

遺伝的アルゴリズムを用いる HMM 音声認識における適合度の検討

宜野座朝尚

高良富夫 (琉球大学 工学部)

1 まえがき

隠れマルコフモデル(HMM)は、現在の音声認識技術の中で、その高い認識性能と大語彙、不特定話者音声認識技術への応用の可能性の点から最も有望な手法として盛んに研究されている[1]。しかし、この方法では最適な構造の選別法が確立しておらず、最適な構造の決定は、多くは経験に基づいて行われていた。

そこで我々はHMMに対し、探索・最適化の一手法である遺伝的アルゴリズム(GA)を用いて最適な構造を決定する音声認識法を提案し、その有効性を示した[2]。さらにこの方法は、語彙単語をセットにして進化させる方法に拡張された[3]。

しかし、これまでGAによるHMMの構造決定法では、最適な適合度の与え方について必ずしも十分検討されておらず、生成された構造が認識誤り率を最小にする最適な構造ではない場合があった。したがって、GAによるHMMの構造決定では、より最適な構造を探索するための適合度の検討が必要である。

そこで、本研究では、GAの選択淘汰において適合度の評価を、誤り率、独立尤度、尤度セット、および相対尤度を用いて行い、適合度の与え方の有効性について比較検討を行う。

2 隠れマルコフモデルによる音声認識

隠れマルコフモデルと従来のマルコフモデルの違いは、隠れマルコフモデルでは出力シンボルに

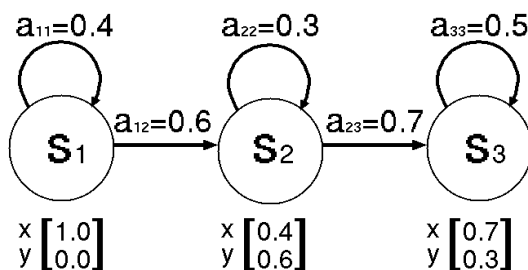


図 1 : HMM構造の例

$$\begin{matrix} & 1 & 2 & 3 \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

図 2 : 図1の行列表現

よって一意に状態の遷移先が決まらないことがある。HMMとして最も良く使用される構造(ここでは基本left-to-right構造と呼ぶことにする)を図1に示す。図中の s_1, s_2, s_3 はそれぞれの状態を表す。図中の矢印の上にあるスカラー値は状態 s_i から s_j 状態への状態遷移確率 a_{ij} を表している。状態についているベクトル値は、シンボル x, y を出力する出力確率を表す。また図1のHMM構造の行列表現は図2のようになる。HMMの行列表現は、遺伝的アルゴリズムの符号化において使用される。

ここで、シンボルの出力の系列を $B = b_1, b_2, \dots, b_T$ 状態遷移系列を $A = a_1, a_2, \dots, a_T$ とする。添字は、出力シンボル、状態とその各時刻を表す。

B の生成確率 $P(B)$ は

$$P(B) = \sum_F P(B | A) P(A) \quad (2.1)$$

である。ここで F は可能な状態系列を表す。また、状態 $i-1$ から状態 i へ遷移する確率を $P(a_i | a_{i-1})$ 、状態 $i-1$ から状態 i へ遷移する際に b_i を出力する確率を $P(b_i | a_i, a_{i-1})$ と表す。よって状態遷移確率は

$$P(A) = \prod_i P(a_i | a_{i-1}) \quad (2.2)$$

となる。また、シンボル系列の条件付き確率は

$$P(B | A) = \prod_i P(b_i | a_i, a_{i-1}) \quad (2.3)$$

となる。これらの式から、式(2.1)は以下のようになる。

$$P(B) = \sum_F \prod_i P(a_i | a_{i-1}) P(b_i | a_{i-1}, a_i) \quad (2.4)$$

隠れマルコフモデルによる音声認識では、入力パターンのパラメータの時系列 B に対して式(2.4)の値を求め、これを最大にするカテゴリを認識結果とする。本研究では式(2.4)を効率よく計算するため、ビタビアルゴリズムを用いて対数尤度を計算し、式(2.4)の近似として用いる。

3 遺伝的アルゴリズムを用いるHMM構造の選択

3.1 HMMのコーディング法

本研究では簡単のため、隠れマルコフモデルの状態数を、予備実験により最も良いとされた8に固定した。また、音声の連続性から自己遷移があり、時間の前後関係から時間的に逆向きの遷移が起らないという制約をつけて図3の上部の行列で示されるHMMの構造を用いる。ここで行は遷移元の状態番号、列は遷移先の状態番号を表し、行列要素の値は遷移が有るとき1、無いときは0である。これを図3の下に示すようにコーディングし、遺伝的アルゴリズムで用いる。上記の音声の制約から、GAでは、行列の上三角要素だけ用いる。

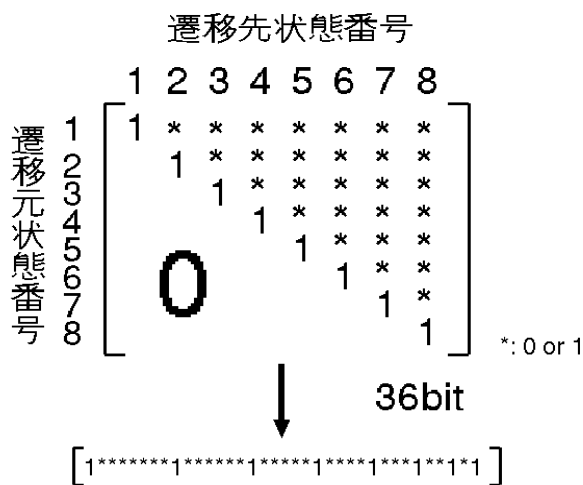


図3：HMM構造のコーディング

3.2 遺伝的アルゴリズムの適用

遺伝的アルゴリズムは、生物の進化の過程に着想を得て構築された確率的探索・学習・最適化のアルゴリズムである。遺伝的アルゴリズムの処理手順は

- () 初期集団の生成
- () 適合度の評価
- () 選択淘汰
- () 交叉
- () 突然変異
- () () ~ () の繰り返しからなる。

以下にこれらの説明と本方法における適用法を示す。

() 初期集団の生成

本研究では、遺伝的アルゴリズムにおける初期集団を30とし、そのうち1個は基本left-to-rightとし、他の29個体はランダムに生成したものを使用する。

() 適合度の評価

最初に各個体について、隠れマルコフモデルのパラメータの初期値を図4のように与える。これらを初期値として訓練用音声データを用いて、forward-backwardアルゴリズムで訓練を行う。全てのデータについて訓練が終了すると、その訓練したパラメータを用いて適合度を求める。

() 選択淘汰

適合度の大きいものほど、次世代に同じ遺伝子を持つ子孫をより多く残す処理である。

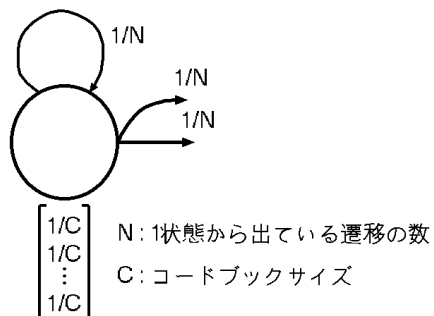


図4：HMMの初期パラメータの与え方

() 染色体の交叉

交叉は、二つの個体の遺伝子を、ランダムに決定した位置で部分的に入れ換える操作で、交叉が生じる確率を交叉率という。本研究では1点交叉を行う。

() 染色体の突然変異

突然変異は、一定の確率で各個体の遺伝子に相当するビットを0ならば1とし、1ならば0に変更する操作で、この生起確率を突然変異率という。

() 繰り返しと終了

これらの操作を終了すると新しい世代の個体集団が作られたことになる。この新しい集団に対して、さらに()~()を行い、新しい世代を作っていく。この世代交替を繰り返して、ある終了条件を満たすと終了し、その時の最良の染色体を解とする。

4 適合度の検討

この節では適合度の与え方を説明する。個体集団サイズを N 、個体 i の順位を I としたとき、適合度 F_i は次式で与える。

$$F_i = N - I$$

また個体 i の順位 I は、個体 i の評価値 f_i で与えられる。以下に誤り率、尤度、尤度セット、相対尤度の場合について、 f_i の与え方を示す。

() 誤り率

各個体に対し、訓練用のデータで認識実験を行い、各個体の誤り率を計算する。訓練用データの個数を M 、エラーの個数を e とすると、 f_i は次式で与えられる。

$$f_i = e/M \quad (4.1)$$

f_i の小さい順に順位をつける。

() 尤度

1つの個体に対し、1つの訓練用データについて、

ビタービアルゴリズムで対数尤度を計算する。次にその対数尤度を訓練データのフレーム数で割り、1フレームあたりでの対数尤度とする。これを全ての訓練用音声データを用いて行い、対数尤度の和を求める。 L_{ij} を訓練用データ j に対する個体 i の対数尤度とすると、 f_i は次式で与えられる。

$$f_i = \sum_{j=1}^M L_{ij} \quad (4.2)$$

ここで M は訓練用データの個数である。 f_i の大きい順に順位をつける。以下同じ。

() 尤度セット

尤度セットはカテゴリ k の個体 i に対して対数尤度の和を尤度の方法と同じように計算する。これを語彙セットに含まれる各カテゴリについて行い、その和を求める。カテゴリの個数を K 、カテゴリの対数尤度を Y_{ik} とすると、 f_i は次式で与えられる。

$$f_i = \sum_{k=1}^K Y_{ik} \quad (4.3)$$

ここで

$$Y_{ik} = \sum_{j=1}^M L_{ij}^{(k)} \quad (4.4)$$

であるので、

$$f_i = \sum_{k=1}^K \sum_{j=1}^M L_{ij}^{(k)} \quad (4.5)$$

() 相対尤度

相対尤度では、一つの個体に対し尤度を使用したときと同じようにすべて訓練用音声データで対数尤度を計算しその和を求める。次に他のすべてのカテゴリの訓練用音声データを用いて対数尤度の和を計算し、カテゴリの個数で平均する。その平均と訓練用音声データで計算した対数尤度の和との差を求める。これを相対尤度と呼ぶことにする。相対尤度を語彙セットに含まれる各カテゴリについて行い、その和を求める。他のカテゴリの訓練用音声データ m に対する個体 i の対数尤度を E_{im} 、カテゴリの個数を K 、訓練用データの個数を M_1 、他のすべてのカテゴリの訓練用音声データの個数を M_2 、相対尤度を R_{ik} とすると、 f_i は次式で与えられる。

$$f_i = \sum_{k=1}^K R_{ik} \quad (4.6)$$

ここで

$$R_{ik} = \sum_{j=1}^{M_1} L_{ij} - \frac{1}{K-1} \sum_{m=1}^{M_2} E_{im} \quad (4.7)$$

5 認識実験

適合度の検討のため比較実験を行った。訓練用に成人男性話者16人、女性話者20人の計36人が各2回発声した1桁数字音声(zero, one, ..., nine, oh)を用いた。認識対象として、別の成人男性話者20人、女性話者20人が各2回発声した1桁数字音声を用いた。また、遺伝的アルゴリズムで使用したパラメータは、個体集団サイズが各カテゴリについて30個体、世代交替は30回とし、交叉確率を60%、突然変異確率は3%とした。なお初期集団の中には隠れマルコフモデルにおいてよく使用される基本left-to-right構造を1個体挿入した。

結果を図5,6に示す。誤り率を e 、尤度を l 、尤度セットを ls 、相対尤度を r 、基本L-Rを b と表す。closed test では、どの適合度を用いても基本left-to-right構造の性能をこえる構造がみつかった。open test では誤り率を用いたとき、基本left-to-right構造の性能をこえることができないようである。また尤度セットを適合度に用いたとき、closed test では15世代以降、open test では19世代以降、認識率が収束していることがわかる。これは語彙をセットにしたことにより、死構造の割合が増え探索範囲が狭くなったためだと思われる。相対尤度を適合度に用いるとclosed test とopen test はともに、認識率は横ばい傾向ではあるが、探索範囲が狭くなっても認識率の値は収束していないようである。尤度を用いるとclosed test では認識率は横ばい傾向であるが、open test では世代を経るに従い認識率の上昇がみられる。また尤度を用いる場合は語彙セットではないので、死構造の増える割合が尤度セットを用いたときに比べて低いので、認識率の値は収束していないようである。

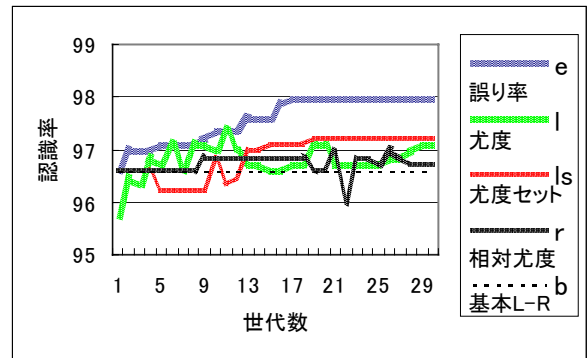


図 5：各世代におけるclosed testの結果

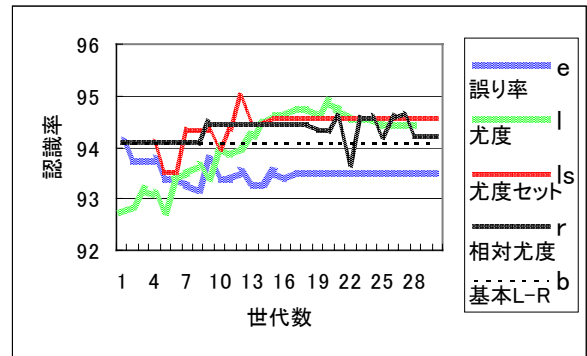


図 6：各世代におけるopen testの結果

6 むすび

本研究では、GAを用いた音声認識について、最適なHMM構造を決定するため、誤り率、尤度、尤度セット、相対尤度を用いて適合度を検討した。その結果、尤度セットを適合度に用いた場合、死構造に対する処理が必要であることがわかった。今後の課題としては、尤度セットと相対尤度のハイブリットな適合度の検討が挙げられる。

参考文献

- [1]中川聖一：“確率モデルによる音声認識”，pp.28-89, コロナ社(1998).
- [2]Takara, T., Higa, K., Nagayama, I.: “Isolated Word Recognition Using the HMM Structure Selected by the Genetic Algorithm”, ICASP97, SPCH7L. 1, Vol. 2, pp.967-970,(1997).
- [3]Takara, T., Ngaki, E.: “Spoken Word Recognition Using the Artificial Evolution of a set of vocabulary”, ICSLP2000, Vol. 1, pp.222-225,(2000).