

基于信息几何的统计回馈神经网络 非线性自适应预测控制¹⁾

戴宪华

(汕头大学电子工程系 广东 515063)

摘要 研究统计回馈神经网络(SRNN)的非线性自适应预测控制. 基于混合统计模型, 利用信息几何的处理方法, 将SRNN的参数估计转化为一般的线性ARMA系统的最小均方误差参数估计算法, 最终获得SRNN参数估计. 获得RNN预测的参数估计以后, 可以十分方便地利用线性ARMA系统的控制规律来设计SRNN的预测控制规律, 解决了非线性SRNN预测参数估计、复杂非线性系统控制规律设计等问题. 在研究单隐元SRNN的基础上, 进一步探讨了多隐元SRNN的自适应预测控制问题.

关键词 统计回馈神经网络, 信息几何, 线性近似.

RECURRENT NEURAL NETWORK PREDICTION CONTROLLER BASED ON THE INFORMATION GEOMETRY

DAI Xianhua

(Dept. of Electronic Eng., Shantou University, Guangdong 515063)

Abstract The paper primarily investigates the adaptive prediction controller modeled by the statistic recurrent neural network(SRNN). Using a mixed statistic model to approximate single neuron of the SRNN, one can easily obtain both parameter estimation and control rule of the adaptive controller based on the procedure of information geometry. On the basis of adaptive controller of single hidden unit, the paper further investigates the controller problems of the complicated SRNN with multiple hidden units.

Key words Statistic recurrent neural network, information geometry, linear approximation.

1 引言

在实际工程应用中, 经常遇到的许多实际物理系统或多或少地存在非线性特性. 在众多的非线性模型中, 统计回馈神经网络是一类十分有效的非线性系统描述方法, 它的研究以及在控制、信息处理等方面的应用一直是众多学者研究的重点.

1) 国家自然科学基金资助课题.

与线性系统的预测误差控制技术相比,非线性系统自适应控制的研究一直处于初期发展阶段.目前,尚无普遍适用的控制方法可有效地解决一般非线性控制问题的全局收敛性等问题.

在处理非线性问题的众多方法中,基于局部线性化或局部线性近似的处理方法一直是十分有效的处理方法.它的最大优点是,非线性系统局部线性化后,可以直接利用大量成熟的线性系统控制技术方法,解决非线性系统的自适应控制问题.本文就是基于这一思想,结合信息几何的处理方法,将非线性系统控制问题转化为一系列线性系统的控制问题,最终利用线性系统的自适应控制方法解决非线性 RNN 的自适应控制问题.

2 单隐元 RNN 的预测控制

2.1 单隐元 RNN 预测的线性近似模型

考虑如下具有单隐元的简单 RNN 系统

$$\begin{cases} y_1(t) = y'_1(t) + n(t), \\ y'_1(t) = g\left(\sum_{i=1}^I a_i(t)y'_1(t-i) + \sum_{m=0}^{M-1} b_m(t)x(t-m) + \alpha\right), \end{cases} \quad (1)$$

其中 $x(t)$, $y_1(t)$ 分别为 RNN 的输入和输出信号; $a_i(t)$, $b_m(t)$ 为受控系统的参数,它们是未知的,需要在预测中估计; $g(\cdot)$ 为 RNN 的激励函数; $n(t)$ 为加性噪声,这里假定它满足高斯分布.

采用单隐元 RNN 预测模型,式(1)所描述系统的预测输出为

$$\hat{y}_1(t) = g\left(\sum_{i=1}^I \hat{a}_i(t)y_1(t-i) + \sum_{m=0}^{M-1} \hat{b}_m(t)x(t-m) + \alpha\right), \quad (2)$$

其中参数 $\hat{a}_i(t)$, $\hat{b}_m(t)$ 为非线性 RNN 预测的待估计参数.

在式(1),(2)中,由于 $g(\cdot)$ 为非线性算子,直接根据式(2)计算控制器的控制输入信号通常也是十分困难的.但根据式(1),(2)可知,当 $g(\cdot)$ 为线性算子时,单隐元 RNN 退化为线性 ARMA 系统.在一定的条件下,线性 ARMA 系统预测的参数估计可以收敛于全局最优解^[2,3,4].而基于线性 ARMA 系统的预测控制设计也较为简单,可直接利用预测控制或极点配置控制设计方法来获得全局收敛的控制规律^[4].因此,若能将单隐元 RNN 转化为线性 ARMA 系统,将可能获得较好的参数估计收敛结果及简单的预测控制规律.下面研究如何将式(2)转化为线性系统.

当激励函数为连续泛函时,它在定义域的充分小区域将可利用它的一阶 Taylor 展开近似.因此,RNN 的激励函数取为软非线性函数,如 Sigmoidal 函数时,式(2)可用下面一系列线性系统近似

$$\hat{y}_1(t) = \sum_{k=1}^K y_{1,k}(t)\delta_k, \quad (3)$$

$$y_{1,k}(t) = g(y_{1,k}^0) + [y'_1(t) - y_{1,k}^0] \frac{\partial g(y'_1(t))}{\partial y'_1(t)} \Big|_{y'_1(t)=y_{1,k}^0} = g(y_{1,k}^0) + l(y_{1,k}^0)[y'_1(t) - y_{1,k}^0], \quad (4)$$

$$y'_1(t) = \sum_{i=1}^I \hat{a}_i(t)y_1(t-i) + \sum_{m=0}^{M-1} \hat{b}_m(t)x(t-m) + \alpha, \quad (5)$$

其中 $\delta_k = \begin{cases} 1, & y'_1(t) \in A_k \\ 0, & \text{Otherwise} \end{cases}$, A_k 表示 $g(\cdot)$ 定义域中以 $y_{1,k}^0$ 为中心的小邻域, 并且满足,

$\bigcup_{k=1}^K A_k = A$, A 为 $g(\cdot)$ 的定义域, k 称为 RNN 的状态指标.

从上面模型描述可以看出, 当线性近似的数目 K 充分大时, 公式(3)—(5)描述的线性系统将完全可以近似预测系统(2), 近似误差随着 K 的增大而趋于零. 这样, 就可以将非线性 RNN 预测转化为一系列线性 ARMA 系统的预测问题.

2.2 单隐元 RNN 线性近似模型的几何

在 SRNN 预测模型的训练学习中, 训练数据或指导信号为 $y_1(t)$. 根据上面的线性近似模型, 若已知训练数据 $y_1(t)$ 可作为式(3)—(5)描述的线性系统的指导信号, 那么参数估计即为一般的线性 ARMA 系统预测的参数估计问题^[2,3,4]. 但是, 通常并不能预先知道训练数据 $y_1(t)$ 可作为 K 个线性系统中的哪一个线性系统的指导信号. 因此, 模型参数估计是一不完全观测数据 (incomplete-data) 的参数估计问题. 那么参数估计包含如下两个方面: 1) 确定当前训练数据 $y_1(t)$ 可能作为哪一个线性系统的指导信号或期望信号, 即定位或状态指标 k 的确定问题, 这是离散概率事件的模式识别问题. 2) 确定每个线性系统的训练数据后, 利用线性 ARMA 系统预测算法估计 $\hat{a}_i(t), \hat{b}_m(t)$.

为解决上面的两个问题, 将引入混合统计模型. 根据线性近似模型(3)—(5)知道在某一时刻 RNN 的训练数据 $y_1(t)$ 将作为 K 个线性近似系统中的一个线性系统的指导信号. 同时, 由于 RNN 的输出包含高斯加性噪声干扰, 那么受控系统的输出 $y_1(t)$ 可以利用如下统计模型来描述^[1,5,6]

$$p(y_1(t) | \hat{y}_1(t)) = \sum_{k=1}^K \pi_k p(y_1(t) | y_{1,k}(t)), \quad (6)$$

$$p(y_1(t) | y_{1,k}(t)) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{[y_1(t) - y_{1,k}(t)]^2}{2\sigma^2}\right), \quad (7)$$

其中 $\sum_{k=1}^K \pi_k = 1$. 根据上面的分析, 知道状态指标 k 是一离散随机变量 (也称为隐含变量),

为此引入变量 $\delta_k(z) = \begin{cases} 1, & y'_1(t) \in A_k \\ 0, & \text{Otherwise} \end{cases}$, 那么单隐元 RNN 的联合统计分布可写为

$$p(y_1(t), z) = \sum_{k=1}^K \delta_k(z) \pi_k p(y_1(t) | y_{1,k}(t)) = \prod_{k=1}^K [\pi_k p(y_1(t) | y_{1,k}(t))]^{\delta_k(z)}. \quad (8)$$

对应的对数似然函数为

$$\log p(y_1(t), z) = \sum_{k=1}^K \delta_k(z) \left[\log \pi_k - \frac{1}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} [y_1(t) - y_{1,k}(t)]^2 \right]. \quad (9)$$

$$\text{令 } r_{1,k}(t) = \delta_k(z), \quad r_{2,k}(t) = y_1(t) \delta_k(z), \quad r_{3,k}(t) = y_1^2(t) \delta_k(z), \quad (10)$$

$$\theta_{1,k}(t) = \log \pi_k - \frac{1}{2\sigma^2} y_{1,k}^2(t), \quad \theta_{2,k}(t) = \frac{y_{1,k}}{\sigma^2}, \quad \theta_{3,k}(t) = -\frac{1}{\sigma^2}, \quad (11)$$

$$\mathbf{r}(t) = [r_{1,1}(t), \dots, r_{1,K}(t), r_{2,1}(t), \dots, r_{2,K}(t), r_{3,1}(t), \dots, r_{3,K}(t)]^T, \quad (12)$$

$$\boldsymbol{\theta}(t) = [\theta_{1,1}(t), \dots, \theta_{1,K}(t), \theta_{2,1}(t), \dots, \theta_{2,K}(t), \theta_{3,1}(t), \dots, \theta_{3,K}(t)]^T.$$

那么, 联合统计分布式(8)可写为

$$p(\mathbf{r}(t), \boldsymbol{\theta}(t)) = p(y_1(t), z) = \exp(\boldsymbol{\theta}^T(t) \mathbf{r}(t) - \varphi(\boldsymbol{\theta}(t))). \quad (13)$$

显然, 观测量 $\mathbf{r}(t)$ 在自然参数 $\boldsymbol{\theta}(t)$ 下为一指数族分布. 但参数 $\boldsymbol{\theta}(t)$ 为 SRNN 参数 $\hat{a}_i(t)$,

$\hat{b}_m(t)$ 的非线性函数,因此,SRNN 参数构成的集合为自然参数流形的子流形,是一曲指数族^[1,7].

在分布函数(13)式中,观测量 $r(t)$ 包含隐含变量,观测量的统计期望 η 为另一与 $\theta(t)$ 对等的参数坐标系,有关指数族与曲指数族的几何性质可参见文献[1,7].

2.3 RNN 参数估计

从以上的分析,可利用最大似然估计方法获得自然参数 $\theta(t)$ 的估计. 然后,再利用 $\theta(t)$ 与参数 $\hat{a}_i(t), \hat{b}_m(t)$ 的关系获得 RNN 参数估计. 但是,在上面的指数族分布中,观测量 $r(t)$ 包含隐含变量,因此,不能直接利用最大似然估计方法估计参数. 这里采用基于信息几何的 em-投影算法.

a) e-投影

根据上面的观测矢量,其中状态指标 k 为隐含未知的, $y_1(t)$ 是已知可观测的. 那么利用 t 时刻的自然参数估计 $\hat{\theta}(t)$, 可求得观测量 $r(t)$ 的估计为

$$\hat{r}_{1,k}(t) = \alpha_k, \quad \hat{r}_{2,k}(t) = \alpha_k y_1(t), \quad \hat{r}_{3,k}(t) = \alpha_k y_1^2(t), \quad (14)$$

$$\hat{r}(t) = [\alpha_1, \dots, \alpha_K, \alpha_1 y_1(t), \dots, \alpha_K y_1(t), \alpha_1 y_1^2(t), \dots, \alpha_K y_1^2(t)]^T, \quad (15)$$

其中 $\alpha_k = p(k | y_{1,k}(t))$ 表示训练数据 $y_1(t)$ 可作为第 k 个线性 ARMA 系统指导信号的概率,由下式估计

$$\alpha_k = \frac{\pi_k \exp\left(-\frac{(y_1(t) - y_{1,k}(t))^2}{2\sigma^2}\right)}{\sum_{i=1}^K \pi_i \exp\left(-\frac{(y_1(t) - y_{1,i}(t))^2}{2\sigma^2}\right)}. \quad (16)$$

b) m-投影

根据信息几何的理论,知道在指数族分布中,观测量的期望参数为一充分统计量,因此,可以利用期望参数来获得自然参数的估计. 根据观测量的表示式,观测量的期望参数估计为

$$\hat{\eta}(t+1) = \rho \hat{\eta}(t) + (1 - \rho) \hat{r}(t), \quad 0 \leq \rho \leq 1. \quad (17)$$

当 $\rho=0$ 时,式(17)为无记忆的期望参数估计. 那么,自然参数估计为

$$\hat{\theta}(t+1) = \hat{\theta}(t) + \epsilon G^T [\hat{\eta}(t+1) - \eta(\theta(t))], \quad (18)$$

其中 G^T 为 Fisher 信息矩阵 G 的逆矩阵, $G = E \left[\frac{\partial}{\partial \theta_i} \log p(r, \theta) \frac{\partial}{\partial \theta_j} \log p(r, \theta) \right]_{3K \times 3K}$, $\eta(\theta(t))$ 为参数估计 $\hat{\theta}(t)$ 下的期望参数.

令 $\hat{\beta}(t) = [\hat{a}_1(t), \dots, \hat{a}_I(t), \hat{b}_0(t), \dots, \hat{b}_{M-1}(t)]^T$ 为 RNN 预测的参数估计,并设参数 $\hat{\beta}(t)$ 与自然参数 $\hat{\theta}(t)$ 的关系为 $B = \frac{\partial \theta}{\partial \beta}$. 那么,参数 $\hat{\theta}(t)$ 的估计为

$$\hat{\beta}(t+1) = \hat{\beta}(t) + \epsilon G^T B [\hat{\eta}(t+1) - \eta(\theta(t))]. \quad (19)$$

根据文献[1,7],式(2)~(10),式(19)中的 $\eta(\theta(t)) = \frac{\partial \varphi(\theta)}{\partial \theta} = 0$. 当式(17)中的 $\rho=0$, 即观测量的期望参数取为无记忆估计时,式(19)的 $B \hat{\eta}(t+1)$ 可展开为

$$B \hat{\eta}(t+1) = y(t) e(t), \quad (20)$$

其中 $e(t) = \sum_{k=1}^K l(y_{1,k}^0) \alpha_k [y_1(t) - y_{1,k}(t)]$, $y(t) = [y_1(t), \dots, y_1(t-I), x(t), \dots, x(t-M+1)]^T$.

将式(20)代入式(19),并与文献[2,3,4]中的一般线性 ARMA 系统预测参数估计算法比较,可以看到 SRNN 预测参数估计公式与一般的线性 ARMA 系统预测的参数估计基本相同,只是其中的误差与线性 ARMA 系统稍有不同.因此,通过近似处理,将非线性 SRNN 预测转化为线性 ARMA 系统预测问题.线性 ARMA 系统预测参数估计的各种算法和算法收敛性质分析可直接应用到 m-投影参数估计中.同样利用信息几何的批处理方法,可以求得 π_k, σ^2 的估计^[1,6,7].注意这里的 em-投影算法是交替迭代获得参数估计的^[1,7].

2.4 RNN 最小均方预测误差控制规律

在上述线性近似模型及信息几何的描述下,有两种方法可以计算 RNN 的最小均方预测误差控制规律.设 RNN 控制器的期望信号为 $d_1(t)$.

方法1.在获得了 RNN 参数估计 $\hat{\beta}(t)$ 后,RNN 预测可写为

$$\begin{aligned} \hat{y}_1(t) = & \sum_{k=1}^K y_{1,k}(t) \alpha_k = \sum_{k=1}^K \alpha_k g(y_{1,k}^0) + l(y_{1,k}^0) [y_1'(t) - y_{1,k}^0] = \\ & \sum_{k=1}^K \alpha_k [g(y_{1,k}^0) - l(y_{1,k}^0) y_{1,k}^0] + \sum_{k=1}^K \alpha_k l(y_{1,k}^0) \left[\sum_{i=1}^I \hat{a}_i(t) y_1(t-i) + \right. \\ & \left. \sum_{m=0}^{M-1} \hat{b}_m(t-m) + \alpha \right]. \end{aligned} \quad (21)$$

令 $\hat{C}_1(t) = \sum_{k=1}^K \alpha_k [g(y_{1,k}^0) - l(y_{1,k}^0) y_{1,k}^0]$, $\hat{C}_2(t) = \sum_{k=1}^K \alpha_k l(y_{1,k}^0)$, 那么式(21)可简化为

$$\hat{y}_1(t) = \hat{C}_1(t) + \hat{C}_2(t) \left[\sum_{i=1}^I \hat{a}_i(t-i) + \sum_{m=1}^{M-1} \hat{b}_m(t) x(t-m) + \alpha \right]. \quad (22)$$

显然,式(22)为一具有时变系数的线性 ARMA 系统,因此它的控制规律,如控制输入信号可采用线性 ARMA 系统的自适应预测控制规律的设计方法来获得^[4].

方法2.在上面的 RNN 预测训练数据的分布(13)式中,自然参数 $\theta(t)$ 为受控系统输入信号 $x(t)$ 的函数.因此,可通过先估计自然参数,然后利用自然参数与受控系统输入信号的关系,求出 RNN 控制器的输入信号.

将指数分布(13)式中的观测量 $r(t)$ 改造为

$$r(t) = [\alpha_1, \dots, \alpha_K, \alpha_1 d_1(t), \dots, \alpha_K d_1(t), \alpha_1 d_1^2(t), \dots, \alpha_K d_1^2(t)]^T. \quad (23)$$

利用上面的 em-投影算法可求出 $y_{1,k}(t), k=1, \dots, K$ (它们为受控系统输入 $x(t)$ 的函数), 然后根据关系式(4),即可求出 RNN 控制系统的输入信号.

从上面的讨论看出,非线性 RNN 的控制问题完全转化为线性 ARMA 系统的预测控制问题.同时,也可根据线性 ARMA 系统的稳定性分析方法分析非线性 RNN 控制系统的稳定性问题.关于稳定性分析,限于篇幅,本文不再讨论.

3 具有多隐元的复杂 RNN 预测控制

根据单隐元 SRNN 模型,多隐元的 RNN 预测的局部线性近似模型示于图1.对于多隐元 RNN 的预测若能分解为多个单隐元 RNN 预测,那么多隐元 RNN 的预测控制问题也就转化为多个单隐元 RNN 的预测控制问题.

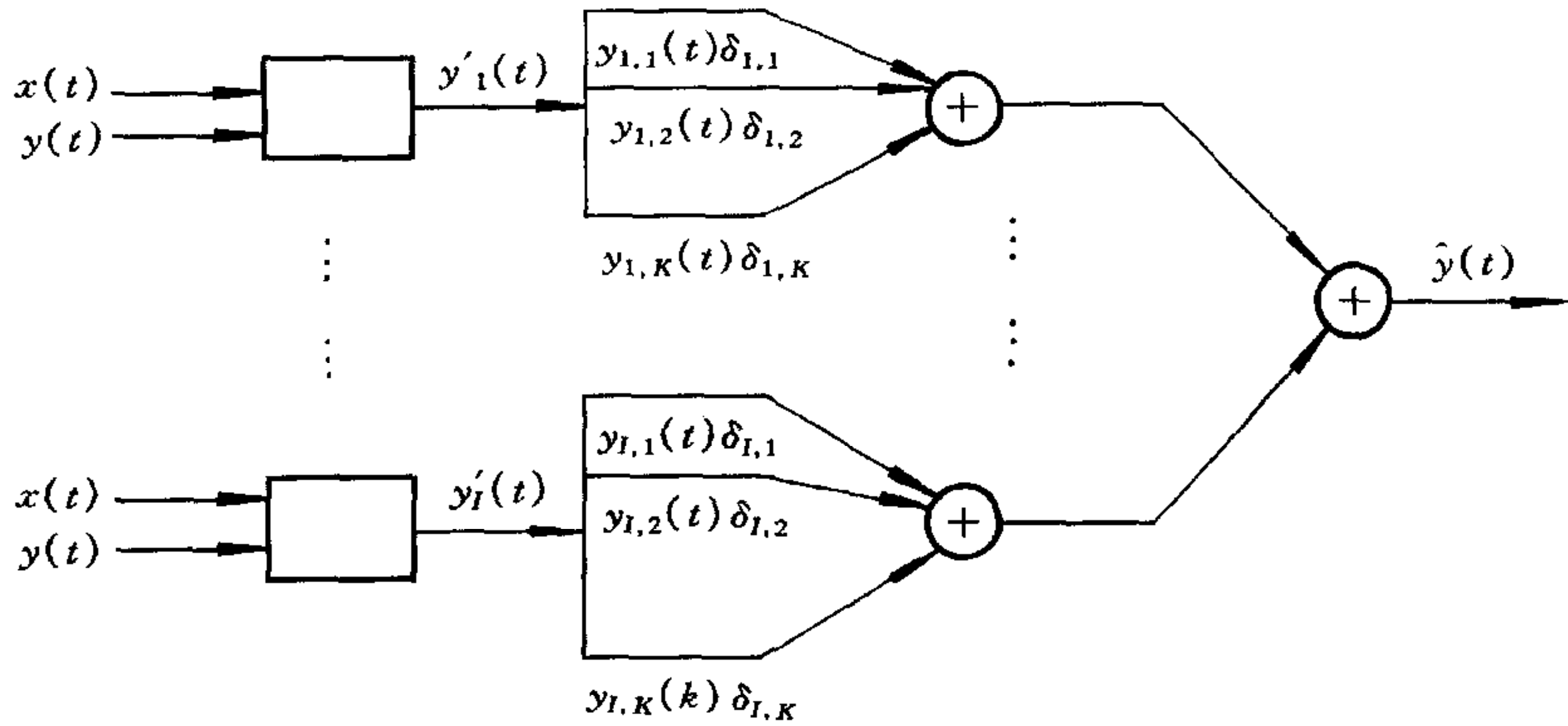


图1 多隐元的 RNN 预测局部线性近似模型

在上面的模型中,多个单隐元 RNN 的输出合成作为 RNN 预测的输出. 因此,如果 RNN 每一隐元对应的训练数据或期望信号确知,那么多隐元 RNN 的参数估计、自适应控制即为多个单隐元 RNN 的分别控制及参数估计. 但是,多隐元 RNN 预测每一隐元的训练数据、期望信号并没有给出,需要在参数估计的过程中估计每一隐元的训练数据、期望信号. 这里,采用如下简单的估计方式给出各隐元的训练数据、期望信号. 第 l 隐元在 $t + 1$ 时刻的训练数据、期望信号分别为

$$y_l(t + 1) = y(t + 1) - \sum_{i=1, i \neq l}^l \hat{y}_i(t + 1),$$

$$d_l(t + 1) = d(t + 1) - \sum_{i=1, i \neq l}^l \hat{y}_i(t + 1), \quad l = 1, \dots, I, \quad (24)$$

其中 $\hat{y}_i(t + 1)$ 为利用 t 时刻参数估计的第 i 隐元的一步前后预测输出信号^[4].

在多隐元 RNN 的每一隐元获得了训练数据及期望信号后,多隐元 RNN 预测的参数估计以及预测控制也就转化为多个单隐元的参数估计、控制问题,可利用本文第二部分提出的方法解决.

注意在多隐元 RNN 转化为多个单隐元控制问题以后,它是一个多输入、多输出 (MIMO) 的线性 ARMA 系统,需利用 MIMO 线性 ARMA 系统的参数估计、控制方法来解解决参数估计及控制问题^[4].

4 计算机模拟实验结果

实验模型如下

$$y'(t) = \sum_{i=1}^2 w_i(t) g \left(\sum_{n=1}^3 a_n(t) y'(t - n) + \sum_{m=0}^2 b_m(t) x(t - m) \right), \quad (25)$$

$$y(t) = y'(t) + n(t), \quad (26)$$

其中 $n(t) = 0.8n(t - 1) + w(t)$, $w(t)$ 为零均值高斯白噪声. RNN 预测取与式 (25), (26) 同阶的预测形式,每一隐元的局部线性化数目取为 11. RNN 控制的期望信号取为方波.

图2为部分实验结果,其中(a)为 RNN 控制的期望信号 $d(t)$ 和受控系统 (25), (26) 的控制输出信号; (b) 为 RNN 控制器控制输入信号; (c) RNN 预测系统的参数估计学习曲

线,它是30次独立的计算机模拟结果的平均,曲线表征了 RNN 预测参数估计或系统辨识的收敛性能.从图2可看出,本文提出的 RNN 预测控制器的设计方法是完全可行的.

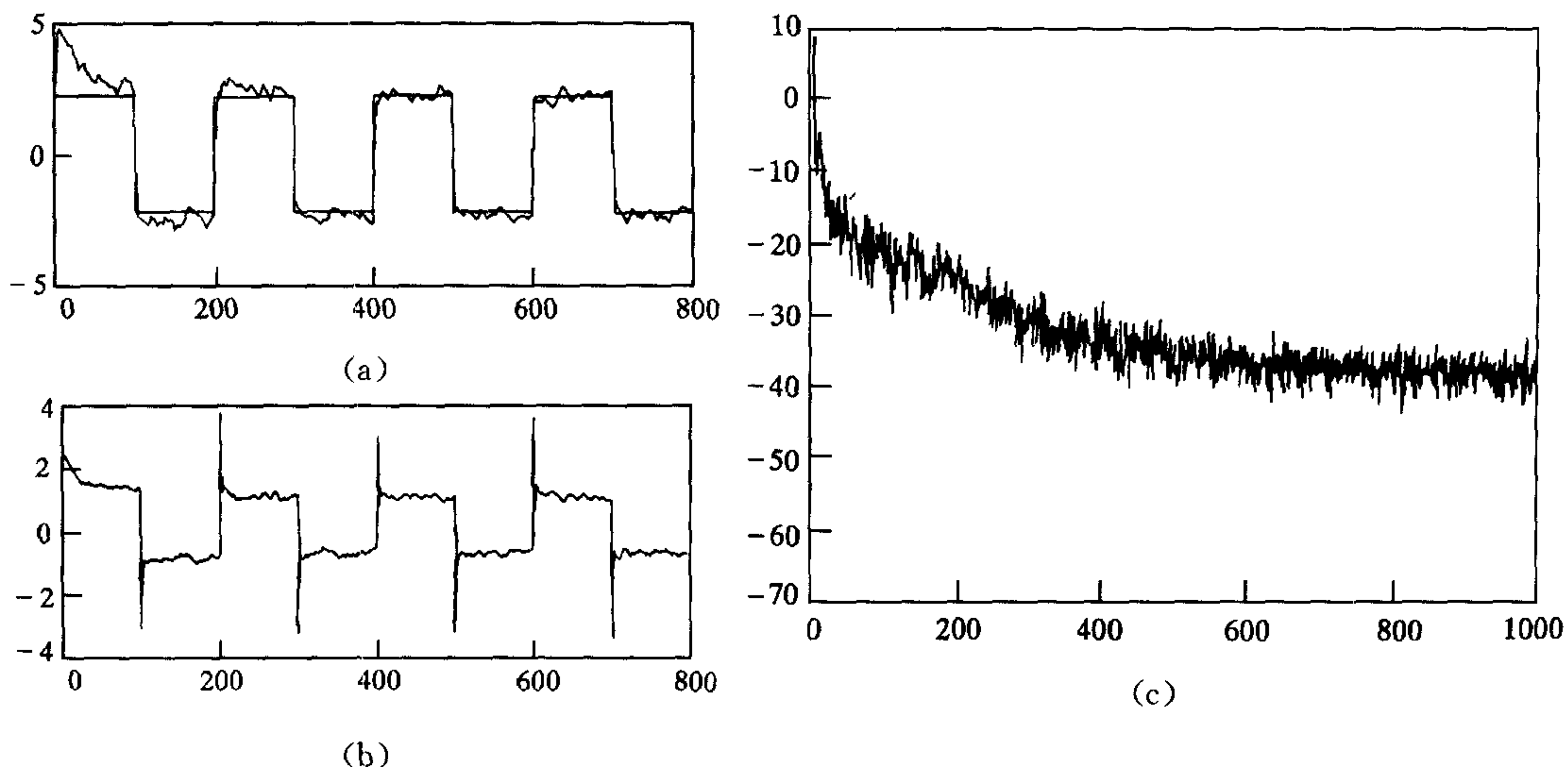


图2 预测控制实验结果

5 结论

本文利用 SRNN 的非线性激励函数的分段局部线性近似方法,将非线性 SRNN 转化为混合统计模型表示的统计预测问题.利用信息几何的 em-投影算法,SRNN 预测转化为线性 ARMA 系统的预测问题,最终利用线性 ARMA 系统预测控制的结论解决非线性 SRNN 的预测控制问题.

参 考 文 献

- 1 Shun-ich Amari. Information geometry of the EM and em algorithms for neural networks. *Neural Network*, 1995, 8(9):1379—1408
- 2 Anderson B D O. Stability of adaptive systems: passivity and averaging analysis. Cambridge, London, The MIT Press, 1984
- 3 Kumar P R. Stochastic systems estimation, identificaion and adaptice control. Englewood Cliffs, NJ; Prentice-Hall, Inc. 1986
- 4 Graham C. Goodwin Adaptive filtering prediction and control. Englewood Cliffs, NJ; Prentice-Hall, Inc. 1984
- 5 Xuan J. Information geometry of maximum partial likelihood estimation for channel equalization. In: ICASSP'96: 3533—3566
- 6 Roy L. Streit maximum likelihood training of probabilistic neural network. *IEEE Trans. Neural Network*, 1994, 5(5):764—783
- 7 Shun-ich Amari. Differential geometry of curved exponential family-curvature and information loss. *The Annual of statistics*, 1982, 10(2):357—385

戴宪华 1992年毕业于东南大学,获博士学位.1993—1994年在华南理工大学通信与电子系统博士后流动站工作;1995年至今,在汕头大学电子工程系任教,同年获得副教授任职资格.主要科研领域:非线性信号处理,语音信号处理,通信系统均衡,自适应控制等.