1.4214	1.4780	0.8731	-0.5363	-0.6346	0.0078	1.0000	0.0000	A	A	0.9164
87	0.8407	1.7573	-0.5356	0.8162	-0.6346	-0.7698	1.0000	0.0000	A	A	0.8526
88	1.4214	0.6401	0.1688	1.4173	0.9687	0.0078	1.0000	0.0000	A	A	0.8605
89	-0.3206	-0.7564	0.1688	2.0185	0.9687	-0.7698	0.9999	0.0001	A	A	0.6814
90	-0.3206	-0.7564	0.1688	2.0185	0.9687	-0.7698	0.9999	0.0001	A	A	0.6814
91	0.8407	-0.4771	0.1688	0.0648	-0.6346	-1.5474	1.0000	0.0000	A	A	0.4143
92	-0.3206	-1.8736	-0.5356	-0.3860	-0.6346	0.7854	0.0002	0.9998	B	B	0.5117
93	0.8407	-0.4771	0.5356	0.0648	-0.6346	0.0078	0.9998	0.0002	A	A	0.5905
94	0.2601	-0.7564	-0.5356	0.0648	-0.6346	-0.7698	0.0001	0.9999	B	B	0.4457
95	0.2601	-0.7564	-0.5356	0.0648	-0.6346	0.0078	0.0002	0.9997	B	B	0.5776
96	-0.9012	-1.0357	-0.5356	1.5676	-0.6346	0.0078	0.9993	0.0004	A	A	0.6561
97*	0.2601	0.3608	-1.2400	-0.533	-0.6346	0.0078	0.9999	0.0001	A	A	0.6868
98*	1.4214	1.4780	-1.2400	-0.5363	-0.6346	0.7854	1.0000	0.0000	A	A	0.8672
99*	-0.3206	0.1978	0.1688	0.0648	-0.6346	1.5630	0.0078	0.9933	B	B	0.9131
100*	0.6109	-0.7564	0.8731	-0.3860	0.9687	-0.7698	0.0005	0.9994	B	B	0.7019
101*	0.8407	1.7573	0.1688	-0.2358	-0.6346	1.5630	1.0000	0.0000	A	A	0.9721
102*	0.2601	0.3608	0.1688	1.5676	0.9687	0.7854	1.0000	0.0000	A	A	0.9236

* 代表预测样本 Reprerents prediction sample.

表 2 训练结束后的权值和阈值

Table 2 Trained weight and threshold values

W_1	-3.4918	-4.0313	2.0561	0.3252	-9.0534	0.2812				
	10.2118	-0.5088	-0.2124	0.7001	-0.4083	3.2499				
	-0.9195	2.2816	-1.1589	4.0543	-5.2103	6.0495				
	-5.1294	1.8524	6.0451	2.0439	6.0862	6.0584				
	-5.8476	0.8012	5.3088	3.4311	2.0391	2.7206				
	-6.1105	5.9558	-0.0700	0.4186	1.2527	5.9815				
	2.7117	-1.7032	2.0290	6.5229	0.4159	-3.9209				
	-1.7814	-8.6296	-3.7334	0.1704	3.8801	2.9676				
	2.3141	1.5453	6.9495	1.2542	-0.4842	1.3146				
	5.3163	8.5338	3.0708	-0.9907	-2.1618	4.2556				
W_2	9.6906	6.6394	5.7493	4.7616	3.9570	1.7059	3.6369	-3.0828	-7.7356	-3.3442
	1.3665	-4.4495	5.0562	3.1675	4.0960	-1.7265	-0.0856	2.5110	-0.1758	2.1572
	1.2260	-2.1863	5.4615	2.8609	0.2124	-1.8408	1.7256	-1.3265	0.6455	-3.0689
	2.2283	-2.0546	1.0017	4.8756	-5.1208	0.4563	-0.4146	0.8880	3.5318	2.58365
	-0.7523	3.4026	9.9692	-5.6001	5.6286	6.8924	-10.3413	-8.5653	6.0072	-9.5730
	-2.7950	-2.5769	2.9072	-2.2224	-2.1307	-1.6781	1.8583	-1.0273	1.3928	2.5906
	-5.8650	-2.7757	-2.6821	1.2904	-0.5601	-0.5125	1.7447	3.2823	2.0316	3.8558
	-0.0385	-0.4502	-5.4274	-6.1650	-1.8091	-2.4342	1.1734	-4.5516	-1.4999	3.5334
	-1.8024	0.7795	-3.7678	-2.9809	-2.7653	-0.3805	2.8180	2.1785	3.0415	4.1704
	-2.1227	1.2388	-1.7138	0.8750	-0.4688	3.3700	2.6872	-0.3998	0.8164	1.9644
W_3	-14.6900	2.0124	-0.8540	4.8963	17.9613	4.0151	4.7880	-4.0611	2.1563	1.3297
	14.7178	-3.1299	2.9885	-4.2903	-17.9967	-2.7725	-4.3808	4.7292	2.1612	-1.1470
b_1^{**}	3.3745	7.8568	1.9439	1.4433	0.5582	3.4495	7.2241	-2.3056	1.8810	3.5612
b_2^{**}	-8.2195	1.6739	2.7879	1.2676	1.8940	6.3743	-0.7072	0.9066	5.9447	7.4506
b_3^{**}	-2.8765	-0.2508								

* * * $b_1^{**}, b_2^{**}, b_3^{**}$, 分别为 b_1, b_2, b_3 的转置矩阵。

2.2 LOGIT 模型的构建

2.2.1 LOGIT 模型的一般原理

LOGIT 模型是用来刻画属性变量(Y)(一般只取有限个离散值)取某一特定值的可能性大小(概率) p 与自变量 $x_1, x_2, x_3, \dots$ 之间关系的模型,因而通常的用于连续变量的回归分析已不适用。利用 logit 变换:

$$\text{Logit}(p) = \log(p/(1-p)) \quad (4)$$

可将属性变量取某个值的概率 p 的 logit 变换表示为自变量的线性函数,即

$$\text{Logit}(p) = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k \quad (5)$$

或等价地

$$p = \exp(b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k) / [1 + \exp(b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k)] \quad (6)$$

拟合 Logit 模型不采用通常的最小二乘法,而是采用最大似然法。衡量 Logit 模型拟合好坏的常用指标是预测和观测数据中一致对(concordant)与不一致对(discordant)的比例。考虑数据对

$$(Y_i, P_i), (Y_j, P_j), Y_i \neq Y_j$$

若 $(Y_i - Y_j)(p_i - p_j) > 0$, 称数据对为一致的,

若 $(Y_i - Y_j)(p_i - p_j) < 0$, 称数据对为不一致的,

若 $(Y_i - Y_j)(p_i - p_j) = 0$, 称数据对为平分的。

这是衡量拟合的 Logit 模型,能否反映实际的趋势的一种方法。模型的整体有效度以对数可能率表示,其值等于 $[-2\log\text{likelihood}]^{[9]}$ 。

2.2.2 LOGIT 模型用于马尾松毛虫概率测报建模

选取与马尾松毛虫林间发生较为有关的 6 个生境因子,树高(x_1)、树龄(x_2)、坡向(x_3)、坡度(x_4)、植被(x_5)和郁闭度(x_6)。以性外激素诱捕器诱蛾量代表林间实际发生情况,分为偏重、偏轻两类。数据的标准化按公式(1),标准化后的数据见表 1。对上述数据,在 SAS6.12 上编程作 Logistic 回归,其 LOGIT 模型为:

$$\text{logit}(p) = 1.3270 - 0.2989x_1 + 0.8118x_2 + 0.4964x_3 + 0.1415x_4 + 0.1115x_5 + 0.6830x_6 \quad (7)$$

(其中, p 为偏轻发生的概率)。

模型的一致对(concordant)率为 74%,对数可能率-2LOG L 为 12.653($p=0.0489<0.05$)。依据公式(6),计算出 102 个诱捕点马尾松毛虫偏重发生的概率,结果见表 1。

3 小结与讨论

(1)BP 网络具有自适应性、抗干扰和容错性强能力,个别学习样本的分量偏差对学习效果的影响较小,因此模型的稳定性较好。

(2)BP 网络模型是一种非线性模型,不需要关于各种变量之间关系的额外知识,通过网络的学习,精确地建立待测量与影响因子之间的函数关系,从而实现对待测量的经验性预测。比如本例对马尾松毛虫发生程度的类别测报,在生产上比较实用。

(3)采用 BP 网络模型进行预测,预测精度要优于多元线性回归等传统方法。但 BP 网络模型的权重对影响因子的重要性很难作出直观的解释,是其一大弱点。

(4)本文构建的马尾松毛虫发生类别的 LOGIT 概率测报模型,能够给出危害程度的概率值,在林业生产上具有重要的指导作用,其预测能力略低于 BP 网络模型。

参考文献

- [1] 陈昌洁. 松毛虫综合管理. 北京:中国林业出版社,1990. 5~18.
- [2] 侯陶谦. 中国主要害虫综合防治. 北京:科学出版社,1979. 370~400.
- [3] Widrow B. 30 years of adaptive neural networks:perceptron, Madaline, and B-P. Proc. IEEE, 1990, 78(9):1415~1442.
- [4] 徐克学. 生物数学. 北京:科学出版社,1999. 76~83.
- [5] Rumelhart D E & McClelland J L. *Parallel Distributed Processing-Exploration in the Microstructure of Cognition*, Vols 1 and 2. MIT Press, 1986. Cambridge, MA.
- [6] Lippmann, R. P. An introduction to computing with neural nets. IEEE Acoust. Speech Signal Process, 1987, 4(2):1~22.
- [7] 楼顺天,施 阳. 基于 MATLAB 的系统分析与设计——神经网络. 西安:西安电子科技大学出版社,1998. 2~114.
- [8] 王 伟. 人工神经网络原理——入门与应用. 北京:北京航空航天大学出版社,1995. 52~76.
- [9] 彭昭英. 世界统计与分析全才 SAS 系统应用指南. 北京:北京希望电子出版社,2000. 387~411.
- [10] 李祚泳,彭荔红. 基于人工神经网络的农业病虫害预测模型及其效果检验. 生态学报,1999, 19(5):759~762.
- [11] 闻 新,周 露,王丹力,等. MATLAB 神经网络应用设计. 北京:科学出版社,2000. 208~212.