論　文

Ergodic HMM を用いた未知・複数信号源クラスタリング問題の検討

村上　仁一[†]　　杉山　雅英[††]　　渡辺　秀行[†]

Study of Unknown-Multiple Signal Source Clustering Problem using Ergodic HMM

Jin'ichi MURAKAMI[†], S. SUGIYAMA[††], and H. WATANABE[†]

あらまし
入力された信号系列を複数個の信号源に分割し、信号源を推定する問題は、音声情報処理や言語処理などの極めて広い分野において存在する。本論文では、この問題を未知・複数信号源クラスタリング問題と名付け、Ergodic HMM をもちいた解決方法を提案する。 Ergodic HMM において、状態を信号源に、状態から出力されるシンボル系列を信号系列に対応させた場合、Viterbi 復号法によりセグメンテーション位置とカテゴリ識別を同時に推定することができる。次に、この問題の応用として複数話者発話の識別を調べた。この実験の結果、 Ergodic HMM の初期パラメータの値が重要であること、および話者性を有する特徴ベクトルとして長時間窓幅の LPC ケプストラムが有効であることが示された。

キーワード
全状態結合型隠れマルコフモデル, 話者識別, クラスタリング

1. ま え が き

異なる N 個の信号源より生成された信号系列が、どの信号源から生成されたのかを分割・識別する問題を、未知・複数信号源クラスタリング問題とする。この問題は、音声処理分野に限らず言語処理などの分野でも重要なテーマである。たとえば音響単位の自動決定問題 [1] や、話者の言語を識別する問題、発話様式 (単語・文発声・連続発声等) の識別問題、音声・非音声の識別問題、さらに言語における品詞ラベルの自動的な作成および形態素解析 [2] などがこの問題に相当する。

この未知・複数信号源クラスタリング問題は大きくわけて以下の 4 つの部分問題から構成される。

(1) カテゴリを特徴付ける特徴量の問題

(2) カテゴリ遷移の セグメンテーション位置を決める問題

(3) セグメンテーションされた各ブロックを N 個のカテゴリに識別する問題

(4) カテゴリ数 N を推定する問題

本論文では、カテゴリ数 N が既知の場合に、観測された信号系列に対し、自動的にセグメンテーションを行ない、セグメンテーションされた各区間のカテゴリを識別する問題を扱う。

ところで HMM は、非定常信号源の一つのモデルとして、特に音声認識の分野で広く用いられている [4]。このモデルはオートマトン制御の下で確率的定常信号源を次々に切替えることにより、非定常信号源を表現している。音声認識では、音声の特性を考慮して、left-to-right 型の HMM が用いられる。しかし、話者認識や言語のモデリングにおいては、全ての状態が全ての状態に接続している Ergodic HMM が使用されている [5]。この Ergodic HMM を未知・複数信号源クラスタリング問題に利用した時、カテゴリーが状態に相当し、信号系列は状態から出力されるシンボル系列と考えることができる。

実験では未知・複数信号源クラスタリング問題の応用として複数話者発話の識別問題を検討する。実験から、男性話者 4 名の場合、長時間窓分析を用いた LPC ケプストラムを用いることにより、音声資料により異

[†] ATR 音声翻訳通信研究所, 京都府
ATR Interpreting Telephony Reseach Labs. , Kyoto-fu, 619-02 Japan
[††] 会津大学 コンピュータ理工学部、 会津若松市
School of Computer Science and Engineering nThe University of AIZU, Tsuruga, Ikki-machi, Aizu-Wakamatsu-shi Fukushima, 965 Japan

なるが、フレーム単位で約 67.5%の平均識別率が得られることを示す。次に異なる初期モデルから尤度の高いモデルを選択することにより、約 78%の平均識別率が得られることを示す。最後に残された課題と、その解決の展望について報告する。

なお過去に行なわれた類似した研究としては、文献 [6] においてセグメンテーション位置およびカテゴリ数が既知の場合に、Universal コードブックおよびその出現頻度による Kullback 情報量を使用して話者を識別する方法を報告した。文献 [7], [8] では本論文と同様な複数話者発話の識別問題を扱っている。しかし 1 話者の音響パラメータを 1 つのガウス分布であると仮定し、全音声データに対し VQ clustering を続けることによって問題の解決を図っている。また、通常の話者識別問題 [5] は事前に多量の音声を用いて話者モデルを作成し、それとは異なる認識用の入力音声に対して逐次識別を行なう問題であるのに対し、本研究は連続した音声中に含まれる複数の信号源を、その音声のみを用いて事前学習なしで分割・分類する問題を扱っている。

2. 未知・複数信号源クラスタリング問題

2.1 問題の定式化

図 1 に示すように、n 次元ベクトルで与えられる信号の系列を $\boldsymbol{X} = (\boldsymbol{x}_t)(t = 1, 2, ..., T)$ とする。この系列は、K 個のブロック $\boldsymbol{X}_k$ $(k = 1, 2, ..., K)$ からなり、各ブロックが $N(\leq K)$ 個のカテゴリ $\boldsymbol{C}_j$ $(j = 1, 2, ..., N)$ のいずれかから生じた系列であるとする。ここで、ブロック $\boldsymbol{X}_k$ の構成要素を $(\boldsymbol{x}_{t_{k-1}+1} ... \boldsymbol{x}_{t_k})$ (ただし、$t_0 = 0, t_K = T$) 、継続長を M_k $(= t_k - t_{k-1})$ とする。

未知・複数信号源クラスタリング問題は、与えられたベクトル信号系列に対し、信号源の特性の違いに着目してブロックの切れ目の位置 $t_k (k = 1, 2, ..., K-1)$ を探し、この K 個のブロックを $N(\leq K)$ 個のカテゴリに識別しカテゴリ数 N を推定することである。

3. Ergodic HMM を用いた解法

3.1 Ergodic HMM

カテゴリ数 N が既知で、セグメンテーション未知のカテゴリ識別を考える時、Ergodic HMM の適用が考えられる。この場合、カテゴリが状態に対応し、信号系列は状態から出力されるシンボル系列と考えることができる。そして問題点は次の 2 つになる。

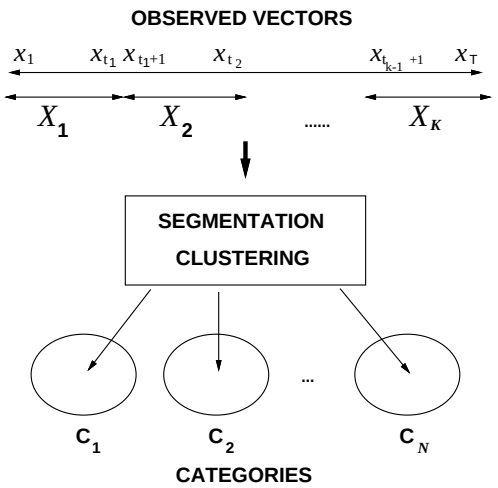

図 1 未知・複数信号源クラスタリング問題の定義
Fig. 1 Definition of N-signal Problem

(1) 信号系列 $\boldsymbol{X}$ の尤度を最大にする HMM のパラメータ $\boldsymbol{M}$ を推定する問題 (モデルの推定、または学習)。なお、パラメータ $\boldsymbol{M}$ は初期状態確率 $\boldsymbol{\pi} = (\pi_i)$、状態遷移確率 $\boldsymbol{A} = (a_{ij})$ 及びシンボル出力確率 $\boldsymbol{B} = (b_j(l))$ で構成されるとする。

(2) HMM のパラメータ $\boldsymbol{M}$ が信号系列 $\boldsymbol{X}$ を出力する可能性の高い状態遷移系列を推定する問題 (最適状態遷移系列の推定)。

Ergodic HMM による解法の概略を、図 2 に示す。

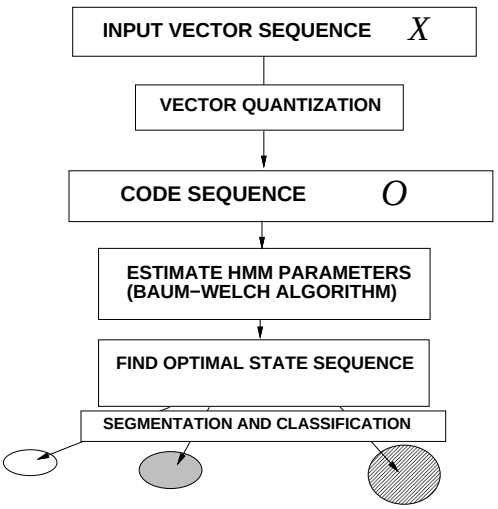

図 2 Ergodic HMM を利用した未知・複数信号源クラスタリングの手順
Fig. 2 Flow chart that includes an Ergodic HMM for the N-signal source problem

3.2 HMM のパラメータ推定

HMM のパラメータ $\boldsymbol{M}$ の推定手法として、Baum-Welch のアルゴリズムが知られている。これは EM アルゴリズム [4] を $\boldsymbol{M}$ の推定に適用したもので、学習データの尤度が最大になるように HMM のパラメータを繰り返し演算で計算する。ただし、この計算は局所的最大点の方向に進むため [4]、初期パラメータの設定が重要になる。

3.3 最適状態遷移系列の推定について

ここでは、Baum-Welch アルゴリズムで求まったパラメータ推定値から、信号系列を生み出す最適状態遷移系列を推定する問題を考える。この推定手法として、Viterbi 復号法と Forward 復号法が考えられる。この復号法を次に示す。この復号法から計算された最適状態遷移系列 $\boldsymbol{S}^* = \{s_1^*, ..., s_T^*\}$ から、セグメンテーション位置とカテゴリ識別が直接得られる。

3.3.1 Viterbi 復号法

推定された HMM のパラメータ $\boldsymbol{M}$ がコード系列 $\boldsymbol{O}$ を出力する可能性の高い最適状態遷移系列は、Viterbi 復号法により効率的に求まる [4]。Viterbi 復号法を次に示す。

(1) 全ての $i \in \{1, ..., N\}$ に対し、

$$\delta_1(i) = \log \pi_i + \log b_i(\boldsymbol{o}_1), \phi_1(i) = 0$$

とおく。

(2) 時間軸 $t = 2, ..., T$ に沿って、全ての $j \in \{1, ..., N\}$ に対し

$$\delta_t(j) = \max_i [\delta_{t-1}(i) + \log a_{ij} + \log b_j(\boldsymbol{o}_t)],$$

$$\phi_t(j) = arg \max_i [\delta_{t-1}(i) + \log a_{ij}]$$

(3) 最適状態遷移系列に対する対数尤度及び T 時刻目の最適状態を次式で求める。

$$\max_{\boldsymbol{S}} P(\boldsymbol{O}, \boldsymbol{S} \mid \boldsymbol{M}) = \max_j \delta_T(j),$$

$$s_T^* = arg \max_j \delta_T(j)$$

(4) 時間軸 $t = T-1, ..., 1$ に沿って、次式により最適状態遷移系列を得る。

$$s_t^* = \phi_{t+1}(s_{t+1}^*)$$

3.3.2 Forward 復号法

最適状態遷移系列の推定には Viterbi 復号法の他に Forward 復号法が考えられる。このアルゴリズムは、始めに HMM のパラメータ $\boldsymbol{M}$ がコード系列 $\boldsymbol{O}$ を出力する時の尤度を、各状態からの総和で計算する。次に最適状態遷移系列は各時刻における最大の尤度を持つ状態とする。この Forward 復号法を次に示す。

(1) 全ての $i \in \{1, ..., N\}$ に対し、

$$\delta_1(i) = \pi_i \times b_i(\boldsymbol{o}_1)$$

$$s_1^* = arg \max_i \delta_1(i)$$

(2) 時間軸 $t = 2, ..., T$ に沿って、全ての $j \in \{1, ..., N\}$ に対し

$$\delta_t(j) = \sum_i [\delta_{t-1}(i) \times a_{ij} \times b_j(\boldsymbol{o}_t)]$$

$$s_t^* = arg \max_j \delta_t(j)$$

(3) 全ての可能な状態遷移系列に対する尤度及び T 時刻目の最適状態を次式で求める。

$$\sum_{\boldsymbol{S}} P(\boldsymbol{O}, \boldsymbol{S} \mid \boldsymbol{M}) = \sum_j \delta_T(j)$$

$$s_T^* = arg \max_j \delta_T(j)$$

4. 複数話者発話の識別実験

本論文では未知・複数信号源クラスタリング問題の応用例として、複数話者発話の識別問題を取り上げる。この場合、信号系列は LPC ケプストラムの音響パラメータに、各カテゴリは各話者に、セグメンテーション位置は話者の遷移に対応する。

4.1 音 声 資 料

実験には、単語音声 (ATR 5240 単語データ) の波形データを用いて疑似的な音声資料を作成する。話者は 4 名とし、同一話者の異なる単語音声を 8 個つないで 1 ブロック (1 発話) とする。なお、各単語の前後の無音区間は削除する。 1 セットの音声資料は、ブロックごとに話者間でランダムに連接して、 1 話者につき 8 ブロックで作成する。従って 1 セットの音声資料は 32 ブロックで構成され、話者は 31 回遷移する。音声資料の例を図 3 に示す。

この音声資料を 8 セット作成し実験に用いる。音声資料の 1 セットの平均発話時間は約 150 秒である。

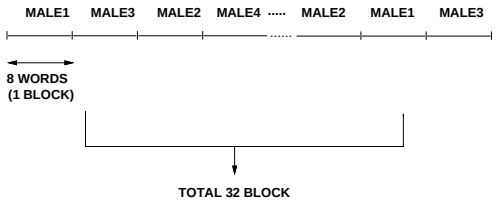

図 3　実験に使用した音声資料の例
Fig. 3　Fig.3 Example of speech data (Speaker: 4 males)

4. 2　音響パラメータ

本実験では、話者特徴量を表す音響パラメータとして LPC ケプストラム (1 次 ~ 16 次) を用いる。音声分析条件を表 1 に示す。 Universal コードブックはテストデータ (音声資料) から Euclid 距離を用いて作成する。

表 1　音声分析条件

音声特徴量	LPC ケプストラム係数
標本化周波数	12 kHz
LPC 分析	14 次
打ち切り次数	16 次
分析窓長	21.3 ms (256 点)
フレーム更新周期	10.7 ms (128 点)
高域強調	$(1 - 0.97z^{-1})$
コードブックサイズ	256

4. 3　HMM の初期パラメータ

HMM には多くの種類があるが、本論文では各状態がシンボルを出力する状態出力タイプ (Moore 型) の離散型 (discrete) HMM を考える。 また状態数 (カテゴリ数) は、話者数と同数の 4 状態とする。

なお、Baum-Welch の学習アルゴリズムは局所的最大点の方向に進むため、初期パラメータによって識別率が大きく変動することが知られている [9]。そこで初期パラメータを以下の 3 種類の方法で計算し、識別実験を行なう。

(1) 実験 1 : (全パラメータに真値を与えた場合)

HMM パラメータ $\boldsymbol{M}$ の初期パラメータとして、初期状態確率 $\boldsymbol{\pi}^{(0)}$ と状態遷移確率 $\boldsymbol{A}^{(0)}$、とシンボル出力確率 $\boldsymbol{B}^{(0)}$ 全てに真値を与える。なお話者および発話時間を決めて音声資料を作成しているため、真値はこれから直接計算できる。

(2) 実験 2 : (シンボル出力確率に真値を与えた場合)

初期パラメータとして、シンボル出力確率 $\boldsymbol{B}^{(0)}$ は真値を与えるが、初期状態確率 $\boldsymbol{\pi}^{(0)}$ と状態遷移確 率 $\boldsymbol{A}^{(0)}$ は等確率にする。

(3) 実験 3 : (シンボル出力確率にランダムな値を与えた場合)

初期パラメータとして、初期状態確率 $\boldsymbol{\pi}^{(0)}$ と状態遷移確率 $\boldsymbol{A}^{(0)}$ を等確率に、シンボル出力確率 $\boldsymbol{B}^{(0)}$ をランダムに与える。なお、 $\boldsymbol{B}^{(0)}$ を等確率に与えた場合、Baum-Welch の学習アルゴリズムは動かない。

4. 4　識別率の評価方法

Forward 復号法や Viterbi 復号法から得られる最適状態遷移系列はカテゴリの番号であって、カテゴリと話者の関係は未知である。そのため以下の式で識別率を算出する。

$$R = \frac{1}{T} \max_{\sigma} \sum_{t=1}^{T} d(\tau(\boldsymbol{x}_t), \sigma(S_t)) \qquad (1)$$

ここで、 τ は最適状態遷移系列、σ は $(1, 2, \ldots, N)$ の任意の置換、S_t は各発話の正解カテゴリ番号、d は値が一致した時のみ 1 それ以外は 0 である関数である。本論文では、 各カテゴリ番号を σ で置換し、$N!$ 通りの置換について正解率を算出しその中の最大値を識別率とする。従って話者が 4 名の場合 $24(= 4!)$ 通りの組み合わせを調べる。

なお、複数話者発話の音声データでは、LPC 分析のフレーム更新周期の間に話者が遷移する。このフレームでは話者を一意に決めることができない。 そこで話者が遷移したフレームは、 どちらの話者が選択されても正解にする。

5. Ergodic HMM による複数話者発話の識別の実験結果

LPC ケプストラムを分析窓長 21.3ms で計算した時の実験結果を図 4 に示す。平均識別率は、実験 1 、 2 に関しては 8 セットの音声資料の平均値で、実験 3 に関しては、 8 セットの音声資料それぞれに対し 16 回の異なる初期モデルで実験した、合計 128 回の平均値である。この図において、縦軸が平均識別率で横軸が HMM の学習回数である。□ は実験 1 の、△ は実験 2 の、○ は実験 3 の結果である。また、実線で■ ▲ ● は Viterbi 復号法による平均識別率で、破線で□ △ ○ は Forward 復号法による平均識別率である。この図からわかることを以下に示す。

(1) Viterbi 復号法と Forward 復号法の平均識別率を比較すると、差は小さい。しかし、パラメータの全てを真値にした実験 (実験 1) では、Forward 復号法では 94.0% であったのに対し、Viterbi 復号法では

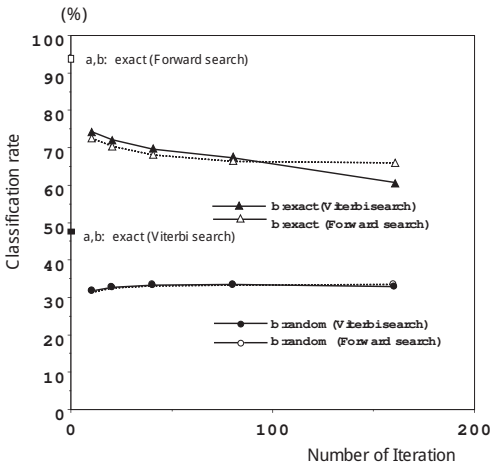

図 4　学習回数と平均識別率の関係
Fig. 4　Relationship between number of iteraions and classification rate

48.3%しか得られな かった。この原因として Viterbi 復号法は誤った経路を選択した場合、最後まで経路を間違えてしまうのに対し、 Forward 復号法は間違った経路を選択しても、その後正しい経路を選択する可能性があるためと考えている。

（ 2 ） シンボル出力確率 $\boldsymbol{B}^{(0)}$ のみ真値にした実験（実験 2 ）では平均識別率約 75%が得られた。しかし、学習回数を繰り返すに従い、平均識別率は低下した。また $\boldsymbol{B}^{(0)}$ をランダムにした実験（実験 3 ）では平均識別率で 30%から 35%と低い値になった。この値は学習回数を増加してもあまり向上しない。この原因として、今回の実験は音素認識で使用される分析条件で実験を行なったため、パラメータは主に音素のカテゴリを特徴づけるパラメータになっていて、話者を特徴づけるパラメータになっていないためと考えている。

5.1　話者特徴量と長時間窓分析

話者識別の研究から話者識別には長時間平均スペクトルが有効であることが知られている [10]。従って分析窓長を長くすることによって話者の識別性能が向上することが予想される。そこで分析窓長を変化させたときの識別率の変化を調べる。Universal コードブックサイズは 256 と 64 で行なう。また、いずれの実験でもフレーム更新周期は分析窓長の半分とする。平均識別率は、各音声資料に対し、乱数により 16 個のランダムな初期モデルを作成して、その各々について試行を行なった音声資料 8 セットの平均値、すなわち計 128 回の試行に対する平均値である。HMM の学習は 160 回繰り返して終了する。その他の実験条件は実験 3 と同一である。この実験結果を図 5 に示す。この図において縦軸は平均識別率、横軸は分析窓長である。

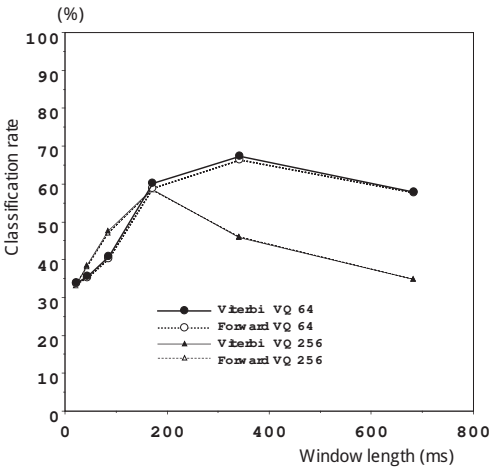

図 5　分析窓長と平均識別率の関係
Fig. 5　Relationship between window length and classification rate

この図からわかることを以下に示す。

（ 1 ） 分析窓を長くするに従い、平均識別率は向上するが、ある値を越えると低下する。

（ 2 ） 分析窓長が 128ms 以下ではコードブックサイズ 256 の方が 64 よりも識別性能が高い。しかし分析窓長が 128ms 以上ではコードブックサイズ 64 の方が 256 よりも識別性能が高い。

（ 3 ） コードブック 64、分析窓長 341ms において最も高い識別率が得られる。

これらの結果は、次のような原因によると考えている。分析窓長を広げるとケプストラムのパラメータに含まれる話者特徴量は増加する。 従って、分析窓長を広げると識別性能が向上する。しかし、分析窓長を広げると得られるデータ量が減少する。例えば分析窓長 683ms ではデータは平均約 440 個である。そのため Ergodic HMM のパラメータの推定精度が低下する。よって分析窓幅がある閾値を越えると、識別性能は低下する。

なお、データ量の減少を防ぐために、フレーム更新周期を短くすることが考えられる。しかし、この場合フレーム更新周期の間に話者が遷移するデータが増加する。このデータは 2 人の話者の特徴量が入るため、不安定な特徴量を含む。したがってフレーム更新周期を短くしてデータ量を増加させても、識別率は向上し

ないことが予想される。実際に、フレーム更新周期を変化させて行なった実験で、この予想が確かめられた。

5.2 コードブックサイズ

ここでは Universal コードブックサイズの数を変化させたときの認識性能の変化を示す。コードブックサイズの数は推定するパラメータとパラメータ推定に使用されるデータ量との関係で決定される。複数話者発話の識別は不特定話者であるためコードブックサイズが 256 では小さい可能性がある。一方、サイズを大きくした場合、出現するコードの個数が少なくなるためシンボル出力確率の信頼性が低下すると思われる。実験は分析窓長 341.3ms と 170.7ms で行なった。その他の実験条件は前の実験と同一である。この実験結果を図 6 に示す。これからコードブックサイズの数が 64、分析窓長 341ms のとき認識性能が最大になることがわかる。

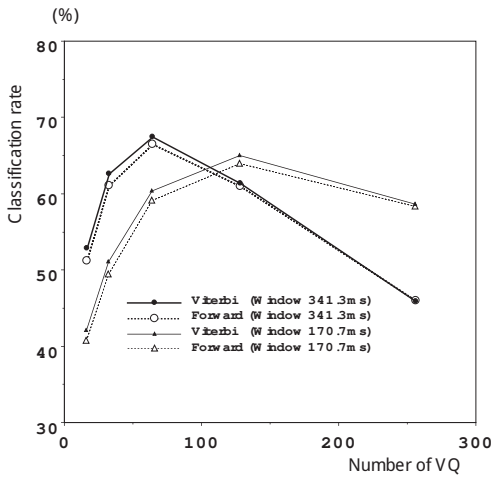

図 6 コードブックサイズと平均識別率の関係
Fig. 6 Relationship between number of codebook sizes and classification rate

5.3 対数尤度に対する識別率の変化

Baum-Welch の学習アルゴリズムは局所的最大点の方向に進むため、初期パラメータによって学習後のパラメータは大きく変化する。そして学習データに対する尤度も変化する。そこで尤度に対する識別率の関係を調べた。実験は分析窓長 341.3ms、コードブックサイズを 64 とし、 HMM の学習は 160 回繰り返した。他の実験条件は実験 3 と同様である。 1 セットの音声資料に対し、異なる 16 個の初期モデルを実験した 8 セットの音声資料の、合計 128 回の実験結果を図 7 に示す。縦軸は識別率を示し、横軸は対数尤度を示している。この図から HMM の尤度と識別率には相関があることがわかる。

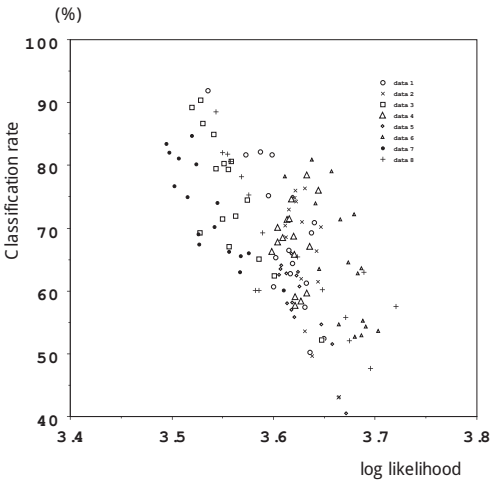

図 7 尤度と識別率の関係
Fig. 7 Relationship between likelihood and classification rate

5.4 初期モデルの選択

前の実験から、尤度の高いモデルは識別率が高いことがわかった。そこで異なる初期モデルをランダムに 16 個作成し、 Baum-Welch 学習をした後、尤度の最も高いモデルを選択したときの、音声資料 8 セットに対する平均識別率を図 8 に示す。この図では、横軸がコードブックのサイズで縦軸は平均識別率である。 また分析窓長は 341.3ms である。その他の実験条件は実験 3 と同様である。この結果から、尤度の高い HMM を選択することにより平均識別率が約 10%程度向上し、最適な条件で 78.8%が得られることがわかった。

6. 考　　察

本論文では、未知・複数信号源クラスタリング問題の中で複数話者発話の問題を取り上げた。しかし、多くの問題が未解決である。これらについて述べる。

（1） Baum-Welch アルゴリズムにおける初期パラメータの設定

HMM の尤度関数は様々な局所的最大点が存在する。従って初期パラメータの設定は重要である。今回の実験では尤度の高い HMM のモデルを選択することにより 78.8%の平均識別率を得た。しかし、分析窓長 341.3ms でコードブックサイズの数が 64 のときシンボル出力確率に真値をいれて Baum-Welch 学習をおこない、平均識別率を計算したところ 96.2%が得られ

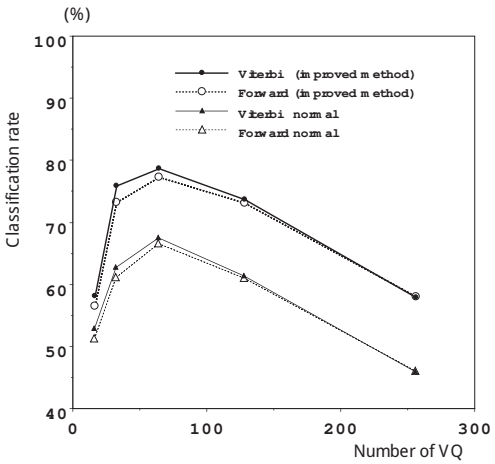

図 8　初期モデルを選択したときの平均識別率の変化
Fig. 8　Comparison of selection method and base-line method

た。したがって、別の初期パラメータの計算方法を考案することにより、より高い識別性能が得られる可能性がある。

（２） 識別率の評価方法

話者数が多い場合の識別率の評価方法を考える必要がある。今回の実験では全ての可能性を調べて、最も高い値を識別率としたが（式 (1)）、この方法では、話者数が多くなるに従い、組み合せの数は階乗で多くなる。そこで評価方法の高速化が必要である。これに対しては、分岐限定法などの組み合せ最適化の適用が考えられる。

（３） 話者識別の時間分解能

複数話者発話の音声データでは、LPC 分析のフレーム更新周期の間に話者が遷移する。このフレームは話者を一意に決めることができない。従って話者識別の時間分解能はフレーム更新周期に依存する。この時間分解能を向上させる方法を考える必要がある。

（４） Ergodic HMM の状態数

一般に音声データには無音区間がある。この区間は、物理的に話者を特定できない。この解決方法として、Ergodic HMM の状態の数を話者の数より１つ多くして、無音区間のみを生成する状態を入れる方法が考えられる。

（５） 連続分布型 HMM

HMM には多くの種類がある。言語のような離散型のデータを扱う場合は離散型 HMM が好ましいが、話者識別などで連続量を扱うには連続型 HMM が好ましいと考えられる。この場合コードブックサイズの問題や VQ 歪などの問題点がなくなるため、識別率は向上すると考えられる。

（６） カテゴリ数 N の推定

今回の実験では、話者数（カテゴリ数）を４として実験を行なった。そして、この話者数は事前にわかっていると仮定した。この話者数を推定する方法を考える必要がある。

7. むすび

本論文では、信号系列を複数個の信号源に分割する問題を取り上げ、Ergodic HMM を用いた解法を示した。応用例として複数話者発話の識別をあげ、実験により識別性能を示した。この結果、以下の事柄が示された。

（１） セグメンテーション位置とカテゴリ識別を同時に推定する、Ergodic HMM による解法では、パラメータ推定の際の収束計算の初期パラメータの設定が重要である。

（２） HMM のパラメータの初期状態確率、状態遷移確率、シンボル出力確率のうち、シンボル出力確率が最も重要である。

（３） 複数話者発話の識別においては 341ms 程度の長時間窓分析した LPC ケプストラムを用いることにより、より良好な識別性能が得られる。

（４） 尤度の高いモデルを選択することにより平均識別率は向上する。

謝　辞

研究の機会を与えて頂いた北海道大学電子科学研究所永井信夫教授、三木信弘助教授、ATR 自動翻訳電話研究所樽松明社長（現在電気通信大学教授）、熱心に御討論していただくと同時に有益な御助言をいただいた音声情報処理研究室の皆様に感謝致します。

文　献

[1] 児島宏明, 田中和世, 速水悟, “単語音声サンプルからの階層的な音韻概念の獲得”, 音響学会講演論文集, 2-P-4 (1991-10).

[2] 村上仁一, 嵯峨山茂樹, “HMM を用いた形態素解析”, 情報処理学会講演論文集, 4F-01 (1992-10).

[3] 松井和子, 古井貞熙：“エルゴード的 HMM による話者認識”, 音響学会講演論文集,3-6-14 (1991-10).

[4] X.D. Huang, Y. Ariki and M.A. Jack, “Hidden Marcov Models for Speech Recognition”, Edinburgh University Press, Edinburgh (1990).

[5] 松井和子, 古井貞熙, “エルゴード的 HMM による話者認

識", 音響学会講演論文集, 3-6-14 (1991-10).
[6] 杉山雅英, 村上仁一, 渡辺秀行, "N 信号源モデルに基づく音声の区分化識別問題ー話者特徴の違いに基づく区分化音声の識別", 電子通信学会論文誌 D-2 Vol.J76-D-2 No12, PP.2477-2485 (1993-12)
[7] G.Yu, M.Siu, H.Gish, "An Unsupervised, Sequential, Learning Algorithm for the segmentaion of Speech Waveform with Multiple Speakers", Proc. of ICASSP92, 2-189 (Apr.1992).
[8] H.Gish, G.Yu, "Clustering of Speakers Engaged in Dialog", Proc. of Speech Research Symposium XII (Jun.1992).
[9] K.Lari, S.J.Young, "The estimation of stochastic context-free grammars using the Inside-Outside algorithm" , speech recoginition, Computer Speech and Language, vol. 4, pp.35-56 (1990).
[10] 古井貞熙, 板倉文忠, "単語の統計的パラメータによる話者認識", 信学会論文誌, 56A-11, pp.717-724 (1973-12).

Abstract

In this paper, we consider signals originated from a sequence of sources. More specifically, the problems of segmenting such signals and relating the segments to their sources are addressed. This issue has wide applications in many fields. This report describes a resolution method using an Ergodic Hidden Markov Model (HMM), in which each HMM state corresponds to a signal source. The signal source sequence can be determined by using a decoding procedure (Viterbi algorithm or Forward algorithm) over the observed sequence. Baum-Welch training is used to estimate a HMM parameters from training material. As an application of a multiple signal source identification problem, an experiment is performed on unknown speaker identification. Result show a classification rate of 79% for 4 male speakers. The results also indicate that the model is sensitive to the initial values of the Ergodic HMM and that employing the long-distance LPC cepstrum is effective for signal preprocessing.