短文

# 基于多 ANN 模型的复杂系统长时段预报

高　峰　　李人厚
（西安交通大学系统工程研究所　西安　710049）

**摘要**　提出一种多 ANN 结构的极值聚类训练算法，并将这种方法应用于复杂系统长时段预报. 采用这种方法，可以提高长时段预报精度、增强模型的可靠性. 以这种模型为基础可以进一步建立基于多 ANN 模型的复杂系统预测控制.

**关键词**　多 ANN 结构，多层前馈网络，非线性系统预报.

## 1　引　言

多 ANN 结构，是由多个相对独立、相互连接、协同作用的子 ANN 或称 ANN 模块组成的系统. 它是一种新型的神经元网络系统，是在更高层次上对人脑结构的模拟. 人脑正是由于这种层次化、模块化的组织形式才产生出复杂、高级的智能行为.

用多 ANN 结构的方法可以得到高度可靠的神经元网络系统，这种思想曾被许多 ANN 研究者用来解决复杂问题[1-4]. 为了解决复杂函数的逼近问题，本文作者也曾采用了多 ANN 方法[5]，极大地改善了 ANN 的逼近能力.

## 2　多 ANN 结构

多 ANN 结构如图1所示，内部通常包含两种类型的网络：一是实现某种映射关系的子网络，二是协调各子网络之间关系的门控网络. 一般在一个多 ANN 结构中，有多个子网络，而仅有一个门控网络. 各子网络具有相同的输入但生成各自独立的输出，而整体的输出是门控网络对各子网络输出值的加权和：

$$y = \sum_{s=1}^{n} g_s \cdot y_s, \qquad (1)$$

其中 $y_s$ 表示第 $s$ 个子网络的输出值. 对于多 ANN 结构来说子网络的结构

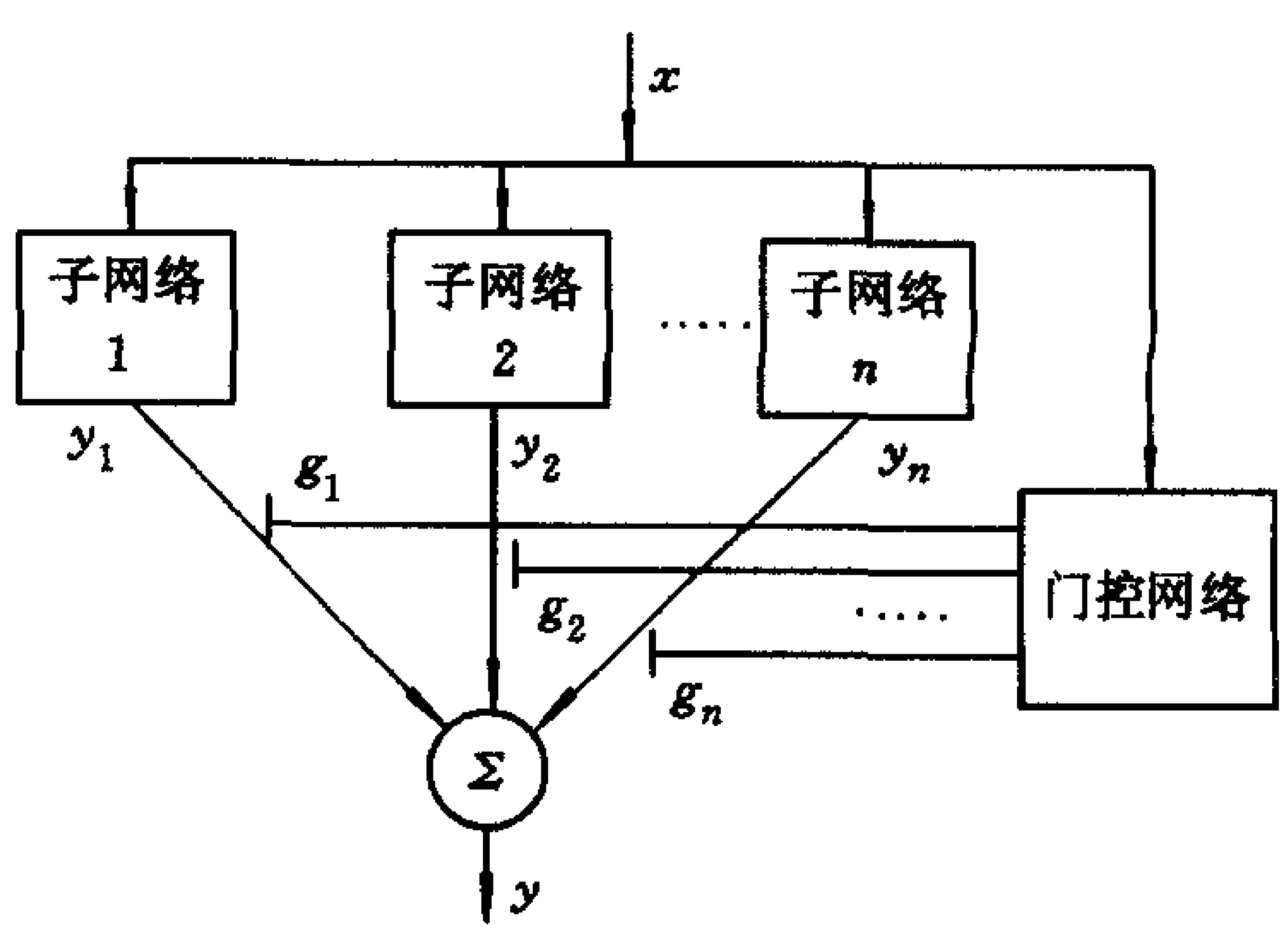



图1　多 ANN 结构示意图

收稿日期　1996-01-29

形式可以是任意的,MFNN, RBF,FLN,CMAC 等都是可取的,只要它们具有良好的对非线性映射的逼近特性. 这里为了使问题简化,假设各子网络一律采用三层前馈网络的形式,用式

$$y_s = W_s^2\sigma(W_s^1x + B_s^1) + B_s^2 \tag{2}$$

统一描述. 门控网络的输入与各子网络的输入相同,对应于每一个子网络有一个输出量,(1)式中 $g_S$ 表示其第 $s$ 个输出. 门控网络的作用是分配各子网络在整体映射关系中的角色. 根据当前时刻的输入值,按照某种原则产生输出,即各子网络输出的加权值,进而决定对各子网络的选择性. 门控网络的特性决定了整个多 ANN 结构的工作情况.

# 3　基于极值聚类的多 ANN 训练算法

## 3.1　极值聚类算法

现有的多 ANN 结构聚类实现方法,往往只是对输入空间进行聚类,没有考虑样本的输出特征,得到的划分仅仅体现了样本输入的分布关系,不能体现出函数映射本身的简单性或复杂性.

通过对 MFNN 的分析可以知道,MFNN 实现的映射是由各隐节点决定的一组饱和上升曲面加权叠加. 划分子区域必须使各子区域中函数易于用这种曲面叠加形式来逼近,这就是 MFNN 对函数简单性的要求. 从这一特点出发,如果划分使得各区域中函数都是单极值的,那么在各区域中用子网络逼近时,隐节点的方向几乎都是指向极值点或与此相反,隐节点之间的耦合现象非常少,逼近比较容易.

为得到这种划分形式,就要对样本集合进行极值聚类. 对此提出一种简捷而有效的搜索样本集局部极值点的方法,可以对样本集中蕴涵的函数关系进行极值聚类.

设样本集合为$\{(x_i,y_i),i=1,2,\cdots P\}$,则极值聚类算法如下:

1) 设置初始值. 确定样本划分类数 $K$,选择 $K$ 个初始聚类中心.

2) 按样本输入值的邻近关系,确定样本集中各样本的归属关系. 对于第 $i$ 个样本$(x_i,y_i)$,如果$\|x_i-c_k\|<\|x_i-c_j\|$, $j=1,2,\cdots,K$, $C_k$ 和 $C_j$ 为第 $k$ 和第 $j$ 个样本子集的聚类中心,则$(x_i,y_i)\in D_k$,$D_k$ 表示第 $k$ 个样本聚类中心的子集.

3) 按下面的关系更新聚类中心:搜索局部极大值时,对第 $k$ 个聚类,如果有$(x_i,y_i)\in D_k$,满足对$\forall\ (x_j,y_j)\in D_k$,有 $y_j\leq y_i$,则选择$(x_i,y_i)$作为新的第 $k$ 个聚类的聚类中心. 即取聚类域中具有最大的样本输出值的样本作为新的聚类中心. 同理,搜索局部极小值时,取聚类域中具有最小的样本输出值的样本作为新的聚类中心.

4) 检查聚类中心是否发生了变化:是,则返回2.

5) 检查聚类中心是否临近,即是否存在$\|c_i-c_k\|<\varepsilon$,$(\varepsilon>0)$,是则归并最相邻的两类,然后返回2;否则算法结束.

以这一算法为基础,可以这样完成对训练集的划分:先求出样本集的局部极大值中心,再求出局部极小值中心,将所求出的中心合并为样本集极值中心集,再按邻近原则(即算法第2步)重新划分样本区域.

极值聚类与其它多 ANN 结构中子网络功能划分和协作方法不同的是,不仅利用了样本集的输入,同时利用了样本集的输出,进行的是针对函数关系的聚类. 通过划分降低

了函数的复杂性，所确定的子任务是子网络易于实现的，具有明确的几何意义，子网络的数目也是由问题的复杂性所决定的.

**3.2 多 ANN 结构的极值聚类训练算法**

由此可以得到基于极值聚类的多 ANN 训练算法：

1）对整体样本集按极值聚类算法进行划分，选择初始分类数 $K$=样本总数/50，经分类及归并后可以得到 $K$ 个极值聚类中心 $c_k, k=1,2,\cdots,K$. 由邻近原则确定各子类的子样本集 $D_k$；

2）初始化各子网络 $ANN_k$，在各训练子集 $D_k$ 内，测试初始逼近误差 $E_k$；

3）找出逼近误差最大的子网络 $ANN_k$，在相应的训练子集 $D_k$ 内，按某种训练算法的训练 $T$ 步(这里取 $T=50$)；

4）检查是否所有 $E_k$ 均小于规定误差限，是则训练结束，否则返回3.

在网络工作阶段，只需对当前输入值由邻近原则确定其所属类，激活该类对应的各子网络得到网络的输出值，并以此作为整体网络输出值.

因为，MFNN 对一个紧集上的任意连续函数具有任意逼近性，每个子网络可以任意逼近由子样本集所反映的函数关系，这样由这些子网络的共同作用就可以得到对整体样本集的任意逼近.

# 4 基于多 ANN 结构的复杂系统长时段预报

多 ANN 模型中，可以认为门控网络与各子网络是在不同精度上对系统的描述. 门控网络完成的是对系统输入的模式分类，是对系统的一种粗线条的、定性的描述. 各子网络是对各输入模式类的输出估计，是对系统的一种定量的描述. 因此，可将门控网络的训练与各子网络的训练分开，前者采用极值聚类算法离线进行. 这里提出的算法是在多 ANN 长时段预报模型中对各子网络的修正算法.

对一个实际对象，可以量测到一组实际的输入—输出数据 $\{u(t), y_p(t) \mid t=1,2,\cdots,T\}$. 通过适当的实验设计，它可以在相当大的程度上反映出我们所关心的对象性态. 根据这组数据 ，可以离线建立对象的长时段预报模型.

这里以并联连接的 NARMA 模型作为长时段预报模型，以 ANN 模型描述如下：

$$\begin{cases} y_m(t) = \sum_{s=1}^{S} g_s(X(t)) \cdot y_s(X(t), \theta_s) \\ X(t) = [y_m(t-1)^T, \cdots, y_m(t-K_y)^T, u(t-1)^T, \cdots, u(t-K_u)^T]^T \\ \quad\;\; = [X_1(t)^T, \cdots, X_{K_y}(t)^T, X_{K_y+1}(t)^T, \cdots, X_{K_y+K_u}(t)^T]^T, \end{cases} \tag{3}$$

其中 gs 表示门控网络的第 s 个输出量，ys 表示第 s 个子网络的输出值，网络形式如第2节所述.

在训练时，1到 P 步输出预报模型辨识的指标函数定义为

$$\begin{aligned} J(P) &= \frac{1}{2} \sum_{t=1}^{P} \sum_{s=1}^{S} g_s^2(X(t)) \cdot \| y_p(t) - y_s(X(t), \theta_s) \|^2 \\ &= \frac{1}{2} \sum_{t=1}^{P} \sum_{s=1}^{S} g_s^2(X(t)) \cdot \| \varepsilon_s(t) \|^2, \end{aligned} \tag{4}$$

其中 $\varepsilon_s(t)=y_p(t)-y_s(X(t),\theta_s)$ 是 $t$ 时刻的第 $s$ 个子网络的估计误差. $J(P)$ 反映了各子网络1到 $P$ 步输出预报的总误差. $J(P)$ 的定义是假定每一个子网络要逼近的仍然是对象的整体特性，只是在响应一个输入时对各子网络的信任度(由 $g_s$ 决定)是不同的.

设已知模型输出初始值 $y_m(-K_y+1),y_m(-K_y+2),\cdots,y_m(0)$，控制输入 $u(-K_u+1),\cdots,u(0)$，$u(1),\cdots,u(P-1)$，和期望输出轨线 $y_p(1),y_p(2),\cdots,y_p(P)$. 长时段预报模型的离线训练问题可描述为：在给定的模型结构(3)下，确定模型参数 $\theta=(\theta_1^T,\theta_1^T,\cdots,\theta_S^T)^T$，使式(4)的指标函数最小.

在给出训练算法前，首先需求出目标函数对各子网络参数的导数

$$\frac{\partial J(P)}{\partial\theta}=\sum_{t=1}^{P}\sum_{s=1}^{S}\left(-g_s(X(t))^2\cdot\left(\frac{\partial^+ y_s(X(t),\theta_s)}{\partial\theta^T}\right)^T\cdot\varepsilon_s(t)\right), \tag{5}$$

其中 $\frac{\partial^+}{\partial\theta}$ 称为有序偏导(ordered derivatives)[6]，表示对变量 $\theta$ 考虑各种可能存在的复合函数关系之后求得的偏导数. (5)式中的有序偏导可由下面的式子计算得到：

$$\frac{\partial^+ y_s(X(t),\theta_s)}{\partial\theta^T}=\frac{\partial y_s(X(t),\theta_s)}{\partial\theta^T}+\sum_{i=1}^{K_y}\left(\frac{\partial y_s(X(t),\theta_s)}{\partial X_i(t)^T}\cdot\frac{\partial^+ y_m(t-i)}{\partial\theta^T}\right). \tag{6}$$

而网络整体输出对参数的灵敏度方程为

$$\frac{\partial^+ y_m(t)}{\partial\theta^T}=\sum_{s=1}^{S}\left(g_s(X(t))\frac{\partial^+ y_s(X(t),\theta_s)}{\partial\theta^T}\right). \tag{7}$$

从这几个导数计算公式中可以看出，当前时刻输出对模型参数的偏导数，通过网络输入连接依赖于前 $K_y$ 步的输出对参数的偏导数. 换言之当前时刻的误差会通过模型输入连接，反向传播到网络工作的过去时刻中. 这就是沿时间的反向传播的思想. 其中式(6)反应了通过第 $s$ 个子网络的反向传播.

需要指出的是，在极值聚类多 ANN 模型内，还存在一条通过门控网络的反馈通路. 但由于门控网络的输出值是通过求极值得到的，每一时刻 $g_k$ 中只有一个是1，其余全部是0，所以门控网络的输出对输入是不可导的. 在假设 X(t) 只有微小的变化时不影响 $g_k$ 的取值的前提下，可认为 $g_k$ 对输入的导数为0. 因此对极值聚类的多 ANN 模型只考虑误差通过各子网络的反向传播，不考虑通过门控网络的反向传播.

结合 MFNN 的反向传播算法，可得到如下的多 ANN 长时段预报模型离线辨识算法. 在第 θ－1次更新之后，计算第 θ 次权值更新

1. 初始化：$\frac{\partial^+ y_m(0)}{\partial\theta^T}=\cdots=\frac{\partial^+ y_m(-K_y+1)}{\partial\theta^T}=0$；

2. 对 $t=1,2,\cdots,P$，计算

a) 样本正向、反向传播. 由门控网络计算出 $t$ 时刻各子网络的输出加权 $g_S(t)$，对被激活子网络由(6)式计算 $\frac{\partial^+ y_S(X(t),\theta_S)}{\partial\theta^T}$；

b) 计算累计梯度值；

c) $G(t)=G(t-1)+\sum_{s=1}^{S}\left(-g_s(X(t))\cdot\left(\frac{\partial^+ y_s(X(t),\theta_s)}{\partial\theta^T}\right)^T\cdot\varepsilon_s(t)\right)$；

d) 由式(7)更新 $\frac{\partial^+ y_m(t)}{\partial\theta^T}$.

3. 采用某种方法更新各子网络权值阵. 可以选用 BP 型更新公式. 这里采用的是基

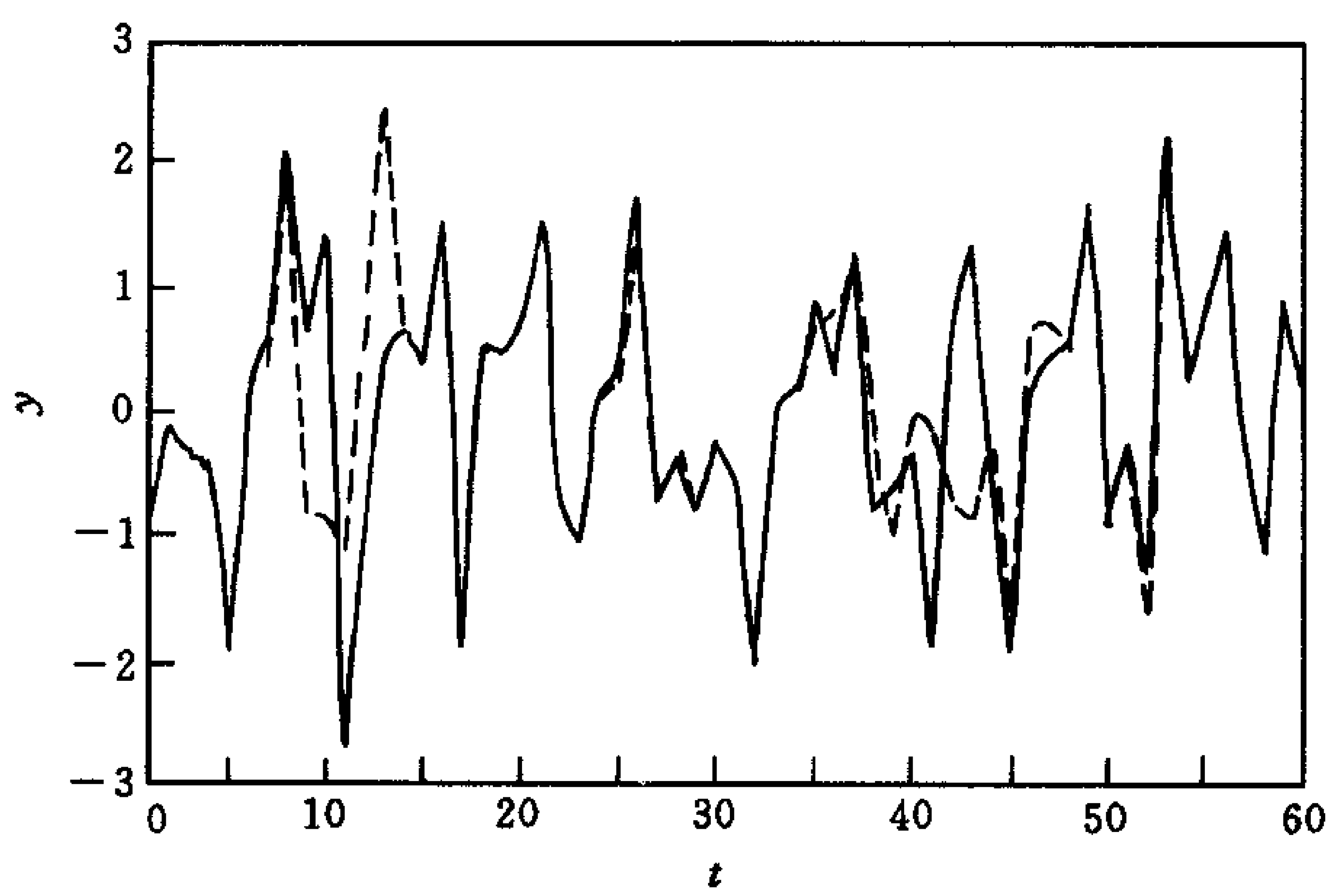
于变尺度法改进 BP 算法.

## 5 数值仿真

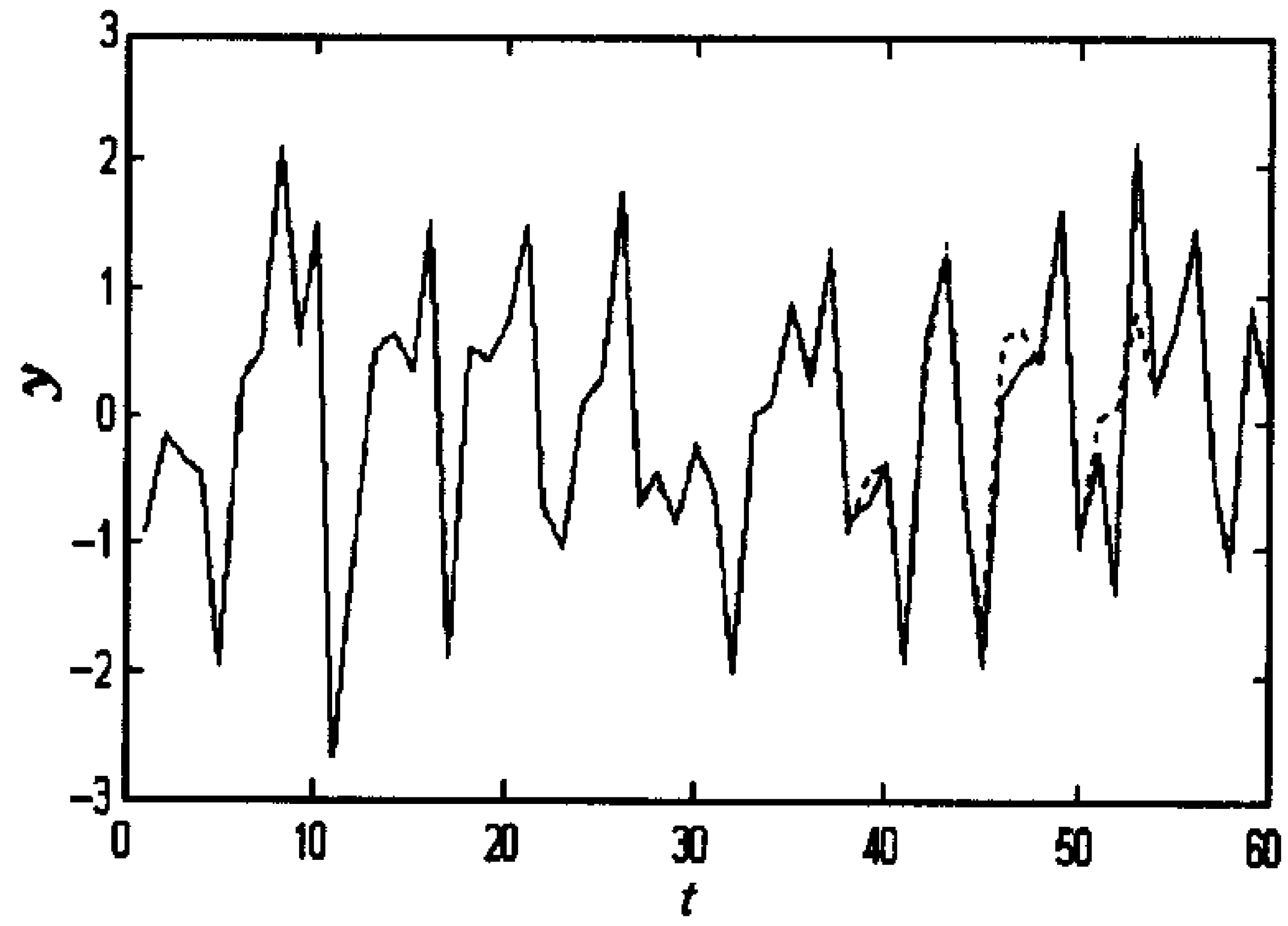
这里比较一下多 ANN 模型与单 ANN 模型的差别. 考虑动态系统

$$y(t+1)=\sin(\pi\cdot y(t))\cdot\exp(-u(t)). \tag{8}$$

首先建立一步预报模型. 设精度要求为均方差 $MSE<0.0001$. 单 ANN 模型,采用隐节点数可变的修正变尺度 BP 训练算法,经过2056次迭代,达到精度要求. 隐节点从6个开始增加到19个. 对多 ANN 模型,采用极值聚类训练算法,先对样本空间进行划分,得到6个极值中心;各子网络分别经过568,543,675,522,520,495次训练达到精度要求,隐节点数目分别为8,7,6,8,8,8个. 由一步预报模型进行长时段预报,分别从第10和40个采样时刻开始进行两次30步预报,结果都不理想. 在一步预报模型的基础上,再采用本文提出的方法进行长时段预报模型的辨识. 对单 ANN 模型,经50次修正均方差为0.4577,之后不再下降. 对多 ANN 模型,经68次迭代,均方差为0.0741. 多 ANN 模型的效果改善比较明显. 如图2(a),(b)所示,图中实线表示对象输出曲线,虚线表示模型预报输出曲线.




(a) 单网络的30步长时段预报模型学习结果




(b) 多网络的30步长时段预报模型学习结果

图2 单网络与多网络模型预报结果对比

## 6 小结

本文提出了一种基于极值聚类的多 ANN 结构训练方法,并将多 ANN 结构应用于复杂系统长时段预报,提出了基于多 ANN 结构的长时段预报模型的离线批方式辨识算法.所提出的多 ANN 结构的极值聚类学习策略,可以大大提高预报精度,对于具有多极值特性的一类复杂系统预报问题是一种可行的方法. 本文的方法可以作为进一步建立基于多 ANN 模型的复杂控制系统预测控制系统的基础.

## 参 考 文 献

[1] Jacobs R A, Jordon M I *et al.* Adaptive Mixtures of local experts. *Neural Computation*, 1991,**3**:79—87.

[2] Cho S B, Kim J H. combining multiple neural networks by fuzzy integral for rocognition. *IEEE Trans SMC*, 1995, **25**(1).

[3] Narendra K S, Levin A U. Regulation of nonlinear dynamical systems using multiple neural networks. , In: Proc. of ACC, 1990, 1609—1614.

[4] Hrycej T. A Modular Architecture for Efficient Learning. Int'l Joint Conf. on Neural Networks, San Diego. 1990, **1**:557—562.

[5] Li Renhou, Gao Feng. Complex function approximatino based on multi-ANN approach. *Chinese Journal of Systems Engineering and Electronics*, 1995,**6**(2):22—31.

[6] Werbos P J. Backpropagation through time:what is does and how to do it? Proc. of IEEE. 1990, **78**:1550—1560.

# LONG-TERM PREDICTION FOR COMPLEX SYSTEM BASED ON MULTIPLE ANN MODELS

GAO FENG LI RENHOU

(*Systems Engineering Institute*, *Xi'an Jiaotong University*, *Xi'an* 710049)

**Abstract** An extremum clustering approach for multiple artificial neural networks is developed in this paper. Then it is used for long-term prediction of complex dynamic system. With this method the long-term prediction precision and model robustness can be improved. This approach can be used for complex system predictive control using multiple ANNs.

**Key words** Multiple ANNs, multi-layer feedforward neural networks, prediction of nonlinear systems.