

相空间重构与神经网络融合预测模型及其在害虫测报中的应用

马 飞¹, 许晓风², 张夕林³, 程遐年⁴

(1. 清华大学生物信息研究所 智能技术与系统国家重点实验室, 北京 100084; 2. 南京师范大学生命科学学院, 南京 210097; 3. 江苏省南通市植保站, 南通市 226300; 4. 南京农业大学 农业部病虫害监测与治理重点开放实验室, 南京 210095)

摘要: 针对混沌时间序列, 建立了相空间重构和 BP 神经网络融合预测模型。经实例验证, 该方法能有效地提高预测精度, 避免了一般混沌预测要求较长时间序列的缺陷, 为害虫的可预测性提供了一种新的依据。

关键词: 相空间重构; 神经网络; 预测模型; 褐飞虱

A Prediction Model Integrating Phase Space Reconstruction and Artificial Neural Network and Its Application in Pest Forecasting

MA Fei¹, XU Xiao-Feng², ZHANG Xi-Lin³, CHENG Xia-Nian⁴ (1. *State Key Lab of Intelligent Technology and System, Institute of Bioinformatics, Tsinghua University, Beijing 100084, China*; 2. *School of Life Sciences, Nanjing Normal University, Nanjing 210097, China*; 3. *Tongzhou City Station of Plant Protection, Tongzhou City 226300, China*; 4. *Key Lab. of Monitoring and Management of Plant Disease and Insects, Ministry of Agriculture, Nanjing Agricultural University, Nanjing 210095, China*). *Acta Ecologica Sinica*, 2002, 22(8): 1297~1301.

Abstract: Insect pests can significantly impact the production of many major crops. Brown planthopper, *Nilaparvata lugens* (stål), is one of the most devastating pests of rice in China. Although substantial efforts have been directed to examine factors affecting the occurrence of Brown planthopper and to predict its future outbreaks, the long-term forecasting power is still lower. We propose to use nonlinear prediction of chaotic time series for the forecasting of insect occurrence. The term “chaotic time series” here is applied in a broad sense, covering observations with complex behavior derived from underlying nonlinear phenomena of a presumed deterministic nature. Numerous complex prediction techniques, including neural networks, projection pursuit regression and radial basis functions, have previously been used for chaotic prediction. However, those methods require long time series data sets and their prediction accuracy is unsatisfactory, limiting the scope of their applications. We have built a model for predicting insect occurrence by integrating phase space reconstruction theory and multilayer backpropagation neural networks. A database of Brown Planthopper, collected by Taihu District Institute of Agricultural Science, Jiangsu Province, was used to test the accuracy of the model. Test results showed that the relative forecasting error was 9.68% and the prediction accuracy was 100%. These results demonstrate that this approach can effectively improve the prediction accuracy and does not depend on long-time series datasets for chaotic prediction.

基金项目: “973”资助项目(G20000016210); “948”资助项目(201065)

致谢: 本文得到美国北卡罗来纳州立大学农业与生命科学学院胡水金先生的指正, 特此致谢。

收稿日期: 2000-09-08; **修订日期:** 2001-08-10

作者简介: 马飞(1968-), 男, 安徽怀远人, 博士, 讲师。主要从事昆虫信息生态、害虫预测预报及生物信息研究。E-mail: bphma@tsinghua.edu.cn

Key words: phase space reconstruction; artificial neural network; prediction model; brown planthopper; *Nilaparvata lugens*(stål)

文章编号:1000-0933(2002)08-1297-05 中图分类号:Q141,Q968.1 文献标识码:A

随着混沌理论和应用技术研究的不断深入,混沌时间序列的建模和预测已成为混沌信息处理研究领域中的热点^[1~3]。然而,在实际问题中由于受到数据太少和噪声等因素的影响,使得各种不变量计算的精度无法令人满意。同时,这也是人们不断在探索的问题。进入 20 世纪 90 年代以来,非线性预测理论尤其是神经网络理论在认识时间序列混沌行为中的应用获得了重大突破^[4~5],并为这一新兴学科的深入发展开拓了新的空间。本文基于相空间重构理论和神经网络理论,提出了相空间重构与神经网络融合预测模型,并通过南通市田间三代褐飞虱高峰期虫量进行预测来检验模型的效果。

1 混沌时间序列的相空间重构

所谓混沌序列,可以看作是考察混沌系统所得到的一组随时间变化的观察量值。设时间序列 $\{x_i\}$,其时间间隔是 Δt (单位时间)。显然,吸引子的结构特性就包含在这时间序列中。为了从实际的单变量时间序列中估计出系统的信息,Packard^[6]在 1980 年提出了用时间序列重构吸引子的相空间图象。

对 n 个变量的动力系统(即一组 n 个一阶微分方程):

$$dx_i/dt = f_i(x_1, \dots, x_n) \quad i = 1, 2, \dots, n \tag{1}$$

通过消元法,使其变换为一个 n 阶非线性微分方程:

$$x^{(n)} = f(x, x^{(1)}, \dots, x^{(n-1)}) \tag{2}$$

变换后,系统的新轨迹: $X(t) = (x(t), x'(t), \dots, x^{(n-1)}(t))$ 描写了同样的动力学,它在由坐标 $x(t)$ 加上其 $(n-1)$ 阶导数 $x'(t), \dots, x^{(n-1)}(t)$ 所张成的相空间中演变。代替这种连续变量 $x(t)$ 和它的导数,可考虑不连续的时间序列和它的 $(n-1)$ 时滞位移 $X(t) = [x(t), x(t+\tau), \dots, x(t+(n-1)\tau)]$,只要将时滞 τ 选作时间序列的长时间尺度,将会保证延滞坐标线性无关。

Takens^[7]在 1981 年提出嵌入定理,表明只要相空间的维数 d 足够大(通常为 $2D_0+1$),它就可以刻画 D_0 维混沌吸引子。

设一维时间序列为: $x(t_0), x(t_1), \dots, x(t_i), \dots, x(t_n)$,将此时间序列延拓成 m 维相空间的一个相型分布:

$X(t_0)$	$X(t_1)$	$\dots$	$X(t_i)$	$\dots$	$X(t_n - (m-1)\tau)$
$X(t_0 + \tau)$	$X(t_1 + \tau)$	$\dots$	$X(t_i + \tau)$	$\dots$	$X(t_n - (m-2)\tau)$
$X(t_0 + 2\tau)$	$X(t_1 + 2\tau)$	$\dots$	$X(t_i + 2\tau)$	$\dots$	$X(t_n - (m-3)\tau)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$X(t_0 + (m-1)\tau)$	$X(t_1 + (m-1)\tau)$	$\dots$	$X(t_i + (m-1)\tau)$	$\dots$	$X(t_n)$

(3)

这里的 $\tau = k\Delta t(k=1, 2, \dots)$ 为延滞时间,(3)式中的每一列构成 m 维相空间中的一个相点,任一相点 $X(t_i)$ 有 m 个分量: $X(t_i), X(t_i + \tau), X(t_i + 2\tau), \dots, X(t_i + (m-1)\tau)$ 。 m 维空间中的一个相点表示了该系统在某个瞬时的状态,而相点的连线构成了点在相空间中的轨道,这条轨线便表示了系统状态随时间的演化。于是,在 m 维相空间中得到了一个相型,从原则上讲,便有了足够的信息把体系的动态展现在多维相空间上来考察。

2 神经网络的基本原理

人工神经网络(ANN)是一种模拟人的神经系统而建立起来的非线性动力学模型,由大量称为神经元的简单信息单元组成,每个神经元不仅从它邻近的其他神经元接受信息,也向邻近于该神经元的其他神经元发出信息。整个网络的信息处理是通过神经元之间的相互作用来完成的。神经网络具有自组织、自学习和联想记忆功能,并具有分布式、并行性及高度鲁棒性等特点^[8]。在害虫预测中,可应用人工神经网络模仿人脑的思维、学习和总结经验的过程,借助建立适当的数据结构的基础上,让人工神经网络系统进行学习、积累知识,进行预测预报。

目前在系统识别、预测预报中 90% 的应用为 BP 网络模型^[9]。该模型一般由充分连接的 3 层以上,即 1 个或多个中间(隐含)层的单元(神经元)组成,且输入层只和中间层连接,中间层再与输出层相连,每一层的各个单元之间没有连接,对于输出信号,先向前传播到中间层的各个单元,经过作用函数后,再将中间层各个单元的信息输出到输出单元,最后给出输出结果。网络的知识(或称智能)存在于基本结构神经元的连接权与阈值上。在拓朴结构确定以后,确定连接权与阈值的过程称为网络学习。对 BP 网络,学习采用有教师的误差反传算法。学习过程是将输出层误差反向逐层传播到输入层,调整连接权与阈值,如此反复进行,直到满足要求的收敛误差时为止。当学习结束后,这些连接权和阈值就确定了^[10]。根据这些参数以及多年害虫发生动态进行推理,即可对未来的害虫发生进行预测。以下简要说明 3 层 BP 网络的算法过程:

(1) 设网络的学习样本输入和期望输出为 A_k 和 $C_k (k=1, 2, \dots, m)$ 。随机地给出网络输入层到隐含层的初始连接权值为一组随机小量 V_{hi} , 同时也随机给出一组隐含层到输出层的连接权函数 W_{ij} 以及隐含层单元的阈值 θ_i 和输出层单元的阈值 γ_i 。

(2) 根据连接权矩阵(初始时刻为给定的一组随机小量)和学习样本的输入计算隐含层新的激活值:

$$b_i = f\left(\sum_{h=1}^n a_h V_{hi} + \theta_i\right) \quad (4)$$

其中, $i=1, 2, \dots, p$, 激活值函数为 Sigmoid 函数:

$$f(x) = 1/(1 + e^{-x}) \quad (5)$$

(3) 计算输出层单元的激活值:

$$C_j = f\left(\sum_{i=1}^p W_{ij} b_i + \gamma_j\right) \quad (6)$$

其中, $j=1, 2, \dots, q$ (初始时刻 W_{ij} 为给定的一组随机小量)。

(4) 计算输出层单元的一般化误差:

$$d_j = C_j(1 - C_j)(C_j^k - C_j) \quad (7)$$

其中, $j=1, 2, \dots, q$, C_j^k 为输出层单元 j 的期望输出。

(5) 计算隐含层单元相对于每个 d_j 的误差:

$$e_i = b_i(1 - b_i)\left(\sum_{j=1}^q W_{ij} d_j\right) \quad (8)$$

其中, $j=1, 2, \dots, p$ 。

(6) 调整隐含层单元到输出层单元的连接权值:

$$\Delta W_{ij} = \alpha b_i d_j \quad (9)$$

其中, $i=1, 2, \dots, p$ 和 $j=1, 2, \dots, q$, α 为学习因子 ($0 < \alpha < 1$)。

(7) 调整输出层单元的阈值:

$$\Delta \gamma_i = \alpha d_i \quad (10)$$

其中, $j=1, 2, \dots, p$ 。

(8) 调整输出层单元到隐含层单元的连接权值:

$$\Delta V_{hi} = \beta \alpha_h e_i \quad (11)$$

其中, $h=1, 2, \dots, n$ 和 $i=1, 2, \dots, p$, β 为动量因子 ($0 < \beta < 1$)。

(9) 调整隐含层单元的阈值:

$$\Delta \theta_i = \beta e_i \quad (12)$$

(10) 重复第(2)到第(9)式的计算步骤,直到 $j=1, 2, \dots, q$, 和 $k=1, 2, \dots, m$, 计算实际输出与期望输出的误差,当全部样本的输出误差小于设定的收敛误差时,训练结束。

3 相空间重构与神经网络融合预测方法建立

对于某一时间序列 $x(t_i)$ ($t=0, 1, 2, 3, \dots, n$), 若要预测 $x(t_n+1)$ 的值, 首先根据上述相空间重构理论, 按后延时间 τ 对一维时间序列 $x(t_i)$ 作漂移, 拓展成多维序列; 其次, 需要构造 BP 网络的结构形式, 然后取

$X(t_m), X(t_m + \tau), \dots, X(t_m + k\tau)$ 序列作为学习网络的期望输出, 其余序列作为网络学习的输入样本。

确定网络的输入输出模式的结构后, 便可以从已知序列值中提取样本模式构成训练集, 近而对网络进行训练, 直至达到一定小的误差要求为止。网络训练结束后便可以把最后 $m-1$ 维相空间加上网络训练结束后的期望输出作为输入模型(其中 m 为选定的输入维数序列的个数)对 $x(t_n+1)$ 的值进行预测, 从而组建了相空间重构与神经网络的融合预测模型。

4 方法的应用

利用江苏省通州市(121°E, 30°N)植保站 1976~1999 年褐飞虱三代高峰日虫量实测数据资料进行预报研究。取 1976~1996 年的实测数据作为神经网络学习的基本原始数据, 并将 1997~1999 年 3a 的褐飞虱三代高峰日虫量资料作为独立样本预报检验资料。

为进一步满足 BP 网络 Sigmoid 函数条件, 对一维褐飞虱三代高峰日虫量观察数据 $x(t_i)$ 作如下处理:

$$y_i = (x(t_i) - x_{\min}) / (x_{\max} - x_{\min})$$

式中, $x_{\max}$ 和 $x_{\min}$ 为样本序列中最小值和最大值, 这样变换后保证了 y_i 序列的值在 0~1 之间。由于本文使用的褐飞虱发生资料较短, 并没计算其混沌吸引子的维数和饱和嵌入维数值; 而是根据相空间重构理论的基本思想, 在确定预测模型的结构时, 直接分别采取 2、3、4、5、6、... 维来重构相空间, 以最好的预测结果来筛选嵌入空间维的数值。经计算机模拟训练并对其结果进行检验, 当重构 9 维相空间时, 增加误差不进一步减小, 因此, 得到嵌入空间维为 9。所以, 将变换后的一维序列, 延拓为 9 维, 表 1 给出了变换后的褐飞虱三代高峰日发生量的基于相空间重构的神经网络学习样本输入矩阵和期望输出。

表 1 标准化 BP 网络训练集

Table 1 Standardization training set of the network (BP)

学习样本输入								期望输出
Learning samples to input								Hope to output
0.2500	0.0438	0.4175	0.8076	0.8463	0.8229	0.9999	0.8905	0
0.0438	0.4175	0.8076	0.8463	0.8229	0.9999	0.8905	0	0.6037
0.4175	0.8076	0.8463	0.8229	0.9999	0.8905	0	0.6037	0.7799
0.8076	0.8463	0.8229	0.9999	0.8905	0	0.6037	0.7799	0.9729
0.8463	0.8229	0.9999	0.8905	0	0.6037	0.7799	0.9729	0.7503
0.8229	0.9999	0.8905	0	0.6037	0.7799	0.9729	0.7503	0.3279
0.9999	0.8905	0	0.6037	0.7799	0.9729	0.7503	0.3279	0.4947
0.8905	0	0.6037	0.7799	0.9729	0.7503	0.3279	0.4947	0.7935
0	0.6037	0.7799	0.9729	0.7503	0.3279	0.4947	0.7935	0.3742
0.6037	0.7799	0.9729	0.7503	0.3279	0.4947	0.7935	0.3742	0.5672
0.7799	0.9729	0.7503	0.3279	0.4947	0.7935	0.3742	0.5672	0.3173
0.9729	0.7503	0.3279	0.4947	0.7935	0.3742	0.5672	0.3173	0.3541
0.7503	0.3279	0.4947	0.7935	0.3742	0.5672	0.3173	0.3541	0.4805

采用的 3 层 BP 网络的输入层为 8 个节点, 隐含层为 9 个节点, 输出层节点为 1。取定网络学习的收敛误差为 0.00005, 学习因子和动量因子分别取为 0.2 和 0.95。经过 6925 次迭代, 收敛误差达到标准, 学习结束。网络学习的拟合值和实测值变化见图 1。由图 1 可以看出, 网络经过学习后拟合效果具有很高的精度。然后, 根据训练好的网络参数, 对未参加学习的 1997~1999 年 3a 的样本进行预报, 并对这 3a 的独立样本预报结果进行检验, 确定其预报精度。并根据当地褐飞虱发生程度等级划分标准: 百穴虫量 < 500 头, 轻发生; 百穴虫量在 500~1000 头, 偏轻发生; 百穴虫量在 1000~2000 头, 中等发生; 百穴虫量在 2000~3000 头, 偏重发生; 百穴虫量 > 3000 头, 大发生, 对褐飞虱发生程度进行预报。表 2 给出了 1997~1999 年褐飞虱三代高峰日

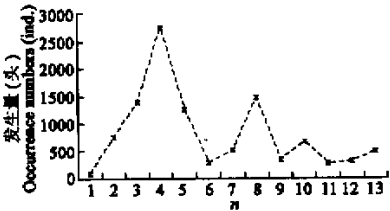


图 1 3 层 BP 网络期望输出(虚线)和实际输出(*)

Fig. 1 The expectance export (broken lines) and real export (*) of the network (BP)

发生量的预报结果和预报结果检验。由表 2 可以得到,预报精度是令人满意的,^{3a} 的预报平均相对误差为 9.68%,其中最大相对误差为 13.51%,最小相对误差为 4.41%,与实际发生程度相符,预报准确率达 100%。

表 2 1997~1999 年褐飞虱三代高峰日发生量的预报结果检验

Table 1 The test for the forecast results of BPH population occurrence numbers at generation three between 1997 and 1999

年份 Year	实测数 Observed data	预报值 Forecast value	相对误差(%) Relative error	绝对误差 Absolute error	转化为发生程度 Simulated occurrence degree	实际发生程度 Observed occurrence degree	检验 Test
1997	827.27	791.2	4.41	56.5	偏轻 ^①	偏轻	✓
1998	374	415.6	-11.12	-41.6	轻 ^②	轻	✓
1999	248	281.5	-13.51	-33.5	轻	轻	✓
2000		292.5					

①Lighter occurrence ②Light occurrence

5 结论与讨论

5.1 本文基于混沌动力系统相空间重构理论和神经网络理论相结合,提出了一种相空间重构与神经网络的融合预测模型,通过对褐飞虱三代高峰日虫量实测数据资料进行预报,实验结果表明,其平均相对误差为 9.68%,预测准确率 100%。这一结果为混沌序列的可预测性及预测模型的选择提供了一种新依据。

5.2 相空间重构与神经网络的融合预测模型能够有效地预测低维混沌时间序列的本质在于:一维时间序列中包含着吸引子的结构,而重构相空间能够估计出系统演化的信息,由于把一维时间序列拓展为多维序列,而多维序列包含着各态的信息,所以使预测更符合实际,同时神经网络所具有的特征使之适于处理信息不全或推理规则不明确的问题,消除了混沌系统对初值的敏感性,而模型的权值是通过实例训练学习得出,避免了某些方法的人为计取权值的主观影响。采用相空间重构与神经网络融合预测模型对害虫进行预测的另一优点是它便于硬件实现,而且克服了一般混沌预测理论要求较长时间序列的缺陷,通用性较好,十分适合于业务预报推广。

5.3 在研究过程中发现,³层 BP 网络方法虽然是一种处理非线性复杂问题的有力工具,但 BP 算法的学习训练过程中收敛速度较慢,而且在网络训练时,有时会在极小点附近产生振荡导致不收敛等问题。为此,作者正在考虑采用进化的遗传算法、小波变换算法和径向基函数算法对上述问题作进一步的深入研究。

参考文献

[1] Zhang J S(张家树),Xiao X C(肖先赐). Predicting low dimensional chaotic time series using Volterra adaptive filers. *Acta Physica Sinica*(in Chinese)(物理学报), 2000, **49**(3):403~407.

[2] Wang M J(王明进), Cheng Q S(程乾生). Application of the Kohonen self-organization networks in chaotic time series prediction. *Systems Engineering—Theory & Practice*(in Chinese)(系统工程理论与实践), 1997, **7**(2):12~17.

[3] Cheng Q S(程乾生), Wu L W(武连文), Wang S Z(王守章). Integrating prediction based on property clustering networks and radial basis function. *Chinese Science Bulletin*(in Chinese)(科学通报), 2000, **45**(11):1211~1216.

[4] Gencay R. Nonlinear prediction of noisy time series with feedforward networks. *Phys. Lett. A.*, 1994, **187**:397~403.

[5] Casdagli M. Nonlinear prediction of chaotic time series. *Physica D*, 1989, **35**:335~356.

[6] Packard N H, Crutchfield JP, Farmer JD, et al. Geometry from a time series. *Phys. Rev. Letts.*, 1980, **45**(9):712~716.

[7] Tankens F. Detecting strange attractors in turbulence. In: Rand DA, Young L S, eds. *Dynamical Systems and Tubulence. Lecture Notes in Math*, 898. Berlin; Springer,1981. 366~381.

[8] Pao, Yoh-han (包约翰). *Adaptive pattern recognition and neural networks*(in Chinese). Beijing; Science Press, 1992. 103.

[9] Zhang L M(张立明). *Artifical neural networks models and its application* (in Chinese). Shanghai: Fudan university press, 1993. 10~51.

[10] He X R(何小荣), Chen B Z(陈丙珍). Study on improving testing results of BP neural networks. *Journal of Tsinghua University* (Sci & Tech) (in Chinese)(清华大学学报(自然科学版)), 1995, **35**(3):31~36.