

神经网络模式在流量之预测

颜正平* 曾仁宏

(中兴大学水土保持研究所·台湾台中市)

摘要 本文以倒传递神经网络模式,进行长期流量预测之应用研究,并以台湾中部德基水库上游松茂流量站历年观测资料进行探讨,初步得到预测结果与实测值比较尚有一倍以上的误差。但以神经网络之潜在应用潜力,寻找系统化网路参数值建立方法以得到更精确预测结果之经验,为吾人相当值得继续研究之课题。

关键词 倒传递神经网络模式 最陡坡降法 双弯曲函数 流量预测

The Prediction of Flow Using Neural Network Model

Cheng—ping Yen Jen—hun Chen

(Institute of Soil and Water Conservation, Chung Hsing University, Taichung Taiwan)

Abstract This paper uses Back—Propagation Neural Network Model to proceed the research of long—term flow prediction. The observed flow data of Son—Mao station on the upstream of De—Ji Reservoir of central Taiwan are used. The primer result is the over prediction nearly 100 % with observed data. But for future potential of Neural Network application, the experience of searching the systematized method of network parameters construction, and to get more accurate prediction results, is worth to study.

Key words back—propagation neural network model the gradient steepest descent method sigmoid function flow prediction

1 前言

仿生学的类神经网络(Neural Network)研究,近年来在应用研究上有了乐观的成果,也为传统线性模式无法处理的问题,开拓出一条崭新的研究方向来。由混沌理论所展现出来的结果,我们了解非线性是一切复杂性之源,由於非线性的作用,才孕育出大自然的万千气象。神经网络模式,正为这种自适应的非线性动力系统,提出一种解决的演算法来。

向人类自身大脑学习的类神经网络,其定义为“是一种平行计算系统,包括软体与硬体,它使用大量简单的相连人工神经元来模仿生物神经网络的能力”。类神经网络是人们在対神经系统不断加深理解的基础上,对神经网路的原理、模型和应用进行研究的一门学科。人脑神经讯息活动的特点与传统冯·诺伊曼(Von Neumann)电脑的差别在传统电脑原理是利用人类的三个逻辑观

① 收稿日期 1995—05—10

※ 中兴大学水土保持学研究所所长兼系主任。

念顺序、回圈与判断的逻辑规则进行运算的,它有极强的算术和逻辑运算能力,但是其形象思维之辨识能力与高等动物赖以生存及适应环境之学习能力却相当薄弱。神经网络正可弥补此不足,其特征为讯息巨量平行处理,讯息处理和储存单元合在一起的物件导向式,具有自我组织自我学习的能力,能接受和处理类比的、模糊的和随机的讯息之容错性,求满意解而不是求精确解之功能之一大型自适应的非线性动力系统^[1]。

类神经网络的应用在分类、诊断与预测上已有相当的成果。流量预测在台湾水资源渐感不足的情况下,逐渐显出其需要性来。过去流量的预测由早期的线性回归模式到考虑时间数列模式其残差含有序列相关关系,即本期之不稳定项受前期不稳定项遗传效应影响之自我回归移动平均整合模式(简称 ARIMA 模式),但其仍为线性模式^[2]。本文以目前类神经网络学习模式中,最具代表性、应用最普遍的非线性倒传递类神经网络模式(Back-Propagation Neural Network Model)进行长期流量预测之应用研究,并以台湾中部德基水库上游之松茂流量站历年月流量记录资料进行探讨,以了解掌握神经网络模式在流量预测等相关自然科学应用之潜在应用潜力与方向。

2 材料与方法

2.1 试验集水区概况

德基水库位于台湾中部大甲溪上游,集水区面积602 km²,有效蓄水容量1.73亿 m³,系台湾中部地区200万人口公共给水主要水源。水库集水区保安林占96%,余为果园、蔬菜、道路等开垦地。而松茂流量站为德基水库上游主要入流量观测站,并提供枯水期水库进水量的主要水源,其上游集水区面积417.08km²,占德基水库集水区面积69%,德基水库集水区分布参图1^[3]。

2.2 试验方法

使用美国史丹福大学教授 David Rumelhart 及 G. E. Hinton 与 R. J. Williams 等于1985年为改善感知机模式(1957)缺乏隐藏层及相关演算法无法解决互斥(XOR)问题而发展出来的倒传递神经网络模式(Back-Propagation Network)法进行分析。并以美国 NeuralWare 公司1993年发表的 NeuralWorks Professional I /PLUS 软体进行数据实作模拟。^{[4][5]}

3 理论基础及演算法^{[6][7]}

倒传递类神经网络(Back-Propagation Neural Network)是一种具有监督式学习能力的多层前授型网路,使用误差逆向传播演算法实现了自我学习。此种神经网络是由 Rumelhart 等在1985年所提出的。其基本原理是利用最陡坡降法(the gradient steepest descent method)的观念,

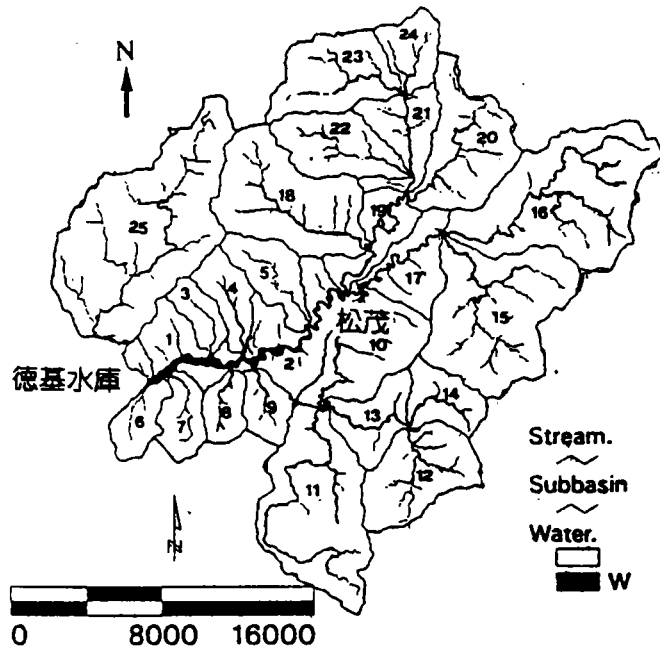


图1 水库集水区分布图

将误差函数予以最小化。其含有隐藏层,使得网路可表现输入处理单元间的相互影响;使用平滑可微分的转换函数,如非线性双弯曲函数(sigmoid function),使得网路可应用最陡坡降法导出修正网路加权值的公式。

倒传递网路的网路架构参图2,包括用以表现网路的输入变数的输入层,用以表现输入处理单元间的交互影响的隐藏层(可不只一层),及用以表现网路输出变数的输出层。其同一层中的神经元彼此并不相连,而不同层间的神经元则彼此相连,且信号的流向是由输入层向输出层单向传播。

倒传递演算法乃应用一个训练范例的一输入值向量 X ,与一目标输出向量 T ,修正网路加权值 W 。而达到学习的目的。其理论基础叙述如下:

在倒传递网路中,第 n 层的第 j 个单元的输值 A_j^n 为第 $n-1$ 层单元输出值 A_j^{n-1} 的非线性函数:

$$A_j^n = f(\text{net}_j^n) \quad (1)$$

式中, net_j^n 为集成函数 $= \sum_i W_{ij} A_i^{n-1} - \theta_j$, f 为转换函数,一般为非线性双弯曲转换函数,则为

$$f(x) = \frac{1}{1 + e^{-x}}. W_{ij} \text{ 为介於第 } n-1 \text{ 层的第 } i \text{ 个}$$

处理单元,与第 n 层的第 j 个处理单元间的连结加权值。 θ_j 为内部临界值或称偏权值。

因为监督式学习旨在降低网路输出单元目标输出值 T 与推论输出值 A 之差距,所以一般以下列能量函数(或称误差函数)表示学习的品质:

$$E = \frac{1}{2} \sum_j (T_j - A_j)^2 \quad (2)$$

式中, T_j 输出层目标输出值; A_j 输出层推论输出值。

网路的学习过程乃在使上述能量函数最小化的过程,通常以最陡坡降法来使能量函数最小化,即每当输入一个训练范例,网路即小幅调整加权值的大小,调整的幅度和误差函数对该加权值的敏感程度成正比,即

$$\Delta W_{ij} = \eta \cdot \delta_j^n \cdot A_i^{n-1} \quad (3)$$

式中, η 为学习速率,控制每次以最陡坡降法最小化误差函数的步幅。 $A_i^{n-1} = W_{ij}$ 所连接之较低层的处理单元之输出值。 $\delta_j^n = W_{ij}$ 所连接之较上层的处理单元之差距量,如转换函数使用双弯曲函数则为

$$= \begin{cases} \text{如果 } W_{ij} \text{ 所连接之较上层的处理单元为输出层} \\ \text{则 } \delta_j^n = (T_j - Y_j) \cdot Y_j \cdot (1 - Y_j) \\ \text{否则 } \delta_j^n = \left[\sum_k \delta_k^{n-1} \cdot W_{jk} \right] \cdot H_j \cdot (1 - H_j) \end{cases}$$

式中, Y 为输出层处理单元的输值; H 为隐藏层处理单元的输值。

同理可得到

$$\Delta \theta_j = -\eta \cdot \delta_j^n \quad (4)$$

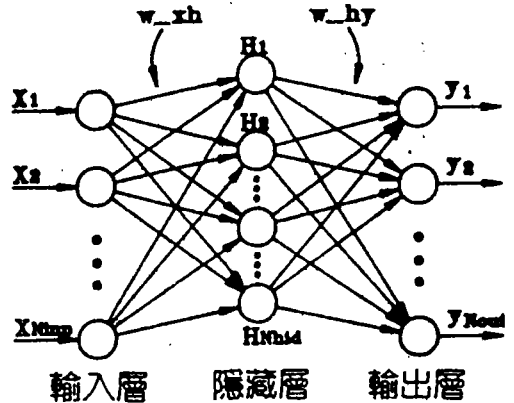


图2 倒传递网路架构

倒传递网路的运作过程分为学习与回想两部份,其学习完后固定其网路连结之加权值及偏权值,再用于回想以进行分类或预测的应用。学习过程在修正网路加权值及偏权值之过程,是将目标输出值与网路的推论输出值之间的误差,一边向后传播,一边加以修正,由于这种特性,故称倒传递。兹将学习及回想两部份之演算步骤整理如下:

3.1 学习演算法

倒传递网路学习演算法共分为 8 个步骤,说明如下:

1. 设定网路的层数及各层间神经元数目等参数。
2. 以均布随机乱数设定网路的初始加权值及初始偏权值。
3. 输入训练样本及目标输出值。
4. 计算网路的推论输出值。
5. 计算输出层与隐藏层的差距量。
6. 计算各层间的加权值修正量及偏权值修正量。
7. 更新各层间的加权值及偏权值。
8. 重复步骤3至步骤7,直到网路收敛(误差不再有明显变化),或执行一定数目的学习循环。

3.2 回想演算法

倒传递网路回想演算法分 4 个步骤,说明如下:

1. 设定网路的层数及各层间神经元数目等参数。
2. 读入已训练好的网路加权值及偏权值。
3. 输入一个测试范例输入向量。
4. 计算网路的推论输出值。

4 结果与讨论

本文以台湾中部德基水库上游松茂流量站的历年(自1971年1月至1993年12月共23年 276个资料)月流量记录进行数据模拟,松茂站之历年月流量资料历线参图3。流量预测使用监督式之倒传递神经网络进行演算预测。以月流量前五年之60个月流量资料当输入值以预测后面6个月之月流量输出值,如此为一组资料。文内以前20年之连续资料共15组当训练组资料进行网路学习,以建立网路之加权值及偏权值;并以最后3组资料(即1991至1993年)当回想值以进行网路执行结果之比较与验证。并分别对起始预测月份由1月至7月分别进行演算,以评估各起始预测月份之情形。

NeuralWork Professional II /PLUS 软体进行网路建立的基本步骤如下^[8]:

- (1) 建立网路:用以设定网路的网路种类、层数、处理单元数、网路连接及网路参数。
- (2) 设定仪表:用以设定监视网路学习过程的绘图视窗。
- (3) 设定网路输出输入(IO/Parameters):用以设定网路的输出,如学习案例来源等。
- (4) 网路初始化(Run/Initialize):用以将网路连结加权值初始化为随机乱数。
- (5) 网路学习(Run/Learn):用以将学习案例交由网路学习,调整网路连结加权值。
- (6) 网路测试(Run/Test):用以将测试案例交由网路测试,探讨网路精确性。
- (7) File/Save: 储存网路架构,以便以后之应用。

由该软体以学习至误差稳定约50次循环,学习法则使用 Delta—Rule,并以双弯曲函数(sigmoid function)当转换函数执行,得到各起始预测月份之未来6个月的预测结果。而预测结果与

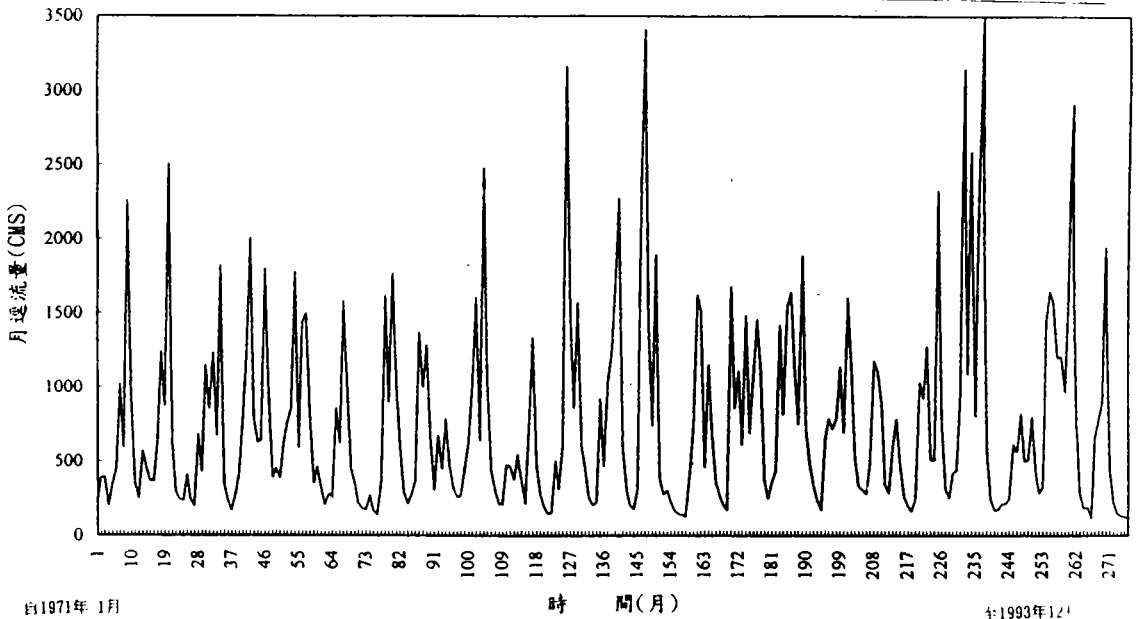


图3 松茂站之历年月径流量(CMS)

实测值之比较,一般系依预测误差来比较之,假设超前 t 期之预测误差定义为⁽⁹⁾:

$$e' = T_{b+t} - A_{b+t} \tag{5}$$

式中, b 为起始点, T_{b+t} 及 A_{b+t} 分别代表观测值及模式预测值。

通常被应用之误差统计量有下列四种:

(1) 平均误差率(Mean Percentage Error, 简称 MPE)

$$MPE = \left(\frac{1}{M} \sum_{t=1}^M \frac{e_t}{T_{b+t}} \right) 100\% \tag{6}$$

(2) 均方误差(Mean Square Error, 简称 MSE)

$$MSE = \frac{1}{M} \sum_{t=1}^M e_t^2 \tag{7}$$

(3) 平均绝对值误差(Mean Absolute Error, 简称 MAE)

$$MAE = \frac{1}{M} \sum_{t=1}^M |e_t| \tag{8}$$

(4) 平均绝对值误差率(Mean Absolute Percentage Error, 简称 MAPE)

$$MAPE = \left(\frac{1}{M} \sum_{t=1}^M \left| \frac{e_t}{T_{b+t}} \right| \right) 100\% \tag{9}$$

松茂流量站使用倒传递网路模式预测结果与实测值比较之预测误差参表1。

表1 松茂流量站使用倒传递网路模式之预测误差

起始预测月份 误差统计量类别	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)	7 (%)
MPE	-178.29	-133.28	-108.62	-154.19	-106.5	-106.82	-109.41
MSE	1117715	835542.2	626627	1102145	689195.1	667445.5	497101.7
MAE	871.58	774.35	624.84	833.75	636.46	616.19	474.44
MAPE	211.41	169.05	130.25	177.66	131.26	133.57	133.68

由表1知,初步使用倒传递神经网络模式在松茂站流量之预测,结果尚有一倍以上的误差,且预测值普遍偏低。检讨其原因为,在使用倒传递网路模式时,有以下几个缺点尚待克服:1. 使用不

同的初始加权值及偏权值会导致不同的收敛结果,即收敛於一局部最小误差值,需以不同的初始值测试以找出一合理的最小误差值;2. 网路架构及网路参数的决定,如隐藏层神经元数目,学习速率等参数的决定,尚缺乏一系统化的方法来遵循,须以试误法为之;3. 网路在学习新样本时有遗忘已学习过样本的趋势;4. 学习速度较慢,造成以试误法寻找合理参数值时之繁琐与困难。另因观测资料较少,用以建立训练网路之学习数目只有15组或许尚不足以代表全局的行为趋势;且因全球气候的变迁及集水区局部地区的开发,亦可能对其后续的流量有所影响,而难以掌握。而在混沌理论中,起始值微小的变化在离散叠代多次之后就有可能产生南辕北辙的结果之蝴蝶效应,故气象学家认为长期预测不可行。本流量站是否符合此混沌行为,亦值得吾人加以探讨;但在本研究中,则尚无明显趋势有起始预测误差优于后面预测误差之情况。而原始资料的随机现象,亦自然使预测结果产生随机之预测结果。

5 结 论

大脑的研究过去因不知从何着手,有如一个黑洞而无法探索。近几年来,因先进国家投入大量的人力进行研究而有了初步基本的了解与掌握,并为进步国家积极推动的研究工作,美国并将90年代称之为"Decade of the Brain",确立了在国家级别上推进脑研究的方针。

本文以目前类神经网络应用最成熟之倒传递神经网络模式,初步研究在流量预测之应用,并以德基水库松茂流量站历年观测资料为例进行探讨,得到使用倒传递网路模式预测尚有一倍以上之误差。但以神经网络潜在的巨大应用潜力及智慧之源的探秘,寻找系统化网路参数值建立之经验为吾人值得继续深入探讨之课题。

参考文献

- 1 焦李成. 神经网络系统理论. 儒林图书有限公司,1991,P14
- 2 曾仁宏、颜正平. 介入时度间数列在坡地开发之应用. 1994年农业工程研讨会论文集. 1994,P387~400.
- 3 萧国鑫等. 德基水库集水区土地利用变迁分析及资讯系统之建立总报告. 工研院能资所,1994,P78
- 4 Using Neural Works — A Tutorial for Neural Works Professional I /PLUS and Neural Works Explorer. Neural Ware, Inc. ,1993
- 5 Reference Guide — Software Reference for Professional I /PLUS and Neural Works Explorer. Neural Ware, Inc. , 1993
- 6 王进德、萧大全. 类神经网络与模糊控制理论入门:全华科技图书股份有限公司,1994,P15~58
- 7 叶怡成. 类神经网络模式应用与实作. 儒林图书有限公司,1993,P69~110
- 8 叶怡成. 类神经网络系统 Professional I Plus 训练教材:中纲科技股份有限公司,1993
- 9 林茂文. 时间数列分析与预测:华泰书局,1992,p321
- 10 Nachimuthu Karunanithi etc. NEURAL NETWORKS FOR RIVER FLOWPREDICTION. Journal of Computing in Civil Engineering, Vol. 8,No. 2,ASCE,1994,P201~220