

利用 EM 算法估计隐含观测量的反馈神经网络学习新方法

戴宪华

(汕头大学电子工程系, 汕头 515063)

摘 要: 研究反馈神经网络(RNN)参数估计的新方法. 利用隐含观测量, 将复杂 RNN 的训练分解为线性输出层和多个单隐元的参数估计. 基于每个隐元激励函数的多点线性近似, RNN 可利用统计混合专家网络模型(ME)描述, 从而将 RNN 的参数估计转化为包含隐含观测量的线性系统的最大似然估计问题, 最后利用期望最大化(EM)算法获得 RNN 的隐含观测量及其参数估计.

关键词: 反馈神经网络; EM 算法; 隐含观测量

中图分类号: TP183 **文献标识码:** A **文章编号:** 0372-2112 (2000) 10-0133-05

A New Method for Training RNN via Hidden Representation Estimated by EM Algorithm

DAI Xian-hua

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

Abstract: A new method for training recurrent neural network (RNN) has been proposed. By introducing the hidden representation or hidden variables into RNN, training the complicated RNN is decomposed into training a set of single neurons and a linear output layer. Based on linear approximation of RNN hidden units, RNN is remodeled with a "mixture of experts" (ME) model. Moreover, training RNN is also changed into a maximum likelihood estimation of the linear systems with hidden variables. Finally, training RNN is fulfilled with the expectation-maximization (EM) algorithm.

Key words: recurrent neural networks; EM algorithm; hidden variables

1 引言

神经网络(NN)的研究及应用为信息处理问题提供了一高性能的解决途径. 优良的 NN 处理性能很大程度上取决于 NN 的自学习能力. 但是, 对于实际的应用问题, 获得良好的收敛于全局最优的训练学习通常是十分困难的^[1,9].

与简单线性系统相比, RNN 的参数估计更为困难. 困难之一, 参数估计是一非线性优化问题. RNN 隐元为一非线性映射关系, 使得 RNN 的实时参数估计收敛分析十分复杂^[1]. 困难之二, 是给定的 RNN 训练数据(RNN 输入、输出关系)仅规定了 RNN 的整体输入输出映射关系. 但由于 RNN 是一个十分复杂的系统, 它通常由多个子系统(隐元)构成. 给定的训练数据对 RNN 系统内部的子系统(各个隐元)的输入、输出映射关系则没有直接规定或描述. 这些隐元的期望输入、输出映射关系通常称为隐含观测量^[2~5].

文献[2]最先提出利用隐表示或隐变量来训练复杂系统, 以后文献[3~5]又作了进一步发展. 目的是利用隐变量分解复杂系统为多个简单系统, 从而简化复杂系统的训练. 统计 ME 模型是一具有较快收敛速度的新型系统模型, 它与统计 EM 算法相结合可以获得较快的收敛速度与良好的收敛性能^[3~7]. 因此, 一个自然的想法是将隐变量或隐表示与统计

ME 模型、EM 算法结合, 提出改进的 RNN 训练算法.

本文基于以上思想, 从统计学角度研究 RNN 系统模型, 利用隐变量或隐表示与统计 ME 模型、EM 算法的结合, 分解 RNN 为多个简单子系统, 推导出包含隐含观测量的似然函数, 最后利用期望最大优化法(EM)获得 RNN 隐含观测量及参数估计.

2 RNN 隐元的全局线性化及统计模型

考虑如下具有单个隐层、多个隐元的 RNN 系统,

$$y(t) = \sum_{i=1}^I \hat{w}_i(t) y_i(t) = \sum_{i=1}^I \hat{w}_i(t) g \left[\sum_{n=1}^N \hat{a}_{i,n}(t) y(t-n) + \sum_{m=0}^M \hat{b}_{i,m}(t) x(t-m) \right] \quad (1)$$

该模型可参看文[1], 它是基于非线性 ARMA 系统的 RNN. 其中 $x(t)$, $y(t)$ 分别为 RNN 输入输出信号, $y_i(t)$ 为 RNN 第 i 隐元输出. $\hat{w}_i(t)$, $\hat{a}_{i,n}(t)$, $\hat{b}_{i,m}(t)$ 分别为 RNN 的待估计参数. RNN 参数估计的目的就是估计并调整参数 $\hat{w}_i(t)$, $\hat{a}_{i,n}(t)$, $\hat{b}_{i,m}(t)$ 使 RNN 的输出信号逼近、跟踪期望信号 $d(t)$, 也称为指导信号. 这里 $d(t) = d(t) + n(t)$, 其中 $d(t)$ 为期望信号, $n(t)$ 为加性噪声, 假定它满足高斯分布. 式(1)中的 $g(\cdot)$ 为 RNN 激励函数, 取为 Sigmoid 函数.

从上模型可见, RNN 的训练集为输入信号 $x(t)$ 和期望输出 $d(t)$. 对于 RNN 复杂系统, 给定的训练集仅规定了 RNN 整体的输入、输出关系, 并没有直接规定 RNN 内部各隐元的输入和输出关系, 即这些约束关系是隐含的. 但对于 RNN 系统式(1), 若能给出各隐元的输出关系使之最佳地满足 RNN 整体的输入输出关系, 那么 RNN 的参数估计将转化为各个单隐元的参数估计, 这一思想对于简化系统参数估计十分有益^[3~5]. 基于此, 本文引入各隐元的期望输出信号作为隐含观测量, 用于简化系统参数估计. 设各隐元的期望输出信号(也称输出指导信号)分别为 $z_1(t), \dots, z_I(t)$. 这些信号为隐含未知的, 估计参数时也需要估计这些隐含观测量.

基于隐含观测量 $z_1(t), \dots, z_I(t)$, RNN 的参数估计被分解为多个单隐元和一个线性输出层的参数估计问题. 由于 $g(\cdot)$ 为非线性算子, 那么单隐元的参数估计仍是一非线性优化问题, 它的实时参数估计算法的均方误差曲面为超二次, 基于梯度下降方向的实时参数估计算法不能确保收敛于全局最优解. 为解决这一问题, 连续激励函数在充分小的区域可利用它的一阶 Taylor 展开式近似. 因此, 式(1)的单个隐元可用下面一系列线性系统近似.

$$y_i(t) = \sum_{k=1}^K y_{i,k}(t) \quad (2)$$

$$y_{i,k}(t) = g(y_{i,k}^0) + \left. \frac{\partial g(y_{i,k}(t))}{\partial y_{i,k}(t)} \right|_{y_{i,k}(t)=y_{i,k}^0} [y_{i,k}(t) - y_{i,k}^0] \\ = g(y_{i,k}^0) + l(y_{i,k}^0) [y_{i,k}(t) - y_{i,k}^0] \quad (3)$$

$$y_{i,k}(t) = \sum_{n=1}^N \hat{a}_{i,n}(t) y(t-n) + \sum_{m=0}^M \hat{b}_{i,m}(t) x(t-m) - i \quad (4)$$

其中 $y_{i,k}^0$ 为 $g(\cdot)$ 定义域中的固定点, $g(y_{i,k}^0)$, $l(y_{i,k}^0)$ 分别为 $g(\cdot)$ 在 $y_{i,k}^0$ 点的函数值和一阶导数, 它们是固定已知的. i, k 表示 RNN 第 i 个隐元 $g(\cdot)$ 定义域中以 $y_{i,k}^0$ 为中心的小邻域, 并且 $A_{i,k} = A_i$, A_i 为 RNN 的第 i 个隐元 $g(\cdot)$ 算子的定义域或输入空间. k 称为 RNN 的状态指标.

从式(2)~(4)可看出, RNN 的每个隐元都利用多个线性系统近似. 根据连续函数的性质, 当线性近似的数目 K 充分大时, 式(2)~(4)描述的线性系统将完全可以近似式(1), 近似误差随着 K 的增大而趋于零. 这样, 就可以将非线性 RNN 的参数估计转化为一组线性 ARMA 系统的参数估计问题^[11]. 但是, 这里的近似并不是多个线性 ARMA 系统的线性并联, 而是多个线性系统以概率求和的形式来近似该隐元. 每个线性系统定义于一小区域 $A_{i,k}$ 解决一函数近似问题. 实质上, RNN 隐元的多线性近似就是一统计混和专家网络(MEN), 式(2)~(4)中的每个线性系统称为一个专家系统^[3~5]. 因此, 可利用 MEN 重新描述 RNN 单隐元, 进而将 RNN 参数估计转化为统计学习问题.

基于 MEN 模型描述^[3~5], 第 i 隐元的统计模型为(这里取为高斯(混合)密度是由于它与通常的均方误差参数估计算法(如 LMS 或 LS)一致, 以下同)

$$p(z_i(t) | x(t), y_i(t)) = \prod_{k=1}^K [B_{i,k} \exp(-y_{i,k} E_{i,k})]^{i,k} \quad (5)$$

有关分布函数的详细情况可参考文献[3~5]. 式(5)中的 i 表示取值于 $\{1, \dots, K\}$ 的离散随机变量, $i=k$ 表示第 i 隐元为其第 k 个专家近似. 由于每个隐元利用 K 个线性系统近似, 那么就需要确定期望信号(隐变量)可作为每个线性系统期望信号的概率, 因此概率 $\text{prob}(i=k)$ 是一新的隐含观测量. $E_{i,k} = [z_i(t) - y_{i,k}(t)]^2$, $i, k = \begin{cases} 1, & i=k \\ 0, & \text{Otherwise} \end{cases}$. $B_{i,k}$ 为概率分布函数的常数, 并且满足概率分布参数的约束. $y_i(t) = [\hat{a}_{i,1}(t), \dots, \hat{a}_{i,N}(t), \hat{b}_{i,0}(t), \dots, \hat{b}_{i,M}(t)]^T$.

基于 MEN 式(5), 第 i 隐元的输出可写为,

$$y_i(t) = \sum_{k=1}^K y_{i,k}(t) \text{prob}(i(t)=k) \quad (6)$$

其中, $\text{prob}(i=k)$ 表示第 k 个专家近似第 i 隐元的概率, 与 $z_i(t)$ 一样是隐变量.

3 RNN 的统计描述——包含隐含观测量的似然函数

RNN 系统式(1)是一确定性系统的描述. 为了利用统计学方法估计 RNN 参数, 须利用统计系统模型重新表示 RNN. 考查 RNN 输出期望信号依赖于隐含观测量的概率描述, 有

$$p(d(t) | z_1(t), \dots, z_I(t), W(t)) = B \exp(-E) \quad (7)$$

其中: $E = [d(t) - Z^T(t)W(t)]^2$, $W(t) = [\hat{w}_1(t), \dots, \hat{w}_I(t)]^T$, $Z(t) = [z_1(t), \dots, z_I(t)]^T$.

将 RNN 参数 $y_i(t)$, $W(t)$, $i=1, \dots, I$ 统一记为矢量 (t) . 根据 RNN 系统可进一步假定:

$$p(d(t) | z_1(t), \dots, z_I(t), y_1(t), \dots, y_I(t), (t)) \\ = p(d(t) | z_1(t), \dots, z_I(t), (t)) \quad (8)$$

根据 RNN 的输出层关系, 在 RNN 收敛时系统输出仅依赖于每个隐元的输出, RNN 整体的期望导信号或训练数据也仅依赖于各隐元的期望信号, 因此假设式(8)通常是合理^[13].

根据 RNN 模型式(1)、线性近似式(2)~(4)及单隐元输出式(6), 有

$$p(Z(t) | y_1(t), \dots, y_I(t), x(t), (t)) = A \exp(-\sum_{i=1}^I (Z(t) - Y_i(t))^T (Z(t) - Y_i(t))) \quad (9)$$

其中: $Y_i(t) = [y_{i,1}(t), \dots, y_{i,I}(t)]^T$, $y_i(t)$ 取式(6). Σ_i 为协方差矩阵.

根据以上推导可以获得关于 RNN 观测量(包括隐含观测量的)联合概率分布为

$$p(d(t), Z(t), y_1(t), \dots, y_I(t) | x(t), (t)) \\ = p(d(t) | Z(t), y_1(t), \dots, y_I(t), x(t), (t)) p(Z(t) | y_1(t), \dots, y_I(t), x(t), (t)) p(y_i(t) | x(t), (t)) \\ = B \exp(-E) A \exp(-\sum_{i=1}^I (Z(t) - Y_i(t))^T (Z(t) - Y_i(t))) p(y_i(t) | x(t), (t)) \quad (10)$$

对应的对数似然函数为:

$$Q(d(t), Z(t), y_1(t), \dots, y_I(t) | x(t), (t)) = -E - \sum_{i=1}^I (Z(t) - Y_i(t))^T (Z(t) - Y_i(t))$$

$$- Y_1(t))^T \left[Z(t) - Y_1(t) + \sum_{i=1}^I p(i|x(t), \bar{Z}(t)) + C \right. \\ \left. + W(t) W^T(t) J(Z(t) - \bar{Z}(t)) \right] \quad (11)$$

式(11)即为 RNN 的统计模型,它以统计概率密度的方式约束了 RNN(包括各个隐元)输入、输出关系.从式(11)也可看出,RNN 参数估计已分解为输出层和隐元的参数估计两部分.

4 RNN 参数估计

上一节利用似然函数推导出了 RNN 的统计模型描述.但是,由于似然函数式(11)包含有隐含观测量,因此不可以直接利用最大似然算法估计参数.在统计估计包含隐含观测量时,有效的估计方法为 EM 方法^[3~5].为方便记: $u(t) = [d(t), Z(t), \dots, Y_1(t)]^T = [u_{\text{obs}}, u_{\text{mis}}]$, 其中 $u_{\text{obs}}, u_{\text{mis}}$ 分别为可观测量和隐含观测量,那么 EM 算法可叙述为:

给定初始参数 $\bar{Z}(0)$, $t=0$ 重复如下操作.

E-运算: 利用首次 M-运算的参数估计,估计隐变量的条件概率,来对数似然函数的期望

$$Q(\bar{Z}|(t)) = E_{u_{\text{mis}}} [Q(u_{\text{obs}}, u_{\text{mis}} | (t))] = \int p(u_{\text{mis}} | u(t), (t)) Q(u_{\text{obs}}, u_{\text{mis}} | (t)) du_{\text{mis}} \quad (12)$$

M-运算: 似然函数期望的最大化,求参数估计,

$$\bar{Z}(t+1) = \arg \max_{\bar{Z}} Q(\bar{Z} | (t)) \quad (13)$$

即由原始的 MLA 转化为 EM.下面,分别讨论 EM 算法的 E-运算和 M-运算.

4.1 RNN 隐含观测量估计 (E-运算)

上面的似然函数式(11)包含了两类隐含观测量,一是各隐元线性化后的状态隐含观测量或概率 $\text{prob}(i=k)$ (由于该概率隐含,似然函数式(11)中的 $Y_1(t)$ 亦是隐含的),二为各隐元的期望输出信号 $z_1(t), \dots, z_I(t)$. 可观测量 u_{obs} 为:训练集 $x(t)$ 和 $d(t)$. 首先,讨论各隐元的期望信号 $Z(t)$ 对应的似然函数期望.

隐变量估计的原则是:寻找 $Z(t)$ 的最优估计使得 RNN 模型式(1)的整体输入输出关系最佳地满足给定的可观测量 $x(t)$ 和 $d(t)$. 根据统计学分析,为了寻找 $Z(t)$ 的最优估计必须首先确定隐含观测量的条件分布函数 $p(Z(t) | d(t), x(t), \bar{Z}(t))$.

首先,考虑给定隐元输出时的 $Z(t)$ 条件概率

$$p(Z(t) | x(t), \bar{Z}(t)) = A_2 \exp \left(- \frac{1}{2} (Z(t) - Y_2(t))^T \Sigma_2^{-1} (Z(t) - Y_2(t)) \right) \quad (14)$$

其中, $Y_2(t) = [y_1(t), \dots, y_I(t)]^T$, $y_i(t)$ 取式(1)中的变量, Σ_2 为协方差矩阵.

利用概率密度递推公式,可推出,

$$p(Z(t) | d(t), x(t), \bar{Z}(t)) = \frac{p(d(t), Z(t) | x(t), \bar{Z}(t))}{p(d(t) | x(t), \bar{Z}(t))} \\ = \frac{p(d(t) | Z(t), x(t), \bar{Z}(t)) p(Z(t) | x(t), \bar{Z}(t))}{p(d(t) | x(t), \bar{Z}(t))} \quad (15)$$

假定: $p(d(t) | x(t), \bar{Z}(t)) = B_3 \exp(-\frac{1}{2} [d(t) - y(t)]^2)$, 将式(7)和(14)代入式(15),有

$$p(Z(t) | d(t), x(t), \bar{Z}(t)) = H \exp \left(- \frac{1}{2} (Z(t) - \bar{Z}(t))^T \left[\Sigma_2^{-1} + \frac{1}{H} \right] (Z(t) - \bar{Z}(t)) \right) \quad (16)$$

其中: $\bar{Z}(t) = [\bar{z}_1(t), \dots, \bar{z}_I(t)]^T = Y_2(t) + [I - W(t) W^T(t)]^{-1} W(t) (d(t) - Y_2(t))$.

实际上, $\bar{Z}(t)$ 就是满足 RNN 整体输入、输出关系的隐含观测量 $Z(t)$ 的最优估计.从式(16)可以直观看出隐含测量估计的物理意义:隐元期望信号估计等于隐元输出信号 $Y_2(t)$ 加上一修正项,该修正项与整体网络的期望输出信号减去其它隐元期望信号的差信号成正比.另外,当 RNN 指导信号不包含加性噪声而且 RNN 已经全局收敛(隐元输出信号收敛于真实期望信号)时,隐元期望信号估计式(16)中的修正项为零,该期望(隐含)信号就等于隐元的输出信号.因此,期望信号估计的物理意义是十分明确的,直观易于理解,并且是合理的.

在求出条件分布函数以后, $z_1(t), \dots, z_I(t)$ 对应的似然函数期望为,

$$Q(\bar{Z} | (t)) = \int \dots \int Q(d(t), Z(t), \dots, Y_1(t) | x(t), \bar{Z}(t), W(t)) p(Z(t) | d(t), x(t), \bar{Z}(t)) dZ(t) \\ = - \left[d(t) - \sum_{i=1}^I \hat{w}_i(t) \bar{z}_i(t) \right]^2 - (\bar{Z}(t) - Y_1(t))^T \Sigma_1^{-1} (\bar{Z}(t) - Y_1(t)) + \sum_{i=1}^I p(i | x(t), \bar{Z}(t)) + C \quad (17)$$

其中 C 包含了积分中产生的方差常数.

在求出 $Z(t)$ 对应的似然函数期望后,下面考虑隐含观测量 i 对应的似然函数期望.由于已经获得隐元的期望输出估计 $\bar{z}_i(t)$, $\bar{z}_i(t)$ 可作为已知信号去估计隐含观测量 i .

由于 i 是一离散随机变量,因而,对应的似然函数期望是一线性求和的形式,

$$Q(\bar{Z} | (t)) = - \left[d(t) - \sum_{i=1}^I \hat{w}_i(t) \bar{z}_i(t) \right]^2 - (\bar{Z}(t) - Y_1(t))^T \Sigma_1^{-1} (\bar{Z}(t) - Y_1(t)) \\ + \sum_{i=1}^I \sum_{k=1}^K p_{i,k}(t) + C \quad (18)$$

其中: $Y_1(t) = [\hat{y}_1(t), \dots, \hat{y}_I(t)]^T$, $\hat{y}_i(t) = \sum_{k=1}^K y_{i,k}(t) p_{i,k}(t)$. $p_{i,k}(t)$ 表示 $\bar{z}_i(t)$ 作为第 k 个线性近似系统(专家)期望信号的条件概率.根据式(5),该概率可由下式求得^[3~7],

$$p_{i,k}(t) = \text{prob}(i=k | \bar{z}_i(t), x(t), \bar{Z}(t)) \\ = \frac{B_{i,k} \exp \left(- \frac{1}{2} (\bar{z}_i(t) - y_{i,k}(t))^2 \right)}{\sum_{j=1}^K B_{i,j} \exp \left(- \frac{1}{2} (\bar{z}_i(t) - y_{i,j}(t))^2 \right)} \quad (19)$$

与似然函数式(11)比较,似然函数期望式(18)已是一完备观测量的似然函数.

4.2 RNN 参数最大似然估计 (M-运算)

在获得对数似然函数的期望后,就可以利用似然函数期望最大化算法估计 RNN 的参数.

首先讨论参数 $W(t)$ 的估计.根据式(18), $W(t)$ 的最优估计满足

$$\max_{W(t)} Q(\bar{Z}(t) | W(t)) = \max_{W(t)} \left[d(t) - \sum_{i=1}^I \hat{\Phi}_i(t) \bar{z}_i(t) \right]^2 \quad (20)$$

那么实时递归的自适应参数估计为

$$W(t+1) = W(t) + \mu [d(t) - \bar{Z}^T(t) W(t)] \bar{Z}(t) \quad (21)$$

其中矢量 $\bar{Z}(t) = [\bar{z}_1(t), \dots, \bar{z}_I(t)]^T$.

在获得了 RNN 输出层参数 $W(t)$ 的估计以后, 根据似然函数式(18), 隐元参数估计为,

$$\min (\bar{Z}(t) - \gamma_1(t))^T \gamma_1(t) \quad (22)$$

由于 $\gamma_1(t) = [\hat{\gamma}_1(t), \dots, \hat{\gamma}_I(t)]^T$, $y_i(t) = \sum_{k=1}^k y_{i,k}(t) p_{i,k}(t)$ 是关于参数的线性函数, 因此参数估计式(22)实际上是一具有指导输出信号的多输入、多输出线性系统的参数估计问题, 可采用相应的线性系统算法估计.

另外, 假定多个隐元输出相互独立, 每个隐元也可单独估计. 以第 i 隐元为例, 在参数 $\gamma_i(t)$ 的估计中, 每个隐元的期望信号估计式(15)中的 $\bar{z}_i(t)$, $i = 1, \dots, I$ 可以对应隐元输出的指导信号. 因此, 每个隐元已经成为具有规定输入、输出关系的单隐元参数估计问题, 即将复杂多隐元 RNN 隐层参数估计转化为多个具有规定输入、输出关系的单隐元参数估计问题.

单隐元参数估计采用与 $W(t)$ 估计相同的处理方式, 可以推出 $\gamma_i(t)$ 的参数估计,

$$\gamma_i(t+1) = \gamma_i(t) + \mu [\bar{z}_i(t) - y_i(t)] Y(t) \quad (23)$$

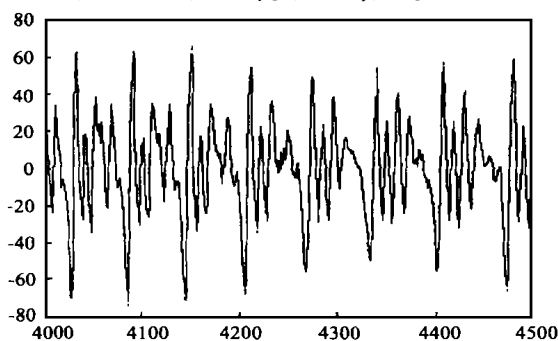


图1 利用 RNN 预测器对语音信号预测处理实验结果, 其中实线为原始语音信号, 虚线为 RNN 预测输出信号

$$Y(t) = \begin{bmatrix} p_{i,k} l(y_{i,k}^0) y(t-1), \dots, p_{i,k} l(y_{i,k}^0) y(t-N), \\ p_{i,k} l(y_{i,k}^0) x(t), \dots, p_{i,k} l(y_{i,k}^0) x(t-M) \end{bmatrix}^T$$

与线性 IIR 滤波相比, 显然 RNN 第 i 隐元参数 $\gamma_i(t)$ 的估计公式与一般的线性 IIR 滤波的预测误差算法参数估计^[11]完全相同. 利用 EM 算法估计隐含信号及隐元线性化可以将非线性 RNN 的参数估计转化为线性 IIR 系统的参数估计问题. 线性 IIR 滤波的各种参数估计方法可直接应用于 RNN 参数的估计. 与传统训练算法^[1,9,12]相比, 上述算法更为简化.

上述统计概率模型中的其它参数 γ_i 和 $B_{i,k}$, B 都可采用一般统计学的批处理方法获得其估计. 由于篇幅限制, 不再讨论.

5 计算机实验

本文实验以语音信号为实验对象, 利用 RNN 实现语音信号的 ADPCM 编码来验证本文提出算法的可行性. 实验中分别采用 RNN 与线性预测器对语音信号处理. 在本文的实验中, RNN 模型规模取为 $I = 3$, $N = 2$, $M = 3$, 共用模型参数 18 个, 语音线性预测器系数数目也取为 18 个, 因此两种处理的模型参数数目相同. RNN 模型的激励函数以零点为中心对称作线性近似处理, 线性化数目取 13.

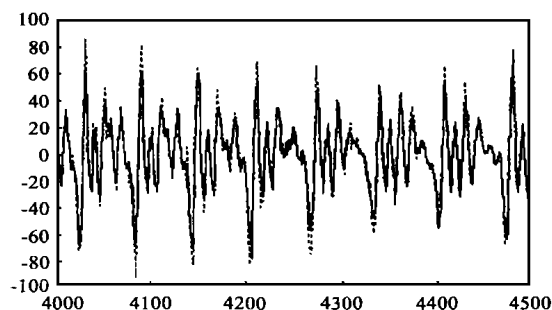


图2 利用线性预测器对语音信号预测处理实验结果, 实线为原始语音信号, 虚线为线性预测输出信号

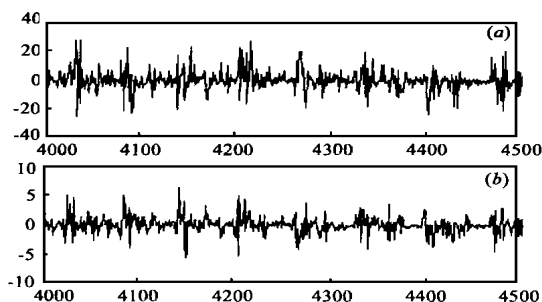


图3 语音信号预测的误差信号, (a) 为线性预测误差信号, (b) 为 RNN 预测误差信号

图1 为 RNN 预测器预测输出信号, 图2 为线性预测器预测输出信号, 图3 为 RNN 预测器预测和线性预测器误差平方信号. 图4 为算法的学习曲线, 它们是 40 次独立实验结果的

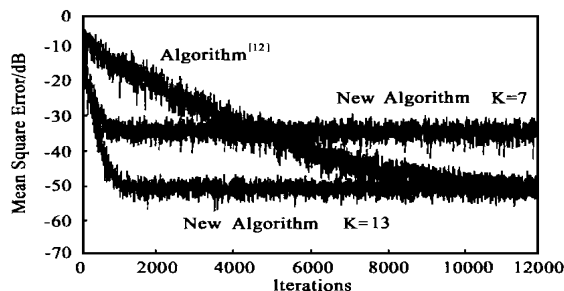


图4 RNN 训练算法的学习曲线, 其中 K 表示 RNN 隐元线性近似的数目

平均. 从实验结果图1、2和3可看出, 利用 RNN 的语音信号预测比线性预测更为有效, 具有较小的预测误差. 本文实验结果与文献[12]的实验结果比较也说明本文提出的 RNN 新算法

是可行的.从算法的学习曲线可看出,新算法的线性近似处理是完全可行的,当线性近似数目较大时($K=13$),线性近似完全可近似原始 RNN.另外,新算法具有更快的收敛速度,比传统 RNN 训练算法^[12]的收敛速度提高约 10~15 倍.

6 结论

本文利用隐表示、RNN 激励函数的线性近似、RNN 的统计概率模型和 EM 算法,提出一种基于隐元输出期望信号为隐含观测量的参数估计新方法.由于引入隐含观测量,使得复杂的 RNN 参数估计分解多个单隐元 RNN 的参数估计,使参数估计问题大大简化.利用 RNN 隐元激励函数的线性近似和 ME 模型,RNN 参数估计转化为包含隐变量的线性系统的最大似然参数估计问题,从而可以利用具有优良收敛性能的 EM 算法获得 RNN 隐变量和参数估计.

参考文献:

- [1] J. T. Connor et al. Recurrent neural networks and robust time series prediction [J]. IEEE Trans. NN, 1994, 5(2).
- [2] Rumelhart, D. E. et al. Learning internal representations by error propagation [A]. In: D. E. Rumelhart, J. L. McClelland and PDP Research Group (Eds.) Parallel Distribution Processing: Explorations in Microstructure of Cognition, Vol. 1: Foundations [C]. Cambridge, Mass: MIT Press.
- [3] S. Ma et al. An efficient EM-based training algorithm for feedforward neural network [J]. Neural Networks, 1997, 10(2).
- [4] Shun-ichi Amari. Information geometry of the EM and em algorithms for neural networks [J]. Neural Networks, 1995, 8(9).
- [5] M. I. Jordan and Lei Xu. Convergence results for the EM approach to mixtures of experts architectures [J]. Neural Networks, 1996, 8(9).
- [6] J. Xuan. Information geometry of maximum partial likelihood estimation for channel equalization [A]. ICASSP '96 [c]: 3533 - 3536.
- [7] H. Li. Probabilistic mapping networks for speaker recognition [A]. ICASSP '96 [C]: 3375 - 3377.
- [8] Roy L. Streit. Maximum likelihood training of probabilistic neural networks [J]. IEEE Trans. on Neural Networks, 1994, 5(5).
- [9] Stephen W. Piche. Steepest descent algorithm for neural network controller and filters [J]. IEEE Trans. Neural Networks, 1984, 5(2).
- [10] Hai-wei Chen. Modeling and identification of parallel nonlinear systems: Structural classification and parameter estimation methods [A]. Proc. of the IEEE [C], 1995, 83(1).
- [11] C. F. N. Cowan. Adaptive Filter [Z]. Prentice-Hall, Inc. Englewood Cliffs, NJ 07632, 1985.
- [12] Simon haykin et al. Nonlinear adaptive prediction of nonstationary signals [J]. IEEE Trans. SP, 1995, 43(2).

作者简介:



戴宪华 1992 年博士毕业于东南大学, 1993~1994 年华南理工大学博士后, 1995 年调入汕头大学工作, 现为汕头大学电子工程系教授. 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.