

基于累积量的两层前馈神经网络盲辨识¹

戴宪华

(汕头大学电子工程系 汕头 515063)

摘 要 由于非线性系统输出是其参数的非线性函数, 直接利用高阶累积量辨识两层前馈神经网络 (FNN) 通常是十分困难的. 为解决这一问题, 该文提出两种基于四阶累积量的 FNN 辨识方法. 第一种方法, FNN 的隐元在其输入空间利用多个线性系统近似, 进而 FNN 利用一统计模型—混合专家 (ME) 网络重新描述. 基于 ME 模型, FNN 参数可利用统计期望值最大化 (EM) 算法获得估计. 第二种方法, 为简化 FNN 的 ME 模型, 引入隐含观测量. 基于隐含观测量估计, FNN 被分解为多个单隐元的训练问题, 进而整体 FNN 可利用一两阶层 ME 描述. 基于单隐元的参数估计, FNN 可利用一具有更快收敛速度的简化算法获得估计.

关键词 ME 模型, EM 算法, 基于累积量的系统辨识

中图分类号 TN-052

1 引 论

盲系统辨识估计作为盲信号处理的重要手段已在包括通信系统盲均衡、盲源分离、独立分量分析、反射地震学等许多领域中发挥重要作用. 目前, 大部分关于盲系统估计、辨识的研究都局限于线性系统^[1-7]. 但是, 实际工程中的许多系统或多或少带有非线性特征, 通常必须利用非线性系统模型来描述. 在众多的非线性系统模型中, 神经网络模型具有良好的系统近似能力, 已成为目前相当成功的一类非线性系统模型. 因此, 研究基于累积量的前馈神经网络 (FNN) 系统盲辨识具有重大的理论和应用意义.

传统的 FNN 指导训练一般是基于给定输入和输出信号构成的完备训练集^[8-13]. 但是, 在前面提到的领域中, 不可能或不便于获得完备的训练集. 通常仅具有系统的输出观测, 而系统的输入是不可观测的, 仅已知输入信号的部分统计知识, 如输入的高阶统计量 (HOS). 在这种情况下, 前馈神经网络 (FNN) 训练实际上是仅基于输出观测量的系统辨识问题, 也称为盲辨识问题^[1-7]. 在盲辨识中, 通常有两种参数估计方法. 第一为基于系统输出和输入 HOS 的系统辨识. 第二为基于系统输出和输入为统计独立条件的系统辨识, 如盲源分离、独立分量分析等. 本文重点研究第一种情况.

对于线性系统, 由于系统输出是其参数的线性函数, 基于 HOS 的参数估计相对简单. 但对于 FNN 非线性系统, 由于系统输出是其参数的非线性函数, HOS 又是系统输出的非线性函数, 因此基于 HOS 的 FNN 参数估计更为复杂, 一般不能直接利用累积量估计参数. 这也是目前极少有关基于累积量的非线性系统盲辨识研究报道的主要原因. 如何解决这一问题, 实现 FNN 基于累积量的盲辨识将是本文研究的主要目标.

2 两层 FNN 及训练集

实际工程中的许多系统如数字卫星通信信道、磁记录信道等都具有非线性特征. 它们中的大部分可利用一个两层 FNN 来描述.

$$y(t) = \sum_{i=1}^I w(i)y_i(t) = \sum_{i=1}^I w(i)g \left[\sum_{m=0}^M w(i,m)x(t-m) - \beta_i \right] \quad (1)$$

¹ 1999-09-13 收到, 2000-07-07 定稿

国家自然科学基金 (69872021)、广东省自然科学基金 (980438) 资助课题

其中激励函数 $g[\bullet]$ 选为 Sigmoid 函数。 $w(i), w(i, m)$ 为 FNN 参数也是待估计参数, M, I 分别为 FNN 输入数目和隐元数目。信号 $y(t), x(t)$ 分别为系统的输出和输入。在训练阶段假定: 系统输出的指导信号为可观测, 但系统输入不可观测, 仅其部分统计信息已知, 如输入信号为四阶白²的统计过程。与其它指导训练相比, 给定的观测信号或训练集是不完备的。当然, 系统可观测量越少, 则系统辨识就越困难。

记 FNN 输出的指导信号(期望信号)为 $d(t)$, 假定它已被加性高斯噪声污染。FNN 盲辨识的目的就是仅利用系统输出的指导信号, 在系统输入为四阶白的条件下估计 FNN 参数。由于高阶累积量对高斯噪声具有抑制作用, 本文将采用四阶累积量匹配原则^[1-6]估计 FNN 参数, 即

$$\min_{\vartheta} \sum_R [c_d(\tau_1, \tau_2, \tau_3) - c_y(\tau_1, \tau_2, \tau_3)]^2 \quad (2)$$

其中 $c_y(\tau_1, \tau_2, \tau_3), c_d(\tau_1, \tau_2, \tau_3)$ 分别为 FNN 输出和输出指导信号的四阶累积量。 R 为非冗余支撑域, τ_1, τ_2, τ_3 表示取值于 R 的时延参数, 以下 (τ_1, τ_2, τ_3) 简记为 (τ_{123}) 。 ϑ 表示 FNN 所有参数集合。累积量定义为累积量产生函数的 Taylor 展开式的系数。四阶累积量为

$$c_y(\tau_{123}) = \mu_y(\tau_{123}) - \mu_y(\tau_1)\mu_y(\tau_2 - \tau_3) - \mu_y(\tau_2)\mu_y(\tau_1 - \tau_3) - \mu_y(\tau_3)\mu_y(\tau_1 - \tau_2)$$

其中 $\mu_y(\tau_{123}), \mu_y(\tau)$ 分别为 $y(t)$ 的四阶和二阶中心矩。

上述系统模型和训练集实际上是从实际通信工程中抽象出来的一类信道模型, 其中系统(1)式用来表示或近似实际的未知传输信道, 指导信号 $d(t)$ 则表示信道接收信号, 系统输入 $x(t)$ 表示传输序列。系统(1)式的逆系统则是信道均衡器。在信道盲均衡中, 传输序列 $x(t)$ 是未知的。

从 FNN 系统模型和累积量关系式可看出, 四阶累积量是 FNN 参数的复杂非线性函数, 通常不能直接利用累积量匹配原则估计 FNN 参数。为解决这一问题, FNN 的隐元需要线性化处理。当 FNN 的激励函数取为连续函数时, 在其输入空间的一充分小区域, 激励函数将可利用线性函数近似。那么, 在激励函数的整个输入空间, 激励函数将可利用多个线性函数近似。基于这一事实, FNN 第 i 隐元可近似为

$$y_i(t) = \sum_{k=1}^K \Delta_{i,k} y_{i,k}(t) \quad (3)$$

$$\begin{aligned} y_{i,k}(t) &= g(y_{i,k}^0) + [y_i'(t) - y_{i,k}^0] \left. \frac{\partial g(y_i'(t))}{\partial y_i'(t)} \right|_{y_i'(t)=y_{i,k}^0} \\ &= g(y_{i,k}^0) + l(y_{i,k}^0)[y_i'(t) - y_{i,k}^0] \end{aligned} \quad (4)$$

$$y_i'(t) = \sum_{m=0}^M w(i, m)x(t-m) - \beta_i \quad (5)$$

其中 $y_{i,k}^0, k = 1, \dots, K$ 为 $g[\bullet]$ 输入空间的一系列固定点。 $\Delta_{i,k} = \begin{cases} 1, & y_i'(t) \in A_{i,k} \\ 0, & \text{其它} \end{cases}$, $A_{i,k}$ 表示以 $y_{i,k}^0$ 为中心的小邻域, 并满足 $\bigcup_{k=1}^K A_{i,k} = A_i$, A_i 表示第 i 隐元激励函数 $g[\bullet]$ 的输入空间。 $g(y_{i,k}^0), l(y_{i,k}^0)$ 为 $y_{i,k}^0$ 点的 $g[\bullet]$ 函数值和一阶导数值。 K 为线性近似的数目。

从以上线性近似可看出, FNN 的每个隐元已被多个线性有限冲激响应系统(FIR)近似。与原始 FNN 相比, 线性系统的四阶累积量将更加容易计算, 这也使得基于累积量的 FNN 训练成为可能。

² 四阶白是指四阶累积量除零时延分量非零外, 其余分量全为零的统计过程

3 基于期望值最大化 (EM) 算法的 FNN 训练

如上所述, FNN 的非线性隐元已被多个线性系统近似。这一近似与一般并联系统不同, 每个近似系统的输出在某时刻都有可能单独成为隐元的输出, 即线性近似并不是简单的线性求和输出的近似, 而是每个近似系统各以统计概率近似原始的隐元。因此, 上述线性近似可利用统计模型描述。

3.1 两层 FNN 的统计模型——简单混合专家 (ME) 网络

由于上述线性近似的每个线性系统都有可能单独近似隐元, 隐元的线性近似系统 (3)-(5) 式是以统计概率方式近似原始的隐元。每个隐元线性近似的输出又作为输出层的输入以线性和的方式产生 FNN 的输出。因此整体 FNN 可利用统计模型描述。根据文献 [8-12], 每个隐元的线性近似以线性和的方式获得 FNN 的输出, 整体 FNN 实际上构成一混合专家 (ME) 网络, 其中每个专家是一线性 FIR 系统, 专家系统输出为

$$\bar{y}_n(t) = \sum_{i=1}^I w(i)y_i(t) = \sum_{i=1}^I w(i)\Delta_{i,i(k)}y_{i,i(k)}(t) \quad (6)$$

其中 $i(k)$ 为第 i 隐元的指标函数, 它取值于 $\{1, \dots, K\}$, 且 $i(k) = k$ 表示第 i 隐元被 (4), (5) 式中的第 k 个线性系统近似。

当 $i(k)$ 取值于 $\{1, \dots, K\}$, 每个专家系统输出又可进一步写为

$$\begin{aligned} \bar{y}_n(t) &= \sum_{i=1}^I w(i)l(y_{i,i(k)}^0) \sum_{m=0}^M w(i,m)x(t-m) + c_n \\ &= \sum_{m=0}^M \left[\sum_{i=1}^I w(i)w(i,m)l(y_{i,i(k)}^0) \right] x(t-m) + c_n \\ &= \sum_{m=0}^M a_n(m)x(t-m) + c_n \end{aligned} \quad (7)$$

其中 c_n 为一常数。当所有隐元的指标函数 $i(k)$ 取遍集合 $\{1, \dots, K\}$, 所有可能的混合专家系统数目为 $N = K^I$ 。因此, (7) 式中的下标 n 将取值于集合 $\{1, \dots, N\}$ 。

基于 N 个形如 (7) 式的专家系统近似 FNN, 两层 FNN 可利用如下 ME 描述^{[8-12] 3}

$$p(c_d(\tau_{123}), \Omega|y(t), \vartheta(t)) = \prod_{n=1}^N [\pi_n q \exp(-\lambda(c_d(\tau_{123}) - c_n(\tau_{123}))^2)]^{\delta_n(t)} \quad (8)$$

其中 λ 表示概率密度的方差, q 为使 (8) 式满足概率分布条件的常数, ϑ 表示 FNN 的所有参数集合。 $\delta_n(t) = \begin{cases} 1, & \Omega = n \\ 0, & \Omega \neq n \end{cases}$, Ω 为以概率 $\pi_n = \text{prob}(\Omega = n)$ 取值于集合 $\{1, \dots, N\}$ 的随机变量, $\Omega = n$ 表示 FNN 为形如 (7) 式的第 n 个专家系统近似。根据概率约束, 有 $\sum_{n=1}^N \pi_n = 1$ 。 $c_n(\tau_{123})$ 表示形如 (7) 式的第 n 个专家系统输出的四阶累积量。根据 (7) 式, $c_n(\tau_{123})$ 可利用

³ 使用高斯混合分布的原因是它们与通常使用的均方误差准则 (2) 式一致, 以下相同。

下式计算^[1-6]:

$$c_n(\tau_{123}) = \sigma_{4r} \sum_{m=0}^M a_n(m)a_n(m+\tau_1)a_n(m+\tau_1)a_n(m+\tau_2)a_n(m+\tau_3) \quad (9)$$

其中 σ_{4r} 为输入信号 $x(t)$ 的四阶累积量。根据输入的假设, $x(t)$ 为四阶白过程, 因此 σ_{4r} 为常数, 可根据 $x(t)$ 的统计分布计算。

(9) 式的累积量实际上是一线性 FIR 系统输出的四阶累积量。与 FNN 输出的累积量相比, (9) 式的累积量显然更为简单, 从 (9) 式计算参数估计的下降梯度也更为简单。这一处理使得基于累积量的非线性 FNN 参数估计成为可能。

3.2 基于 EM 算法的 FNN 训练方法

根据上述统计 ME 模型, 每个专家系统 (7) 式的参数 $a_n(m)$, $n = 1, \dots, N$, $m = 0, \dots, M$ 可利用统计期望值最大化 (EM) 算法估计。EM 算法实际上是一互为条件的两步迭代算法。第一步也称为 E 步: 基于 EM 算法前次迭代中 M 步的参数估计, 估计隐变量 $\Omega = n$ 的概率。第二步也称为 M 步: 在获得了多个专家系统近似 FNN 的概率后, 对应条件概率密度 (8) 式的似然函数已是一完备数据的四阶累积量误差平方函数。这实际上是一线性 FIR 系统基于累积量平方匹配准则的系统辨识问题, 因此可采用线性 FIR 系统的基于累积量的参数估计方法^[1-6]估计专家系统参数 $a_n(m)$ 。在获得了参数估计后, 新的参数估计又作为下一次迭代 E 步的条件, 重复以上迭代过程。由于篇幅限制, 两步 EM 算法不再讨论, 详细可参考文献 [8-12]。

在获得专家系统参数后, 可利用 (7) 式 FNN 参数与专家系统参数的关系获得 FNN 参数估计

$$a_n(m) = \sum_{i=1}^I w(i)w(i,m)l(y_{i,r}^0(k)), \quad n = 1, \dots, N$$

其中 $l(y_{i,r}^0(k))$ 为 (4), (5) 式中的已知参数。上述方程中未知参数的个数为: $I \times (M+1) + I$ 。

通常有: $N = K^I > I \times (M+1) + I$, 因此上述方程是超定的。可采用最小均方 (LS) 拟合法估计 FNN 参数。

与传统局部线性近似的参数估计相比较^[13], 新算法具有完全不同的性质。传统局部线性近似处理通常仅在一处做线性近似处理。新算法与此不同, 它在多点 $y_{i,r}^0(k)$, $k = 1, \dots, K$ 以概率方式线性近似 FNN^[8-12], 构成一专家系统竞争学习的方法。只有当 $\Omega = n$ 的概率为 1, 而其它专家系统概率为 0 时, 新算法的线性近似才退化为传统的线性近似处理。因此, 新算法在解决参数估计的局部解、收敛速度等问题方面将优于传统的线性近似处理。由于篇幅限制, 详细不再讨论。

4 简化两阶层混合专家系统及简化算法

从上面讨论可见, 基于累积量的参数估计计算复杂度为 $O(K^I)$ 量级。因此, 当 FNN 隐元数 I 和每个隐元线性近似的数目 K 充分大时, FNN 的参数估计计算复杂度将十分大。

为了简化上述算法, 两层 FNN 可进一步利用简化两阶层混合专家 (Hierarchical ME) 系统来描述^[8,11], 即每个隐元利用一简单 ME 描述构成 HME 的第一层专家系统。每个隐元的输出作为输入, FNN 的输出层构成 HME 的第二层。但是, 与一般两层 HME 不同, 这里的两层 HME 的第二层是确定性的线性求和, 而不是一般 HME 的统计求和。因此, 第二层可采用一般线性系统的参数估计方法获得参数估计。

以下讨论, 我们假设 FNN 隐元输出互为四阶统计独立 (白的)。因此, 简化 HME 是实际 FNN 的一个近似。但通过下面实验分析知道, 当 FNN 的隐元数目足够大时, 基于四阶独立的 FNN 可完全近似实际的 FNN, 因此, FNN 隐元输出互为四阶独立的假设是合理的有效的。

4.1 第一层 ME

在两阶层 HME 中, 第二层专家系统的门限网络是确定性的, 而非统计概率形式的。因此, 第二层可采用一般线性系统的参数估计方法获得参数估计, 我们仅需要考虑第一层网络。第一层网络包含 K 个专家系统, 每个专家系统对应一个隐元及线性近似 (3)-(5) 式。

为了描述第一层 ME, 隐元期望值输出的四阶累积量引入为隐变量, 记为: $C(\tau_{123}) = [c_1(\tau_{123}), \dots, c_I(\tau_{123})]^T$, 其中 $c_i(\tau_{123})$ 表示第 i 隐元输出信号的期望四阶累积量。

基于隐变量, FNN 训练则可分解为多个单隐元累积量的参数估计和线性输出层的参数估计。每个隐元可利用与 (8) 式相似的方式描述。

$$p(c_i(\tau_{123}), \Theta_i | x(t), \vartheta(t)) = \prod_{k=1}^K [\pi_{i,k} q_i \exp(-\lambda_i (c_i(\tau_{123}) - c_{i,k}(\tau_{123}))^2)^{\delta_{i,k}(\tau_{123})}, \quad i = 1, \dots, I \quad (10)$$

其中 λ_i 表示累积量的方差, q_i 为满足统计概率的常数。 $\delta_{i,k}(\tau_{123}) = \begin{cases} 1, & \Theta_i = k \\ 0, & \Theta_i \neq k \end{cases}$, Θ_i 表示以概率 $\pi_{i,k} = \text{prob}(\Theta_i = k)$ 取值于 $\{1, \dots, K\}$ 的指标随机变量, $\Theta_i = k$ 表示第 i 隐元被 (4), (5) 式中第 k 个线性系统近似。 $c_{i,k}(\tau_{123})$ 表示第 i 隐元第 k 个线性系统输出的累积量, 可利用下式计算^[1-6]。

$$c_{i,k}(\tau_{123}) = \sigma_{4x} [l(y_{i,k}^0)]^4 \sum_{m=0}^M w(i, m) w(i, m + \tau_1) w(i, m + \tau_2) w(i, m + \tau_3) \quad (11)$$

基于隐变量, 两层 FNN 的训练已分解为多个单隐元的训练和一个线性输出层的训练。与原始 FNN 训练比较, 分解后的训练肯定更为简单。基于单隐元的统计模型 (10) 式, 单隐元的训练可采用 3.2 节中的算法实现, 即采用 EM 算法估计单隐元^[1-6]。

但是, 隐变量 $C(\tau_{123})$ 是隐含的, 外部不可观测的。必须在单隐元估计以前估计隐变量。

4.2 隐变量 $C(\tau_{123})$ 估计

隐变量 $C(\tau_{123})$ 实际上是 FNN 的中间变量, 即它是输入信号通过隐元后的输出, 又作为输出层的输入信号产生 FNN 的输出。隐变量的估计原则是: 寻找最优隐变量使整体 FNN 的输入、输出最优满足给定的 FNN 输入、输出信号构成的训练集 $c_d(\tau_{123}), \sigma_{4x}$ 。

为了寻找最优或最可能的隐变量, 需要建立条件概率 $p(C(\tau_{123}) | c_d(\tau_{123}), \sigma_{4x}, \vartheta)$ 。首先考虑

$$p(c_d(\tau_{123}), C(\tau_{123}) | \sigma_{4x}, \vartheta) = p(c_d(\tau_{123}) | C(\tau_{123}), \sigma_{4x}, \vartheta(t)) p(C(\tau_{123}) | \sigma_{4x}, \vartheta(t)) \quad (12)$$

其中 $p(c_d(\tau_{123}) | C(\tau_{123}), \sigma_{4x}, \vartheta(t))$ 取形式⁴:

$$p(c_d(\tau_{123}) | C(\tau_{123}), \sigma_{4x}, \vartheta(t)) = p(c_d(\tau_{123}) | C(\tau_{123}), \vartheta(t)) = B_1 \exp\left(-\frac{1}{2} E_1(\tau_{123})\right) \quad (13)$$

其中 $E_1(\tau_{123}) = \sigma_1 [c_d(\tau_{123}) - |W(t)|^4 C(\tau_{123})]^2$, σ_1 表示方差。 $|W(t)|^4 = [|w(1)|^4, \dots, |w(I)|^4]^T$ 。
 B_1 为常数且满足条件概率约束。

⁴ 显然 (13) 式是基于隐元输出为四阶统计独立 (白)。另外, 取高斯密度意义与 (8) 式相同。

给定输入信号 σ_{4x} 的隐变量 $C(\tau_{123})$ 的条件概率取为

$$p(C(\tau_{123})|\sigma_{4x}, \vartheta(t)) = B_2 \exp\left(-\frac{1}{2}[C(\tau_{123}) - C_h(\tau_{123})]^T \Sigma_2^{-1}[C(\tau_{123}) - C_h(\tau_{123})]\right) \quad (14)$$

其中 B_2 与 B_1 相似, 为一常数. Σ_2 为协方差, $C_h(\tau_{123}) = [c_{h1}(\tau_{123}), \dots, c_{hI}(\tau_{123})]^T$, 及 $c_{hi}(\tau_{123})$ 表示第 i 隐元输出的四阶累积量, 可利用下式计算:

$$c_{hi}(\tau_{123}) = \sum_{k=1}^K p_{i,k}(t) c_{i,k}(\tau_{123}) \quad (15)$$

其中 $p_{i,k}(t)$ 表示 $\Theta_i = k$ 的概率, 即第 i 隐元被第 k 个线性系统近似的概率. 该概率可利用 EM 算法的 E 步估计. 四阶累积量 $c_{i,k}(\tau_{123})$ 取 (11) 式.

除了上述条件概率密度 (13), (14) 式外, 给定 σ_{4x} 的 $c_d(\tau_{123})$ 条件概率密度取为

$$p(c_d(\tau_{123})|\sigma_{4x}, \vartheta(t)) = B_3 \exp\left(-\frac{1}{2}\sigma_3[c_d(\tau_{123}) - c_y(\tau_{123})]^2\right) \quad (16)$$

基于 (13), (14), (16) 式, 关于给定训练集的隐变量条件概率密度可推导为

$$p(C(\tau_{123})|c_d(\tau_{123}), x_m(t), \vartheta(t)) = \frac{p(c_d(\tau_{123}), C(\tau_{123})|x_m(t), \vartheta(t))}{p(c_d(\tau_{123})|x_m(t), \vartheta(t))} = B_h \exp\left(-\frac{1}{2}E_{C,h}(\tau_{123})\right) \quad (17)$$

其中 B_h 为满足条件概率的常数, $E_{C,h}(\tau_{123}) = [C(\tau_{123}) - \bar{C}(\tau_{123})]^T \Sigma_{C,h}^{-1}[C(\tau_{123}) - \bar{C}(\tau_{123})]$, $\bar{C}(\tau_{123}) = [\bar{c}_1(\tau_{123}), \dots, \bar{c}_I(\tau_{123})]^T$ 是隐变量 $C(\tau_{123})$ 的条件期望值. 根据 (13), (14), (16), (17) 式, 可推导出条件期望值为

$$\bar{C}(\tau_{123}) = C_h(\tau_{123}) + \left[\Sigma_2^{-1} + \sigma_1 [|W(t)|^4]^T [|W(t)|^4]\right]^{-1} \sigma_1 |W(t)|^4 [c_d(\tau_{123}) - |W(t)|^4 C_h(\tau_{123})] \quad (18)$$

根据概率密度 (17) 式, 条件期望 $\bar{C}(\tau_{123})$ 实际上就是给定观测集 $c_d(\tau_{123})$, σ_{4x} , $\vartheta(t)$ 的隐变量 $C(\tau_{123})$ 的最优估计, 即使 FNN 满足输入四阶累积量为 σ_{4x} 和期望值输出的四阶累积量为 $c_d(\tau_{123})$ 的隐变量最优估计. (18) 式中的参数 σ_1 和协方差矩阵 Σ_2 可利用时间平均估计, 如

$$\sigma_1 = \hat{\sigma}_1(t) = \frac{1}{P} \sum_{t'=t-P+1}^t [c_d(\tau_{123}) - |W(t')|^4 C_h(\tau_{123})]^2$$

其中 P 为一充分大正整数.

从上面讨论, 概率密度 (13) 式实际上规定了 FNN 输出层的输入、输出关系. 因此 FNN 输出层可根据 (13) 式估计. 需要指出的是, 上述隐变量估计和 FNN 参数估计是一交替迭代的过程, 可简单总结为: 给定初始猜值 $\vartheta(0)$, $t = 0$ 重复如下迭代

步骤 1 基于 $\vartheta(t)$, 利用 (18) 式估计隐变量 $\bar{C}(\tau_{123})$.

步骤 2 利用隐变量估计 $\bar{C}(\tau_{123})$, 根据 (10) 式估计 FNN 单隐元参数, 利用 (13) 式估计输出层参数. 因此获得 $\vartheta(t+1)$.

5 计算机模拟实验

采用数字通信盲均衡器的实验来验证本文新算法。设实际信道输出为

$$d(t) = G \left(\sum_{m=0}^3 b_m(t)x(t-m) - 1.5 \right) + n(t) \quad (19)$$

其中 $b_m(t)$ 为信道参数, $G(s) = 2s/(1+s^2)$ 为信道幅度饱和畸变算子, $n(t)$ 为加性高斯噪声。信道的传输序列 $x(t)$ 取为统计独立、同分布 (iid) 的 4 电平脉冲幅度调制 (4-PAM) 序列, 即 $x(t)$ 以等概率取值于 $\{0, 1, 2, 3\}$ 。这一模型描述了实际工程中的数字卫星通信及磁记录等的实际信道。

在盲均衡中, 假定传输序列 $x(t)$ 未知, 但可根据 4-PAM 的统计信息获得 σ_{4x} 。实验中, 采用两层 FNN 来近似非线性系统 (19) 式。FNN 的输入数目 $M = 4$, 隐元输入限定为 $[-6, 6]$, 每个隐元被以零点为中心的 $K = 15$ 个线性系统近似。

实验图 1 为传输序列、接收信号、均衡器输出及检测判决信号。其中 (c) 为均衡器 (两层 FNN 的逆系统) 的输出, (d) 为根据 (c) 的判决信号, 即: $\hat{x}(t) = \arg \min_{x(t) \in \{0, 1, 2, 3\}} |\hat{G}^{-1}[d(t)] - x(t)|$ 。

图 2 和图 3 分别为算法的学习曲线, 它们是 40 次独立的模拟实验结果 $|d(t) - y(t)|^2$ 的平均。

从图 2、图 3 可看出新的训练算法可有效均衡非线性信道, 恢复传输序列。从实验图也可看出, 当 FNN 的隐元数目足够大时, 隐元输出为四阶独立的假设是合理的。另外, 与后向传播 (BP) 训练算法相比, 新算法的收敛速度提高约 10-15 倍。

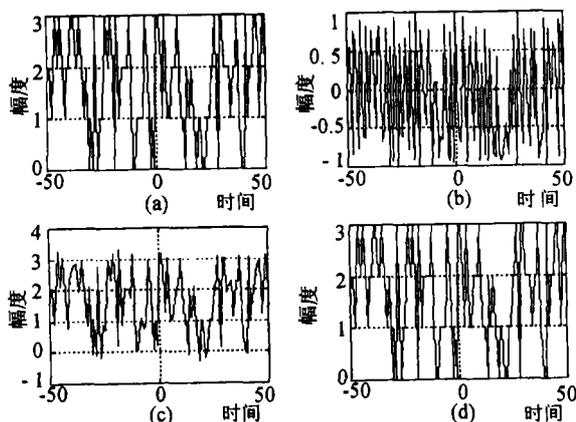


图 1 传输序列和均衡恢复序列 (a) 传输序列, (b) 信道输出, (c) 两层 FNN 逆系统的输出, (d) 根据 (c) 输出判决恢复的序列

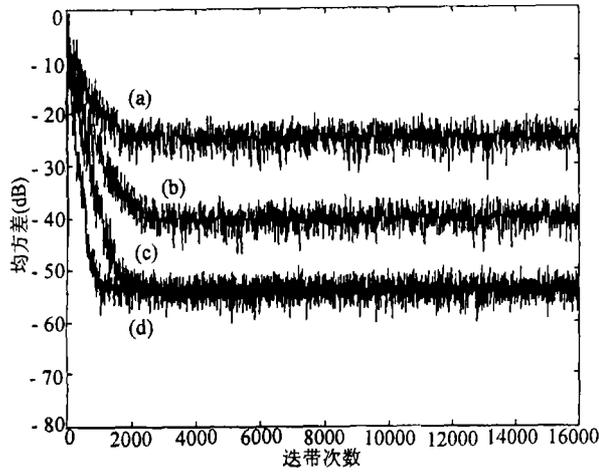


图 2 基于四阶累积量的 FNN 学习曲线 (a)6 个隐元 HME 的学习曲线, (b)8 个隐元 HME 的学习曲线, (c)12 个隐元 HME 的学习曲线, (d) 利用 ME(7),(8) 式的 6 个隐元 FNN 的学习曲线

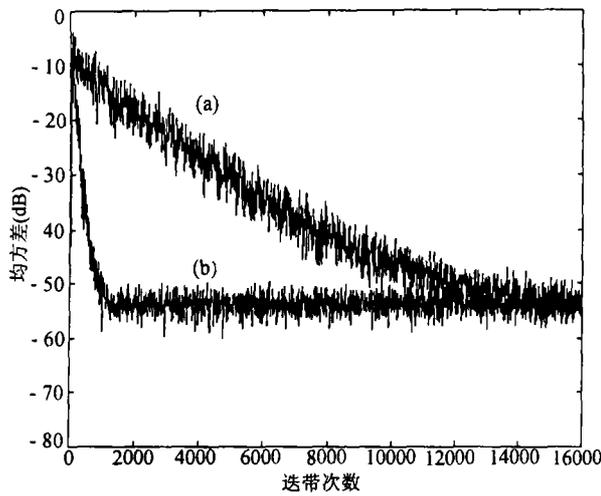


图 3 基于四阶累积量的 FNN 学习曲线和基于 BP 算法的学习曲线 (a)BP 算法学习曲线, (b)新算法学习曲线

6 结论与评论

提出了两种基于累积量的 FNN 训练方法。由于 FNN 隐元的线性近似, 使复杂的 FNN 输出累积量转化为线性系统的累积量, 这使得基于累积量的 FNN 训练成为可能。在新的 EM 训练算法中, M 步参数估计实际上就是线性系统基于累积量的训练, 因此, 线性系统基于累积量的训练方法可直接用来估计 FNN 参数。另外, 新的训练方法也可用于估计其它基于块的非线性系统, 如: 三明治非线性系统、Wiener 系统和 Hammerstein 系统等。

参 考 文 献

- [1] J. M. Mendel, Tutorial on high-order statistics (spectra) in signal processing and system theory, theoretic results and some applications, Proc. IEEE, 1991, 79, 278-305.
- [2] Hames A. Cadzow, Blind deconvolution via cumulant extrema, IEEE Signal Processing Magazine, 1996, 13(3), 24-42.
- [3] D. Hatzinakos, C. L. Nikias, Blind equalization using a tricespectrum-based algorithm, IEEE Trans. on Comm., 1991, COM-39(5), 669-681.
- [4] B. Port, B. Ftiedlander, Blind equalization of digital communication channels using higher-order moments, IEEE Trans. on ASSP, 1991, ASSP-39(2), 522-526.
- [5] J. K. Tugnait, Identification of linear stochastic systems via second and fourth-order cumulants matching, IEEE Trans. on Information Theory, 1987, 33(5), 393-407.
- [6] Jitendra K. Tugnait, Blind equalization and channel estimation with partial response input signals, IEEE Trans. on Comm., 1997, COM-45(9), 1025-1442.
- [7] Vojin, Zivojnic, Minimum fisher information of moment-constrained distributions with application to robust blind identification, Signal Processing, 1998, 65, 297-313.
- [8] Michael I. Jordan, Lei Xu, Convergence results for the EM approach to mixtures of experts architectures, Neural Networks, 1995, 8(9), 1409-1431.
- [9] S. I. Amari, Information geometry of the EM and EM algorithms for neural network, Neural Networks, 1995, 8(9), 1379-1408.
- [10] Sheng Ma, James Farmer, An efficient EM-based training algorithm for feedforward neural networks, Neural Networks, 1997, 10(2), 243-256.
- [11] M. I. Jordan, Hierarchical mixtures of experts and EM algorithm, Neural Computation, 1994, 6(2), 181-241.
- [12] R. A. Jacobs, Adaptive mixtures of local experts, Neural Computation, 1991, 3(1), 79-87.
- [13] B. Widrow, 30 years of adaptive neural networks, perceptron, madaline, and backpropagation, Proc. IEEE, 1990, 78(9), 1415-1442.

CUMULANT-BASED BLIND IDENTIFICATION OF
TWO-LAYER FEEDFORWARD NEURAL NETWORKS

Dai Xianhua

(Department of Electronic Engineering, Shantou University, Shantou 515063, China)

Abstract It is always difficult to train a two-layer Feedforward Neural Networks (FNN) based on the cumulant match criterion because cumulants are nonlinear implicit functions of the FNN parameters. In this work, two new cumulant-based training methods for two-layer FNN are developed. In the first method, the hidden units of two-layer FNN are approximated with multiple linear systems, and further total FNN is modeled with a "mixture of experts" (ME) architecture. With the ME model, FNN parameters are estimated with expectation-maximization (EM) algorithm. The second method, for simplifying the two-layer FNN statistical model, proposes a simplified two-level hierarchical ME to remodel the FNN, in which hidden variables are introduced such that training total FNN is changed to training a set of single neurons. Based on training single neuron, total FNN is estimated in a simplified version with a faster convergence speed.

Key words ME model, EM algorithm, Cumulant-based system identification

戴宪华: 男, 1964年生, 博士, 1993-1994华南理工大学博士后, 教授, 主要研究兴趣: 非线性自适应信号处理、盲信号处理、通信信道均衡等。