

一种用局部非线性预测模式实现的语音编码算法¹

马 竟 韦 岗*

(华南理工大学 广州金鹏集团有限公司 广州 510640)

*(华南理工大学电子与通信工程系 广州 510640)

摘 要 该文提出了一种基于局部非线性模式的语音信号预测模型 (LNLP), 并将其应用于语音编码中。在同样的算法条件下, LNLP 提高了预测增益, 改善了合成语音质量, 而其算法复杂度与常规算法相比并无显著增加。

关键词 局部预测, 非线性预测模型, 语音编码

中图分类号 TN912.3

1 引 言

局部预测模式是一种非常有效的语音信号模型。它实际上是在状态空间中局部拟合全极点自回归 (AR) 模型来实现非线性预测功能, 也即将模型参数当作了状态的函数。从理论上分析, 这种模型可认为是从有噪采样值中进行插值而获取, 那么可以通过有效的插值手段来逼近较为精确的模型^[1]。

现在我们来考虑如何选择这个局部模型。因为插值函数和插值方法有很多种, 如最简单的线性插值、稍复杂一些的样条插值等等。从已有的文献来看^[1,2], 普遍采用了线性函数来拟合这一局部模型, 从而构成局部线性预测 (LLP) 模型。实践中都取得了比线性预测编码 (LPC) 模型要好的预测效果。

对于非线性语音信号而言, 非线性插值函数比线性函数能获得更好的效果。进一步分析, 某些非线性插值函数, 例如径向基函数 (RBF), 能提供一套正则化解, 从而使建模参数减少, 并且保持系统的稳定^[3]。这一想法是促使我们应用 RBF 于局部模型的主要动机。作为一种比较和尝试, 我们将这种非线性预测模型应用于低速率语音编码方案中。

2 语音信号局部预测模型的选择

语音信号的一般预测模型可以这样描述, 对于由输入信号采样值 $x_t (t = 1, 2, \dots)$ 构成的状态矢量 $\mathbf{y}_t = (x_{t-(p-1)}, \dots, x_{t-1})^T$, 其中 p 为矢量的维数, 我们可以构造一函数 $F(\mathbf{x})$, 使得对于 x_t 的最小均方误差 (MMSE) 估计 $\hat{x}_t$ 为

$$\hat{x}_t = F(\mathbf{y}_{t-1}) + e_t \quad (1)$$

(1) 式中 e_t 为稳定的白噪声。从插值观点来看, 对 x_t 的预测就是选择合适的基函数来拟合 $F(\mathbf{x})$ 。下面我们就具体选择 $F(\mathbf{x})$ 基函数的问题作一下研究。

2.1 线性基函数

如果状态更新函数 $F(\mathbf{x})$ 足够平滑, 以至于我们能用其多维 Taylor 级数展开序列在 $\mathbf{x}_i$ 附近的最初几项来比较精确地描述它, 即

$$F(\mathbf{x}) = F(\mathbf{x}_i) + \nabla F^T(\mathbf{x}_i)(\mathbf{x} - \mathbf{x}_i) + \dots \approx \mathbf{b} + \mathbf{a}^T \mathbf{x} \quad (2)$$

¹ 1999-03-29 收到, 1999-10-06 定稿

国家自然科学基金 (69602002, 69772027) 和广东省自然科学基金 (960227, 963037) 资助项目

那么我们能将 $F(x)$ 近似为 x_t 附近的一个常数或线性函数。如果所选择的码本对 (x_{t-1}, y_t) 足够接近 x_t ，此种局部模式将是一种非常精确的拟合。生成这样一种线性拟合 ($F(x) \approx \hat{F}(x) = b + a^T x$)，相当于使参数 b 和 a 匹配于所选择的位于 x_t 附近状态空间区域中的参数对 (x_{t-1}, y_t) 。该模型通常会提供一套关于模型参数的超定线性方程

$$y = b + a^T x_{t-1} \quad (3)$$

这些线性方程可用最小二乘 (LMS) 算法来求解。

该局部模型可以认为是 AR 模型的一种推广，只不过自回归参数 a^T 演变为状态的函数。注意 (3) 式中另外的常数 b ，它是考虑到整个拟合过程中任意局部平均的一个重要参数。

2.2 非线性基函数及局部非线性模式的语音信号预测 (LNLP) 模型建立

上面用线性函数来逼近局部模型的基本假设条件是状态函数要足够平滑。事实上，平滑性是许多物理过程首要考虑的基本因素，也是允许建立逼近关系的最合理、最通用和最少假设限制条件的因素。

拟合问题中用于探索平滑性约束条件的技术手段称之为正则化。特别地，正则化理论与确定函数 F 使得下面式子最小化的问题十分接近

$$h[f] = \sum_{t=1}^N (x_t - f(x_{t-1}))^2 + \lambda \|Pf\|^2 \quad (4)$$

这里 $\|P\|^2$ 是一个包含先验知识的泛函数 (通常为一差分算子)， N 为被处理信号采样点的数目。最小化泛函数 h 会导致下面的解^[4]:

$$f(x) = \sum_{t=1}^N c_t G(x; x_t) \quad (5)$$

其中基函数 G 仅由稳定算子 P 的结构来确定，参数 c_t 由一线性系统来确定:

$$(G + \lambda I)c = y \quad (6)$$

其中 I 是单位矩阵，且 $(y)_{t-1} = x_t$, $(c)_t = c_t$, $(G)_{ij} = G(x_i; x_j)$ 。特别地，如果算子 P 是平稳和旋转不变式， G 将成为一径向函数，即 $G(x_i, y_j) = G(\|x_i - x_j\|)$ 。这种情况下，(4) 式的正则化解将由下式:

$$f(x) = \sum_{t=1}^N c_t G(\|x - x_t\|) \quad (7)$$

得出。对于此解，有两个因素需强调。首先，某些情况下，如同线性插值方案一样，需有如 b 一样的参数附加到 (5) 和 (7) 式中以便涵盖那些位于算子 P 中 0 空间的函数。但对某些径向函数，如高斯函数，并不需要这样，因为 (7) 式是完备的；第二点，实用中有一点重要的约束条件必须考虑到，即基于正则化解的非线性自回归系统 (如 (1) 式的语音预测模型) 必须是稳定的，因为它是一个“逆系统” (即合成系统)。使用高斯径向函数显然遵守这一约束条件^[4]。正则化解能保证非线性合成系统的稳定性。

(7) 式能用一个简单的只有一个隐层的网络计算得到，该网络就是如图 1 所示的被人们称之为径向基函数 (RBF) 的网络。

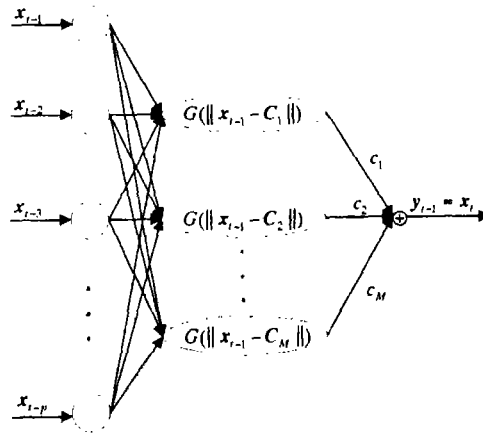


图 1 用于局部预测模式的 RBF 网络

对于线性基函数来讲, 它不能像径向基函数那样保证合成系统的稳定性, 原因就是信号的奇异性问题。因此必须应用奇异值分解或其它一些方法来改善这一特性^[5], 从而需要额外的大计算量。前已述及, 径向基函数能保证系统的稳定性和正则化, 而且文献 [6] 还证明了用高斯径向函数作为隐层单元激活函数的分层神经网络具有普适逼近的特性。这些因素决定了我们选用高斯径向函数:

$$G(x) = e^{-x^2/\sigma^2} \quad (8)$$

其中 σ 称为宽度, 作为 (7) 式的基函数。相应于线性模式的推广, 这里我们也可将 (7) 式作为非线性自回归 (NLAR) 模型的一种通用形式。由此我们得到基于 RBF 的 NLAR 局部预测模型为

$$\hat{x}_t = c_0 + \sum_{t=1}^M c_t G(||x_t - C_t||) \quad (9)$$

式中 C_t 称为中心, 其数目 M 远少于处理数据点的数目 N , 通常我们选取范数 $||\cdot||$ 为一欧几里得范数。至此我们完全建立了基于 RBF 的局部非线性预测 (LNLP) 模型。

(9) 式的估计精度很大程度上取决于中心 C_t 的位置及数目。用聚类的方法获取 C_t 比较费时, 且运算量大; 随机选择 C_t 则不能体现良好的网络性能。一种取得计算复杂性和预测精度之间折衷的优选方案就是采用正交最小二乘 (OLS) 算法^[5] 来选择数量较少的数据点作为中心。这种算法简单、有效, 很适合拟合 RBF 网络。OLS 算法还具有这样一个特性, 即遇到数值病态的情况最少。

3 基于 LNLP 的语音编码

LNLP 的应用有两个主要障碍: (1) 预测器参数的量化; (2) 计算复杂性。第一个问题的解决是通过将其应用于预测器后向自适应的语音编码方案, 从而避免参数量化问题。ITU-T G.728 LD-CELP^[7] 就是这样一种典型算法。我们可以将 LD-CELP 中线性预测器用 LNLP 来代替。至于第二点, 局部模式及相应的小数目的中心使得计算复杂程度已大为降低。应用于 LD-CELP 方案中的另一个好处是其小的子帧处理长度 (每 5 个采样点构成一子帧) 基本上保证了 LNLP 方案的计算量不会超过标准方案太多。

5 性能比较

5.1 预测性能

如果 LNLP 是一种更有效的预测器, 那么它应该能提供更大的预测增益或者说更“白”的预测残差。图 3 是比较结果的一个例子。在这个实验中, 我们使用了三种预测器, 即后向 LP, LLP 及 LNLP, 它们参数的数目相同 (都为 50)。输入信号为一帧 (30ms) 用 8kHz 采样、16bit 量化的语音^[9] (本文中所使用的语音数据都是以此方法获得)。LP 采用加汉明窗、列文逊-杜宾递推算法求解参数, LLP 用加权代价函数求解参数^[1], LNLP 用 OLS 算法来分析。比较结果是 LNLP 达到了这一要求。

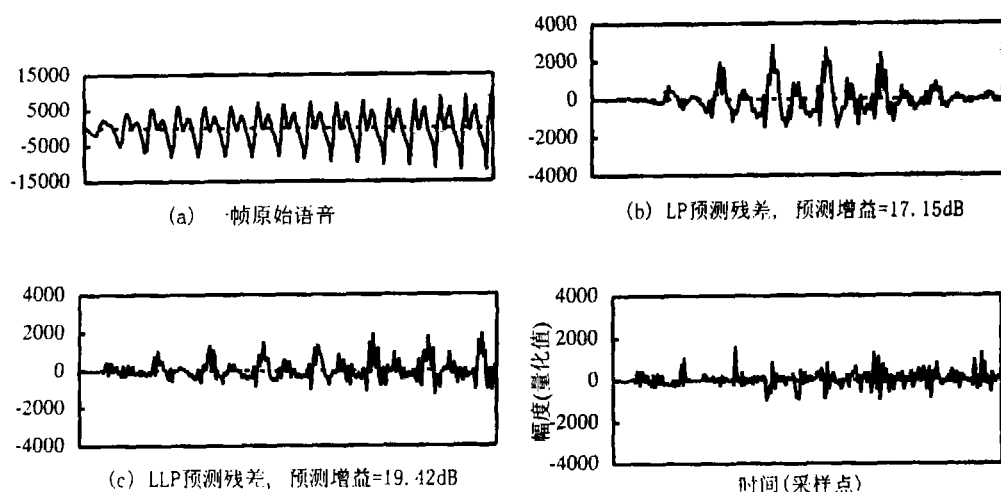


图 3 用一帧语音 (30ms) 比较三种预测器预测性能
(语音信号 8kHz 采样, 16bit 量化)

关于长时相关性预测性能, 由于局部模型方案不需要长时预测器, 我们可采用文献 [10] 的方法来比较一下清、浊音判断的准确程度。即先对一段由 10 男 10 女所说的 48s 的语音进行预测, 得到残差。因为语音基音周期通常在 20~140 采样点之间 (对 8kHz 采样而言), 我们将残差序列以 160 点为单位分段进行清、浊音判断, 把归一化自相关函数第一峰值点在不同门限值之上或之下的语音段数目与相对应的门限作成如图 4 所示的曲线 (语音段数目用百分比表示)。该曲线能清楚说明正确判断或错误判断的概率。显然, 曲线越低, 误判的概率就越低, 说明残差的周期相关性越小, 这种预测模型就越好。图 4 的结果是 LNLP 最好。其中 LNLP 在低门限时显示出优于 LLP。当然如提高门限值, 它们的性能会趋于一致。

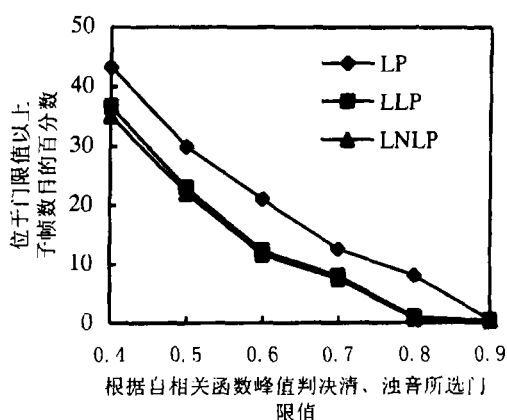


图 4 比较基音周期相关性

5.2 编码性能

这里我们与标准 LD-CELP 比较, 但对它作了一些简化, 去掉一些仿真有困难的过程, 如感觉加权、后滤波等。这时的 LD-CELP 编码的质量比标准方案稍差。

用三种方案对已用于预测性能比较的那一帧语音 (见图 3) 进行编码实验。其中 LLP 编码方法主要根据文献 [10]。结果很明显 (见图 5), LNLP 分析法性能最好。用做基音检测实验的 48 秒语音段来比较合成语音性能, 能得出同样的结论。LNLP 方案所得编码信噪比为 11.23dB, 比 LLP 方案提高 0.4dB, 比 LP 提高 0.7dB。值得一提的是, 在 LNLP 方案中出现数值计算病态情况是 3 次, 要少于 LLP 方案的 8 次, 这也使得 LNLP 方案优于 LLP 方案。

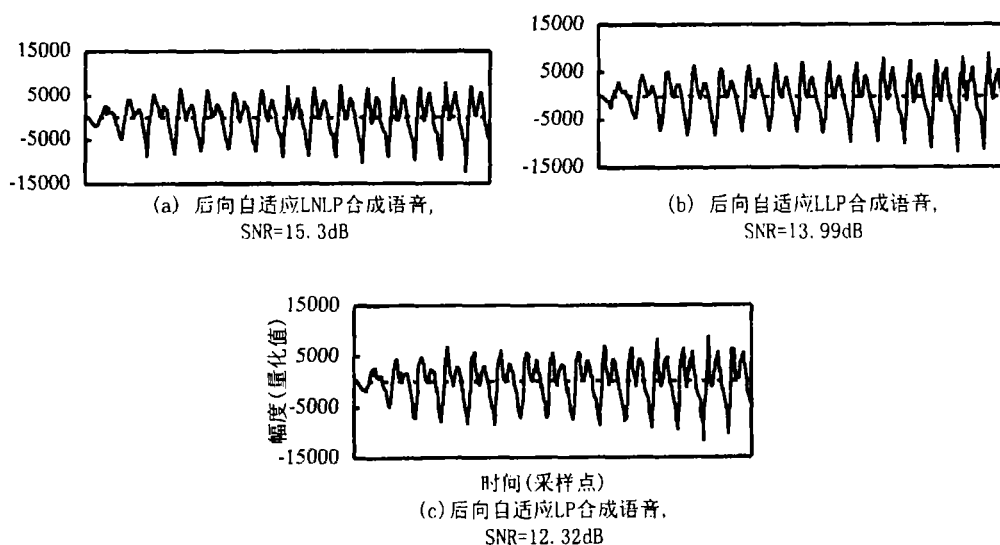


图 5 用同一帧语音 (30ms, 8KHz 采样, 16bits 量化) 比较三种编码方案合成语音性能

6 结 束 语

本文提出了一种局部非线性语音预测 (LNLP) 模型。预测分析和编码实践结果都表明, 它是优于局部线性模式 (LLP) 的一种非线性模型, 特别具有普适特性和较强稳定性, 是一种很有潜力的信号处理模式。

参 考 文 献

- [1] Singer A C, Wornell G W, Oppenheim A V. Codebook Prediction: A nonlinear signal modeling paradigm. ICASSP'92, San Francisco: 1992, V: 325-328.
- [2] 杨行峻, 迟惠生等. 语音信号数字处理. 北京: 电子工业出版社, 1995, 第四章.
- [3] Ogawa S, Ikeguchi T, Matozaki T, Aihara K. Nonlinear modeling by radial basis function networks. IEICE Trans. on Fundamentals, 1996, E79-A (10): 1608-1617.
- [4] Haykin S. Neural Networks: A Comprehensive Foundation. NJ: Macmillan, Englewood Cliffs, 1994, Chapter 2.
- [5] Chen S, Cowan F N, Grant P M. Orthogonal least squares learning algorithm for radial basis function networks. IEEE Trans. on Neural Networks, 1991, 2(2): 302-309.
- [6] Hartman E J, Keeler J D, Kowalski J M. Layered neural networks with Gaussian hidden units as universal approximations. Neural Computation, 1990, 2(1): 210-215.
- [7] Chen J H, Cox R V, Lin Y C, *et al.* A low-delay CELP coder for the CCITT 16kb/s speech coding standard. IEEE J. on Selected Areas in Communication, 1992, 10(1): 830-840.
- [8] Yakowitz S. Nearest-neighbour methods for time series analysis. Journal of Time Series Analysis, 1987, 8(2): 10-26.
- [9] Co. Ltd. International Electronics and Communication Basic Technology Institute. Tokyo: ATR speech database for research, 1997.
- [10] Cucchi, Fratti M, Ronchi M. On improving performance of analysis by synthesis speech coders. IEEE Trans. on Speech and Audio Processing. 1996, 4(3): 243-247.

A SPEECH CODING ALGORITHM WITH LOCAL NONLINEAR PREDICTION MODEL

Ma Ni Wei Gang*

(South China Univ. of Technology, Guangzhou Jinpeng Group Co. Ltd., Guangzhou 510640)

*(Dept. of Electron. and Comm. Eng., South China Univ. of Technology, Guangzhou 510640)

Abstract A new signal processing algorithm based on a local nonlinear prediction model (LNLP) is presented. With the same implementation, the speech predicting and coding based on LNLP gives improved performance and the computational efforts for LNLP analysis is not increased over that for the conventional algorithms.

Key words Local prediction, Nonlinear prediction model, Speech coding

马 霓: 男, 1971 年生, 华南理工大学和广州金鹏集团有限公司联合培养博士后, 现从事移动通信和信号处理方面的研究工作。

韦 岗: 男, 1963 年生, 教授, 博士生导师, 系主任, 主要研究兴趣包括信号处理、神经网络、回波理论、图像处理、模式识别以及移动通信等。