

文章编号: 1001-1498(2001)05-0503-06

鞭角华扁叶蜂滞育幼虫空间数量分布的研究

陈绘画, 崔相富, 郑永祥

(浙江省仙居县林业局, 浙江 仙居 317300)

摘要: 运用人工神经网络方法和空间趋势面分析方法对鞭角华扁叶蜂滞育幼虫的数量与其所处的空间位置之间的关系进行研究, 创建鞭角华扁叶蜂滞育幼虫空间数量分布的BP网络模型和三次趋势面模型。结果表明: 建立的BP网络模型模拟不同空间位置的鞭角华扁叶蜂滞育幼虫数量具有较高的拟合精度, 当隐层神经元数为12个时, 其平均模拟精度为96.90%, 预留的20组样本的平均预测精度为92.35%; 而三次趋势面模型的平均模拟精度为93.01%, 预留的20组样本的平均预测精度为93.24%。

关键词: 人工神经网络; BP算法; 鞭角华扁叶蜂滞育幼虫; 空间格局; 空间趋势面

中图分类号: S718.7

文献标识码: A

空间格局是昆虫种群的重要特征之一。它的形成是由昆虫的生物学特性与特定生境条件相互作用、协同进化的结果。研究空间格局有助于了解昆虫的生态学特性, 对于虫口数量调查的正确取样、种群消长趋势及其扩散范围的预测预报、捕食与猎物关系、种内竞争关系以及制订害虫防治和管理策略都具有重要的意义。因此, 生态学家们利用经典生物统计学和统计生态学提供的手段给出了多种判定昆虫空间格局的准则^[1,2]。这些判定方法的共同点是依赖于样本频次分布或均值与方差关系及其相应的判定准则确定空间分布格局, 仅考虑样本间的数量关系而忽视样本的空间位置; 应用生物地理统计学或地质统计学研究昆虫的空间格局, 虽同时利用了数量和空间位置的关系, 但也只是解决了昆虫种群的空间相关性和依赖性, 仍无法解释昆虫种群随空间位置的数量变化关系^[3~5]。人工神经网络(简称神经网络)是由大量的、功能比较简单的形式神经元互相连接而成的复杂网络系统, 其特色在于信息的分布式存储和并行协同处理, 具有很强的自学习、自组织、自适应及容错性等, 善于联想、综合和推广^[6,7]。趋势面分析是对地质特征的空间分布进行研究和分析的一种方法, 它是用某种形式的函数所代表的曲面来逼近该地质特征的空间分布。趋势面分析由Gittin引入生态学研究之后, 由于该方法简单明确、计算方便等优点, 在生态学研究中得到了广泛的应用^[8,9]。因此, 本文采用人工神经网络和趋势面分析两种方法来研究相对稳定的鞭角华扁叶蜂 [*Chinolyda flagellicornis* (F. Smith)] 滞育幼虫(又称入土幼虫)的数量与空间位置的关系, 为人工神经网络理论和趋势面分析方法在森林昆虫空间格局研究中的应用进行积极的探索。

收稿日期: 2000-12-04

基金项目: 浙江省仙居县科委“八五”攻关项目“鞭角华扁叶蜂综合防治技术研究”内容之一

作者简介: 陈绘画(1968-), 男, 浙江仙居人, 工程师

1 材料与方法

1.1 试验地概况

试验地位于浙江省仙居县萍溪林场,地处 28°56' N, 120°33' E, 海拔高 400~480 m, 年平均气温 17.2℃, 年平均降雨量 1376.8 mm。林分郁闭度为 0.7, 受害程度为重。林下植被为稀疏的扁茅草和小灌木。黄壤土, 土层深 0.4~0.7 m, 坡度 30~40°, 南坡。

1.2 调查方法

1991年8月在浙江省仙居县萍溪林场田塘林区选择一受严重危害的柏木(*Cupressus funebris* Endl)林分, 在一块 80 m × 80 m 的调查样地中, 按随机抽样方式抽取 100 块样方, 样方面积均为 (2 × 2) m²。然后清理样方内及其周围杂草, 在样方一侧挖一条 35 cm 深的沟, 再从沟边开始向样方内细细深挖, 深度达 35 cm 以上, 记录样方虫量及其位置。各样方内虫量的对数值及其位置详见表 1。

表 1 鞭角华扁叶蜂滞育幼虫数量对数值

纵坐 标/m	横 坐 标/m									
	0.5	1.3	2.1	2.9	3.7	4.5	5.3	6.1	6.9	7.7
0.5	6.748*	6.590	7.122	5.707	5.958	5.943	7.290	7.181	5.740	6.450*
1.3	6.384	6.047*	6.758	5.956	5.984	5.762	6.190	5.883	6.125*	6.729
2.1	6.066	5.714	5.442*	5.951	5.886	5.649	6.014	6.176*	5.635	7.249
2.9	7.212	7.266	6.125	6.457*	5.889	6.981	6.763*	6.366	6.290	7.132
3.7	6.784	7.077	7.063	7.006	5.613*	5.587*	6.347	7.032	5.855	5.624
4.5	7.460	6.592	6.945	5.700	6.746*	6.788*	7.448	7.272	6.567	5.826
5.3	5.533	5.846	6.377	6.131*	6.625	5.903	6.495*	6.477	6.948	6.758
6.1	5.846	6.242	6.460*	5.799	6.089	6.933	5.743	5.864*	5.781	7.246
6.9	5.875	6.658*	6.848	6.244	5.956	6.236	6.004	6.746	5.799*	6.127
7.7	6.094*	5.826	6.739	5.762	5.730	6.483	5.886	6.777	6.960	6.692*

注: 纵坐标、横坐标均以 10 m 设置; 带 * 数据作为预留样本, 用于回报检验。

1.3 研究方法

1.3.1 BP 算法 BP (Back Propagation) 算法是一种采用误差反向传播算法的自动学习过程。它不仅有输入层节点(即神经元)和输出层节点, 而且有隐层节点(可以是一层或多层)。对于输入信号, 要先向前传播到隐节点, 经过作用函数后, 再把隐节点的输出信息传播到输出节点, 最后给出输出结果。输入层到隐层的节点作用函数通常取为 sigmoid 函数, 即

$$F(x) = 1/[1 + \exp(-x)]$$

而隐层到输出层的节点作用函数则采用线性激励函数。

这个算法的学习过程, 由正向传播和反向传播组成。在正向传播过程中, 输入信息从输入层经隐单元逐层处理, 并传向输出层, 每一层神经元的状态只影响下一层神经元的状态。如果在输出层不能得到期望的输出, 则转入反向传播, 将误差信号沿原来的连接通路返回, 通过修改各层神经元的权值, 使得误差信号最小。

BP 算法的步骤可概述如下:

(1) 权值和阈值初始化: 随机地给全部权值 W_{lm}^1 、 W_{mn}^2 和神经元的阈值 θ_l 、 ψ_n 赋以初始值。 W_{lm}^1 与 W_{mn}^2 分别是输入层到隐含层、隐含层到输出层之间的连接权值, l 为输入层第 l 个因子, m 为隐含层第 m 个神经元, n 为输出层第 n 个神经元; θ_l 与 ψ_n 分别是隐含层与输出层的阈

值。

(2) 给定输入 $x_j (j=1, 2, \dots, a)$ 和目标输出 y_j 。

(3) 计算实际输出 y_j :

$$y_j = f(\sum_i W_{mn}^2 y_i - \Psi_n)$$

其中 $f(\cdot)$ 为函数, y_i 为隐节点的输出, $y_i = f(\sum_j W_{jm}^1 x_j - \theta_i)$ 。

(4) 修正权值: 从输出层开始, 将误差信号沿连接通路反向传输方向传播, 通过修正各权值 W_{jm}^1 与 W_{mn}^2 (统记为 W_{ij}) 和阈值 θ_i 与 Ψ_n (统记为 θ), 使误差最小:

$$W_{ij}(t+1) = W_{ij}(t) + \eta \delta_{pj} y_j$$

$$\theta_i(t+1) = \theta_i(t) + \eta \delta_{pi}$$

其中 t 为迭代次数, η 为增益项, δ_{pj} 为 j 节点 p 模式的误差项, 这里误差项 E_i 定义为

$$E_i = \frac{1}{2} \sum_j (y_j - y_j)^2$$

且有:) 若 j 为输出节点, 则 $\delta_{pj} = y_j(1 - y_j)(y_j - y_j)$;) 若 j 为隐节点, 则 $\delta_{pj} = y_j(1 - y_j)$

$\sum_k \delta_{pk} W_{jk}$, 式中 δ_{pk} 为隐含层 k 节点 p 模式的误差项。

(5) 达到误差精度或循环次数要求, 则输出结果, 否则回到步骤(2)。

BP 网络的原理和算法详见有关文献[10~12]。

1.3.2 趋势面分析 设变量 x 、 y 和 z 分别表示平面上某点的两个坐标和该点的滞育幼虫数量, 则三次趋势面方程为:

$$z = a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2 + a_6x^3 + a_7x^2y + a_8xy^2 + a_9y^3 \quad (1)$$

式中: a_0, a_1, \dots, a_9 为待定系数。

2 结果与分析

2.1 网络输入量与输出量的确定

本文主要研究鞭角华扁叶蜂滞育幼虫数量与不同空间位置之间的关系, 因此, 输入量选为样方的纵坐标、横坐标, 输出量选为样方内滞育幼虫数的对数值。

2.2 网络隐层神经元个数的确定

隐层神经元个数的选择是一个十分复杂的问题, 不同的研究者有不同的观点^[6, 10~14]。根据文献[12], 采用公式 $j = \sqrt{n+m} + a$ (其中 m 为输出神经元数, n 为输入神经元数, a 为 1~10 之间的常数) 确定隐神经元数。结合网络的泛化功能, 经多次拟合和筛选, 本文的隐层神经元数最终确定为 12 个。

2.3 网络模拟结果

在带有神经网络工具箱的 MATLAB 语言 5.2 版环境下, 建立一个含一个隐层的 BP 网络, 将表 1 中样方位置的纵坐标、横坐标作为输入样本, 样方内鞭角华扁叶蜂滞育幼虫数的对数值作为目标矢量。输入层到隐层的传递函数为 `tansig`, 隐层到输出层的传递函数为 `purelin`, 学习函数为 `learngdm`, 训练函数为 `trainlm` (各训练参数均采用默认值), 利用 `network/sim` 函数计算网络输出。当训练了 150 000 次时, 平均均方误差 (mean squared error) $MSE = 0.0797758$, 达到最小, 于是中断网络学习, 输出 80 组学习信息的网络输出值, 结果列于表 2

中。

从表 2 的结果可以看出,用人工神经网络所建立的模型对数据的模拟结果,平均模拟精度

表 2 BP 模型模拟各样方鞭角华扁叶蜂滞育幼虫数量误差

实测值	模拟值	精度/%	实测值	模拟值	精度/%
6.590	6.585 2	99.93	7.460	7.176 6	96.20
7.122	6.977 7	97.97	6.592	6.560 9	99.53
5.707	5.891 6	96.77	6.945	6.453 3	92.92
5.958	6.068 2	98.15	5.700	6.502 8	85.92
5.943	5.956 4	99.77	7.448	7.210 1	96.81
7.290	7.219 9	99.04	7.272	7.153 1	98.36
7.181	7.181 0	100	6.567	6.758 6	97.08
5.740	5.740 0	100	5.826	5.939 0	98.06
6.384	6.114 2	95.77	5.533	5.547 3	99.74
6.758	6.826 8	98.98	5.846	6.169 6	94.46
5.956	6.093 0	97.70	6.377	6.089 7	95.49
5.984	5.955 6	99.53	6.625	6.328 6	95.53
5.762	5.785 6	99.59	5.903	5.751 6	97.44
6.190	5.613 6	90.69	6.477	6.840 5	94.39
5.883	6.082 2	96.61	6.948	6.864 0	98.79
6.729	6.729 0	100	6.758	6.735 6	99.67
6.066	6.137 1	98.83	5.846	5.775 6	98.80
5.714	6.040 7	94.28	6.242	6.203 5	99.38
5.951	5.951 0	100	5.799	6.003 6	96.47
5.886	5.860 7	99.57	6.089	6.009 7	98.70
5.649	5.821 4	96.95	6.933	6.953 8	99.70
6.014	5.889 8	97.93	5.743	6.254 7	91.09
5.635	5.939 6	94.59	5.781	6.420 5	88.94
7.249	7.206 4	99.41	7.246	6.411 4	88.48
7.212	7.275 9	99.11	5.875	5.857 0	99.68
7.266	6.759 6	93.03	6.848	6.253 9	91.32
6.125	6.202 8	98.73	6.244	6.164 2	98.72
5.889	5.702 5	96.83	5.956	6.097 2	97.63
6.981	6.921 9	99.15	6.236	6.473 8	96.19
6.366	6.221 8	97.73	6.004	5.920 0	98.60
6.290	6.536 5	96.08	6.746	6.502 5	96.39
7.132	6.995 2	98.08	6.127	6.235 8	98.22
6.784	6.966 8	97.31	5.826	5.900 1	98.73
7.077	7.077 1	100	6.739	6.383 3	94.72
7.063	7.094 0	99.56	5.762	6.325 5	90.22
7.006	6.925 1	98.85	5.730	6.266 4	90.64
6.347	7.124 3	87.75	6.483	6.373 5	98.47
7.032	6.528 6	92.84	5.886	5.886 0	100
5.855	5.809 8	99.23	6.777	6.671 2	98.44
5.624	5.594 2	99.47	6.960	6.717 6	96.52
			平均值		96.90

为 96.90%,说明用人工神经网络方法研究鞭角华扁叶蜂滞育幼虫种群空间数量分布,其效果是理想的。利用训练结束后的模拟模型,即可对样地内各样本点的滞育幼虫数量进行预测预报,只要输入样本所处的空间位置的纵坐标、横坐标,就可由模型预测出该点滞育幼虫数量的对数值。而且经过训练后的BP网络,还具有联想学习功能(网络的泛化功能),即可以预测出没有参与建模样本点滞育幼虫数量的对数值。

利用BP网络的联想学习功能,计算出预留的20组样本的输出值,然后进行回报检验,其平均检验精度为92.35%,详见表3。

表 3 BP 模型模拟各样方鞭角华扁叶蜂滞育幼虫数量回报检验

实测值	模拟值	精度/%	实测值	模拟值	精度/%
6.748	6.838 7	98.66	6.746	7.266 0	92.29
6.450	6.774 8	94.66	6.788	6.783 7	99.94
6.047	6.509 5	92.35	6.131	6.044 2	98.58
6.125	7.194 2	82.54	6.495	6.718 6	96.56
5.442	6.389 6	82.59	6.466	6.107 2	94.45
6.176	5.956 4	96.44	5.864	6.461 8	89.81
6.457	6.261 8	96.98	6.658	5.805 3	87.19
6.763	6.445 3	95.30	5.799	6.397 2	89.68
5.613	6.034 0	92.50	6.094	5.934 5	97.38
5.587	7.197 7	71.17	6.692	6.537 4	97.69
			平均值		92.35

2.4 三次趋势面方程

将表 1 中数据代入公式(1),得到鞭角华扁叶蜂滞育幼虫的三次趋势面方程为:

$$z = 6.667 3 + 0.082 8x - 0.270 6y - 0.003 3x^2 + 0.005 9xy + 0.028 9y^2 - 0.003 7x^3 + 0.008 3x^2y - 0.006 9xy^2 + 0.001 3y^3 \quad (2)$$

(剩余方差 0.895, R = 0.324 065)

将观测值分解成趋势部分和偏差部分两部分:

$$z_{\text{观}} = z_{\text{趋}} + z_{\text{残}}$$

趋势部分 $z_{\text{趋}}$ 反映了鞭角华扁叶蜂滞育幼虫空

间分布的区域背景, 是区域性因素引起的; 偏差部分 $z_{残}$ 反映了局部异常, 是局部因素引起的。除少数样本点外, 大部分样本点拟合都比较好, 在残差绝对值较大的样本点中, (0.5, 2.1)、(0.5, 5.3)、(0.5, 6.1)、(4.5, 0.5)、(4.5, 5.3)、(4.5, 6.1)、(7.7, 2.1) 这 7 个样本点的残差是正的, 而且已超过或接近剩余方差的一倍, 这些样本点内滞育幼虫聚集密度较高, 是局部因素和随机因素共同作用而引起的, 这些样本点所在的区域应引起注意。

为便于与用神经网络计算结果进行比较, 对公式 (2) 进行精度检验, 得平均模拟精度为 93.01%; 对预留的 20 组样本进行回报检验, 得平均预测精度为 93.24%。

3 小结与讨论

(1) 趋势面分析既能充分反映由大范围的系统性因素所引起的总的区域性的趋势变化规律(即区域背景), 又能反映由局部因素和随机因素所引起的局部变化趋势; 但人工神经网络比趋势面方程能更充分地利用昆虫种群空间数量分布的信息, 因此, 利用人工神经网络模型模拟昆虫种群空间分布, 能取得更令人满意的拟合效果。

(2) 神经网络的泛化能力是指神经网络对新样本的适应能力。BP 网络的缺点是必须在导师(即期望输出向量)的监督指导下进行学习、训练, 因此在样本间数据差异较小情况下, 网络不仅有很强的拟合能力, 而且有较强的泛化能力; 而在样本间数据差异较大时, 不仅需要较多的隐层神经元才能将网络训练出来, 而且网络的泛化能力较差。例如将本文预留的 20 组检验样本作为训练样本, 80 组训练样本作为检验样本, 当隐层神经元数为 7 个, 训练 10 000 步时, $MSE = 0.000\ 000\ 304\ 492$, 20 组训练样本的精度为 99.997%, 80 组检验样本的精度为 86.96%; 将表 1 中数据转回原来数据(即原来数据不作对数转换), 此时需要 60 个隐层神经元才能将网络训练出来, 80 组训练样本的精度为 55.97%, 用相关系数法检验, 其相关系数为 0.583 757, 20 组预留样本中估计精度达到 70% 以上的只有 8 组, 仅占 40%。

(3) 如果 BP 网络的输出层为 sigmoid 型神经元, 那么整个网络的输出值就限制在一个较小的范围内; 如果 BP 网络的输出层为 purelin 型神经元, 那么整个网络的输出值可以取任意值^[7]。并且理论上已经证明: 具有偏差和至少一个 S 型隐含层加上一个线性输出层的网络, 能够逼近任何有理函数^[13, 14], 也就是说只要增加隐层神经元个数, 就可以逼近任何有理函数。但在实际应用时, 为了提高模型的拟合精度, 需将数据进行变换。如上面提到的要用 60 个隐层神经元才能将没有作对数变换的样本数据训练出来; 而经过对数变换后的样本数据, 只用 12 个隐层神经元来训练 BP 网络, 就能得到满意的拟合效果。

(4) 对于用作函数逼近的 BP 网络, 隐层神经元个数与要逼近的函数的精度和函数本身的波动情况有关, 增加 BP 网络隐层数或隐层神经元个数, 可以提高网络的拟合精度, 但网络的泛化能力随之降低, 例如将本文的隐层神经元数增加到 20 个时, 网络的平均拟合精度提高到 99.55%, 而预留的 20 组样本的平均拟合精度急剧下降为 85.04%。因此, 在建立 BP 网络时, 若样本间数据差异不大时, 则应同时考虑 BP 网络的模型精度和泛化能力, 避免因网络的过拟合而带来的网络泛化能力下降问题。

参考文献:

- [1] 徐汝梅. 昆虫种群生态学[M]. 北京: 北京师范大学出版社, 1987.
- [2] 郑汉业, 夏乃斌. 森林昆虫生态学[M]. 北京: 中国林业出版社, 1995.
- [3] 周国法, 徐汝梅. 生物地理统计学[M]. 北京: 科学出版社, 1998.
- [4] 石根生, 李典谟. 不同松林马尾松毛虫蛹及其寄生天敌群子的空间格局分析[J]. 生态学报, 1997, 17(4): 386~ 392.
- [5] 李友常, 夏乃斌, 屠泉洪, 等. 杨树光肩星天牛种群空间格局的地统计学研究[J]. 生态学报, 1997, 17(4): 393~ 401.
- [6] 焦李成. 神经网络系统理论[M]. 西安: 西安电子科技大学出版社, 1996. 1~ 36.
- [7] 楼顺天, 施阳. 基于MATLAB的系统分析与设计——神经网络[M]. 西安: 西安电子科技大学出版社, 1999. 1~ 126.
- [8] 徐振邦, 娄元仁. 数学地质基础[M]. 北京: 北京大学出版社, 1994.
- [9] 米湘成, 上官铁梁, 张金屯, 等. 典范趋势面分析及其在山西省沙棘灌丛水平格局分析中的应用[J]. 生态学报, 1999, 19(6): 798~ 802.
- [10] 吴简彤, 王建华. 神经网络技术及其应用[M]. 哈尔滨: 哈尔滨工业大学出版社, 1998.
- [11] 王科俊, 王克成. 神经网络建模、预报与控制[M]. 哈尔滨: 哈尔滨工业大学出版社, 1996.
- [12] 袁曾任. 人工神经网络及其应用[M]. 北京: 清华大学出版社, 1999. 118~ 130.
- [13] 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, 1998. 59.
- [14] 闻新, 周露, 王丹力, 等. MATLAB神经网络应用设计[M]. 北京: 科学出版社, 2000. 231.
- [15] 郎奎健, 唐守正. IBM-PC系列程序集[M]. 北京: 中国林业出版社, 1989.

Study on the Spatial Quantitative Distribution of *Chinolyda flagellicornis*

CHEN Hui-hua, CUI Xiang-fu, ZHENG Yong-xiang

(Forestry Bureau of Xianju County, Zhejiang Province, Xianju 317300, Zhejiang, China)

Abstract: The artificial nerve network method and spatial tendency plane analysis method were used to study the relationship between the amount and spatial position of *Chinolyda flagellicornis* diapause larva. The BP (Back Propagation) network model and cubic tendency plane model of the spatial quantitative distribution of *C. flagellicornis* diapause larva were established. The results showed that the BP network model has high precision of fit in simulating the amount of diapause larvae with different spatial position. When the amount of crypto-neurone is 12, the mean simulation precision reaches 96.90%. The mean forecast precision of 20 groups of sample prepared in advance reaches 92.35%. The mean simulation precision of cubic tendency plane model is 93.01%, while the mean forecast precision of 20 groups of sample prepared in advance is 93.24%.

Key words: artificial nerve network; Back Propagation algorithm; diapause larva of *Chinolyda flagellicornis*; spatial structure; spatial tendency plane