

【技術分類】 6 - 1 - 2 音声 / 特徴抽出技術 / モデリング

【 F I 】 G10L3/00

【技術名称】 6 - 1 - 2 - 1 線形予測分析法 (Linear Predictive Coding : LPC)

【技術内容】

本来は標本値の線形結合により将来を予測する方法であるが、音声のモデルとすることもできる。線形予測フィルタまたは逆フィルタの係数 $\{\alpha_i\}$ を調整して、出力 $\hat{x}_t$ の分散 $\sigma^2$ を最小にすることに相当する。音声を周期 $T[s]$ で標本化した信号を $\{x_t\}(t: \text{整数})$ とし、現在の標本値 $x_t$ と、これに隣接する過去の $p$ 個の標本値との間に、次のような線形一次結合が成り立つと仮定する。

$$x_t + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} = \hat{x}_t \quad (1)$$

$\hat{x}_t$ を(2)のようにおけば、(1)式は(3)式のように書ける。

$$\hat{x}_t = -\sum_{i=1}^p \alpha_i x_{t-i} \quad (2)$$

$$\hat{x}_t = x_t + e_t \quad (3)$$

つまり(1)式は、線形予測係数を $\{\alpha_i\}$ とする線形予測モデルであり、 $e_t$ は線形予測残差である。線形予測フィルタ $F(z)$ および逆フィルタ $A(z)$ を、次のように定義する。

$$F(z) = -\sum_{i=1}^p \alpha_i z^{-i} \quad , \quad A(z) = 1 + \sum_{i=1}^p \alpha_i z^{-i} = 1 - F(z)$$

$\hat{X}$ 、 $X(z)$ 、 $E(z)$ をそれぞれ $\hat{x}$ 、 $x_t$ 、 $e_t$ の $z$ 変換とすれば、

$$\hat{X}(z) = F(z)X(z)$$

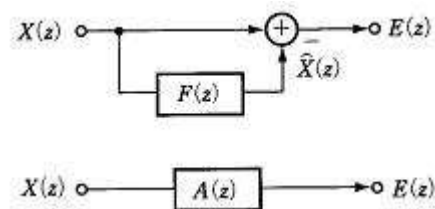
$$X(z)(1 - F(z)) = E(z)$$

$$X(z)A(z) = E(z)$$

と書ける。 $z$ 変換領域での線形予測モデルのブロック図を下に示す。

【図】

線形予測モデルのブロック図



出典：「音声情報処理」,「電子情報通信工学シリーズ」, 1998年6月30日発行、古井貞熙著、森北出版株式会社発行、27頁 図2-5  $z$ 変換領域での線形予測モデルのブロック図

【出典 / 参考資料】

「音声情報処理」,「電子情報通信工学シリーズ」, 1998年6月30日発行、古井貞熙著、森北出版株式会社発行、27頁

【技術分類】 6 - 1 - 2 音声 / 特徴抽出技術 / モデリング

【 F I 】 G10L3/00

【技術名称】 6 - 1 - 2 - 2 隠れマルコフモデル (Hidden Markov Model : HMM)

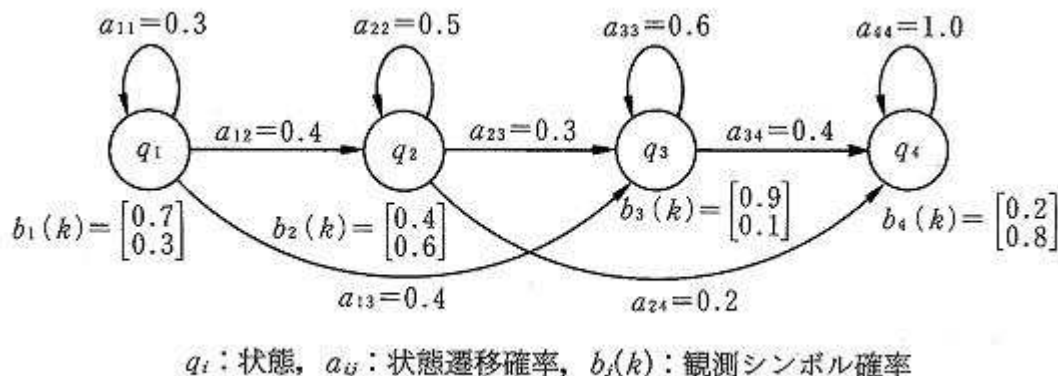
【技術内容】

音声を定常信号源(状態)の連結で表し、時系列をある状態から次の状態への遷移確率で表現する、統計的信号源モデルである。スペクトル時系列の統計的変動をモデルのパラメータに反映させることができる特徴があるが、モデルパラメータを決定するための学習処理がやや複雑になる。音声パターンは時間的に非可逆であるため、一つの初期状態から一つ最終状態へ一方的に遷移していく、left-to-right 型モデルで表されることが多い。図は、最もよく用いられるベイキス型 HMM の例である。一つの音素、音節、または単語を一つの HMM で表す。 $a_{ij}$  は状態  $q_i$  から状態  $q_j$  への状態遷移確率であり、 $i > j$  なら  $a_{ij} = 0$  である。 $b_{ij}(y)$  は、状態が  $q_i$  から  $q_j$  へ遷移するとき、スペクトルパターン  $y$  が観測(出力)される確率を表す。 $\{b_{ij}(y)\}$  を出現確率行列と呼ぶ。

出現確率を連続分布モデルで表すときには、単一ガウス分布(正規分布)または混合ガウス分布(Gaussian Mixture Model : GMM)が用いられるが、パラメータの自由度を減らすために、無相関ガウス分布を用いることも多い。離散分布モデルを用いる場合は、スペクトルパターンをクラスタ化(ベクトル量子化)して代表スペクトルパターン(符号ベクトル)を生成し、各符号ベクトルの出現確率の組合せによって出現確率を表す。

【図】

離散型 HMM の例



出典:「音声と画像のデジタル信号処理」,「デジタル信号処理ライブラリー 3」, 1996年9月30日発行、谷萩隆嗣編著、株式会社コロナ社発行、56頁 図4・3 ベイキスモデルの例

【出典 / 参考資料】

・「音声情報処理」,「電子情報通信工学シリーズ」, 1998年6月30日発行、古井貞熙著、森北出版株式会社発行、97頁

・「音声と画像のデジタル信号処理」,「デジタル信号処理ライブラリー 3」, 1996年9月30日発行、谷萩隆嗣編著、株式会社コロナ社発行、56頁

【技術分類】 6 - 1 - 2 音声 / 特徴抽出技術 / モデリング

【 F I 】 G10L3/00

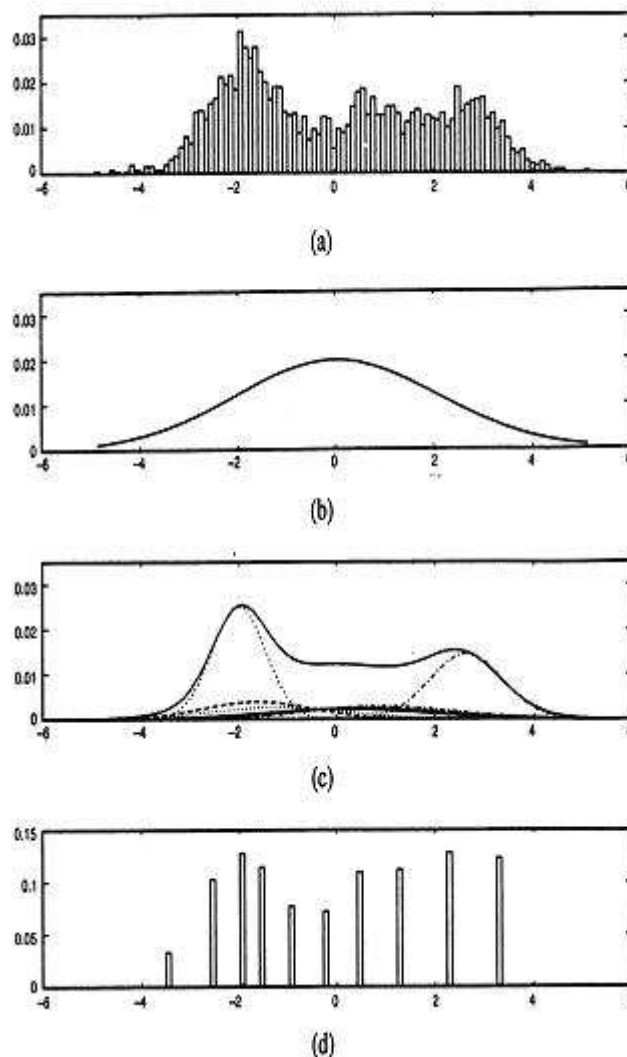
【技術名称】 6 - 1 - 2 - 3 ベクトル量子化 (Vector Quantization : VQ)

【技術内容】

複数のベクトルをまとめて、1つの符号で表わす方式である。音声以外にも、画像などの情報量圧縮方式として広く用いられている。音声の情報処理には、ケプストラム係数をベクトル化したものがよく用いられるが、そのままでは図(a)のようにベクトルの要素数が膨大なものとなる。このベクトルの集合をクラスタ化し、図(d)のような代表パターン(符号ベクトル、セントロイド、テンプレートなどと呼ばれる)の集合を生成する。このようにして、話者の特徴を少数のセントロイド(コードベクトル)で代表させることが可能となり、かつ(b)の単一ガウス分布よりも、元のスペクトル包絡に近い形で近似できる。(c)のGMMは、次の項で説明する。

【図】

音声スペクトルのモデル例 (a) ケプストラム係数のヒストグラム (b) 単一ガウス分布 (c) 混合ガウス分布 (GMM) (d) VQ セントロイド



© 2005 IEEE

出典:「Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models」  
「IEEE Transactions on Speech and Audio Processing Vol.3 No.1 pp.72-83」, 1995年1月発行、

Dauglas A. Reynolds Richard C. Rose 著、IEEE 発行、75 頁 Fig.3 Comparison of distribution modeling

【出典 / 参考資料】

・「音声情報処理」,「電子情報通信工学シリーズ」, 1998 年 6 月 30 日発行、古井貞熙著、森北出版株式会社発行、56 頁

・「Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models」,  
「IEEE Transactions on Speech and Audio Processing Vol.3 No.1 pp.72-83」, 1995 年 1 月発行、  
Dauglas A. Reynolds Richard C. Rose 著、IEEE 発行

【技術分類】 6 - 1 - 2 音声 / 特徴抽出技術 / モデリング

【 F I 】 G10L3/00

【技術名称】 6 - 1 - 2 - 4 混合ガウス分布モデル (Gaussian Mixture Model : GMM)

【技術内容】

音声スペクトルを、複数のガウス分布を混合してモデル化する方法である。D 次元のランダムなベクトル  $\mathbf{x}$  の分布  $b_i(\mathbf{x})$  ( $i=1, \dots, M$ ; component densities: 成分密度) を、次のような正規分布で表わす。

$$b_i(\mathbf{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)' \Sigma_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right\}$$

$\boldsymbol{\mu}_i$  は平均ベクトル、 $\Sigma_i$  は共分散行列である。話者の尤度は、これに重み  $p_i$  をかけた和で表わされる。(図)

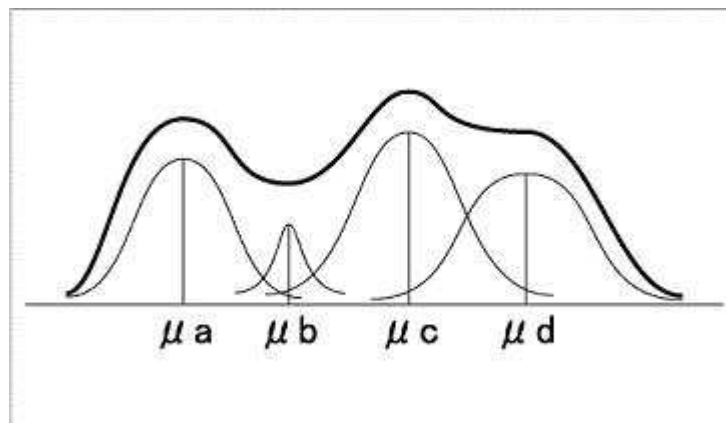
$$p(\mathbf{x} | \lambda) = \sum_{i=1}^M p_i b_i(\mathbf{x})$$

前項の図(c)のように、次の4つのパラメータを用いて、単一ガウス分布やベクトル量子化よりもより忠実に元のスペクトルをモデル化できる。

$$\lambda = \{p_i, \boldsymbol{\mu}_i, \Sigma_i\} \quad i=1, \dots, M$$

【図】

混合ガウス分布モデル(GMM)のイメージ図



本標準技術集のために作成

【参考資料】

「Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models」,  
「IEEE Transactions on Speech and Audio Processing Vol.3 No.1 pp.72-83」, 1995年1月発行、  
Douglas A. Reynolds Richard C. Rose 著、IEEE 発行