

MDL 規準を用いた逐次状態分割法による音響モデル自動構造決定

實廣 貴敏[†] 松井 知子[†] 中村 哲[†]

[†] ATR 音声言語コミュニケーション研究所

〒619-0288 「けいはんな学研都市」光台二丁目2番地2

E-mail: †{takatoshi.jitsuhiro,tomoko.matsui,satoshi.nakamura}@atr.co.jp

あらまし 音響モデルとしての HMM 状態共有構造を作成するアルゴリズムは、一般に分割時に ML 規準を用いている。しかし、基本的にはパラメータが増加するほどゆ一度は増加するため、ML 規準のみでは適切なパラメータ数で停止するのは困難であり、一般に総状態数のような経験的な値が停止条件として用いられてきた。ML 規準を用いる最ゆう逐次状態分割法 (ML-SSS 法) でも同様の問題が存在している。そこで、本研究では、MDL 規準を ML-SSS 法へ分割および停止条件として導入した MDL-SSS 法を提案する。評価実験により、自動的にほぼ最適なモデル構造が得られることがわかり、また、従来法より適切なモデル構造が得られることがわかった。さらなる拡張として音素決定木クラスタリングを用いた SSS 法においても MDL 規準を導入した。音素決定木と Chou アルゴリズムの組合せで、自動的に同等の性能が得られることがわかった。また、2つの係数に対する最適値は手法に依存することがわかった。
キーワード 音声認識, 音響モデル, 構造学習, SSS アルゴリズム, MDL 規準

Automatic Construction of HMM Structures by the MDL Criterion on the SSS Algorithm

Takatoshi JITSUHIRO[†], Tomoko MATSUI[†], and Satoshi NAKAMURA[†]

[†] ATR Spoken Language Translation Research Laboratories

2-2-2 Hikaridai, "Keihanna Science City" Kyoto 619-0288 Japan

E-mail: †{takatoshi.jitsuhiro,tomoko.matsui,satoshi.nakamura}@atr.co.jp

Abstract We propose a new Successive State Splitting (SSS) algorithm based on the Minimum Description Length (MDL) criterion to design HMM topologies automatically. Many methods have been proposed to create tied-state structures of HMMs but they usually need empirical values as stop criteria, for example, the number of states. The SSS algorithm is a mechanism for creating both temporal and contextual variations based on the Maximum Likelihood (ML) criterion. However, it also needs to empirically predetermine the control parameters, including the total number of states as stop criteria. We introduce the MDL criterion to the ML-SSS algorithm so that it can automatically create proper topologies without any heuristic stop criteria. Experimental results using the ATR travel arrangement dialogs show that our extended algorithm can automatically stop splitting and obtain more appropriate HMM topologies than the original one. We also try to extend the MDL-SSS algorithm by using the phonetic decision tree clustering for contextual splitting. The method using a combination of the phonetic decision tree clustering and the data-driven clustering could automatically obtain almost equivalent performances. The optimal two coefficients of the criterion were dependent on each method.

Key words speech recognition, acoustic model, topology training, SSS algorithm, MDL criterion

1. はじめに

音声認識の音響モデル作成において、HMM の状態共有構造を自動生成する方法として、代表的なものに音素決定木クラスタリング [1]、逐次状態分割法 (Successive State Splitting

(SSS) algorithm) [2] がある。前者は音声学的知識による音素カテゴリを利用して音素環境のクラスタリングを行っているのに対し、後者は音素環境方向と時間方向との二方向の状態分割をデータに依存して行う。

音素決定木クラスタリングによる方法では、従来、ゆ一度最

大 (Maximum Likelihood (ML)) 規準で分割する状態を選択する。しかし、一般にモデルパラメータ数が増加すれば、ゆ一度は増加するため、ML 規準のみを停止条件にすることは困難であった。

そこで近年、音素決定木クラスタリングによる方法において、Minimum Description Length (MDL) 規準 [3]、Bayesian Information Criterion (BIC) [4] といった情報量規準を分割条件および停止条件として用いる手法が提案されている。これらの手法では情報量規準の値が改善されなくなるまで状態を分割する。

一方、SSS アルゴリズム [2] は初め、話者依存モデルの HMM 状態ネットワーク (隠れマルコフ網: HMnet) 作成方法として提案され、現在は ML-SSS 法 [5] として不特定話者モデル作成方法へと拡張されている。これらの方法は音素決定木クラスタリングと同様、ML 規準を分割条件に用いており、従来の音素決定木クラスタリングによる方法と同じ問題を含んでいた。

そこで、本研究では ML-SSS 法において、情報量規準として MDL 規準を分割および停止条件として導入する。ここでは提案手法を MDL-SSS 法と呼ぶ。文献 [3] では MDL 規準を音素環境方向の分割に用いるが、提案手法では、音素環境方向および時間方向の両分割に用いる。

2. では、まず従来法である ML-SSS 法について概略を述べる。3. で MDL 規準の定義を述べ、4. で提案手法の分割条件を定義する。5. で比較実験を行う。さらに、6. では、提案法の拡張として、音素決定木を利用した SSS 法において MDL 規準を適用し、同様に評価を行う。

2. ML-SSS アルゴリズム

2.1 ML-SSS アルゴリズムの概略と問題点

まず、従来法として ML-SSS 法 [5] について概説する。図 1 に流れ図を示す。ML-SSS 法では各状態を音素環境方向と時間方向に分割を試み、ML 規準で分割する状態を選択する。これを総状態数 N_s に達するまで繰り返すことで、状態共有構造を生成する。音素決定木クラスタリングでは環境方向の分割に音素カテゴリに関する質問で分割する。これに対して ML-SSS では、P. A. Chou のアルゴリズム [7] を用いて、十分統計量を利用した効率的な計算を行いながら、triphone データごとにクラスタリングを行い、ML 規準で分割を決定している。時間方向分割では、分割対象の状態の分布パラメータを複製して 2 状態を時間方向に連結し、分布パラメータを forward-backward アルゴリズムで再推定する。しかし、理論上、状態を分割するたびに分布パラメータを求めるには、その前後の状態を含めて再推定を行う必要がある。それでは計算量が膨大になるため、ML-SSS では、近似的に対象状態のみについてゆ一度計算をし、分割前の状態滞在確率で重みをかけた値を用いる。

ML-SSS では総状態数 N_s を停止条件として必要とする。また、時間方向分割に対して HMnet 内のパス長制限、つまり triphone での時間方向の状態数制限 N_p を必要とする。この制約がないと ML 規準の性質から、時間方向にはデータのある限り長くなる傾向がある。したがって、あるデータに対して最適

なモデルをひとつ決めるためには、一般にこの 2 つの値を実験的に求める必要がある。学習データ、評価データが変われば、その度ごとに調整しなければならない。

2.2 ML-SSS での分割評価規準

ここで、ML-SSS での分割規準である対数ゆ一度期待値の差分について示す。状態 S_i を状態 S_{i_1}, S_{i_2} に分割する場合、対数ゆ一度期待値の差分は以下のように表せる。

$$G(S_i) = G_{output}(S_i) + G_{trans}(S_i) \quad (1)$$

ここで、 $G_{output}^{(ML)}$ は出力確率による差分、 $G_{trans}^{(ML)}$ は遷移確率による差分を示し、それぞれ以下の式で表される。

$$\begin{aligned} G_{output}(S_i) = & -\frac{1}{2} \{ \Gamma(S_{i_1}) \log |\Sigma(S_{i_1})| \\ & + \Gamma(S_{i_2}) \log |\Sigma(S_{i_2})| \\ & - \Gamma(S_i) \log |\Sigma(S_i)| \} \end{aligned} \quad (2)$$

$$\begin{aligned} G_{trans}(S_i) = & \Xi(S_{i_1}, S_{i_1}) \log a_{i_1 i_1} \\ & + \{ \Gamma(S_{i_1}) - \Xi(S_{i_1}, S_{i_1}) \} \log(1 - a_{i_1 i_1}) \\ & + \Xi(S_{i_2}, S_{i_2}) \log a_{i_2 i_2} \\ & - \Xi(S_i, S_i) \log a_{ii} \end{aligned} \quad (3)$$

ここで $\Gamma(S_i) = \sum_t \gamma_t(S_i)$ は状態 S_i が学習データ中に出現する期待値を示す。 $\gamma_t(S_i)$ は時刻 t に S_i に存在する確率である。 $\Xi(S_i, S_j) = \sum_t \xi_t(S_i, S_j)$ は、 S_i から S_j へ遷移する期待値を示す。 $\xi_t(S_i, S_j)$ は時刻 t に S_i から S_j へ遷移する確率である。 a_{ii} は S_i の自己ループ確率である。

環境方向分割では、計算量削減のため、遷移確率は再推定せず、分割条件でも考慮しない。したがって、出力確率による差分のみを考慮する。時間方向分割では、状態間の遷移確率が増加するため、遷移確率も考慮される。よって、環境方向、時間方向の分割条件をそれぞれ $G_c^{(ML)}(S_i)$ 、 $G_t^{(ML)}(S_i)$ と書くとすると以下のようになる。

$$G_c^{(ML)}(S_i) = G_{output}(S_i) \quad (4)$$

$$G_t^{(ML)}(S_i) = G_{output}(S_i) + G_{trans}(S_i) \quad (5)$$

各状態において式 (4)、(5) を求め、その値が最大になる分割を選択する。

3. MDL 規準

モデルセット $\{\theta^{(i)}, i = 1, \dots, I\}$ に対する MDL 規準 [6] は以下のように定義される。

$$L_i(x) = -\log P(x|\hat{\theta}^{(i)}) + \frac{\alpha_i}{2} \log N_T + \log I, \quad (6)$$

ここで $x = \{x_1, \dots, x_{N_T}\}$ は観測データ、 α_i はモデルのパラメータ数、 $\hat{\theta}^{(i)}$ はモデル i の ML 推定値である。

第 1 項はモデル $\hat{\theta}^{(i)}$ の対数ゆ一度を反対符号にしたものである。より小さい MDL 値を持つモデルがよいモデルである。学習データが固定で、一般にパラメータ数 α_i が大きくなると第 1 項は減少する。しかし、第 2 項は増加するので、MDL 値はある点で極小値を持つ。この点を検出できれば過学習を避けることができる。

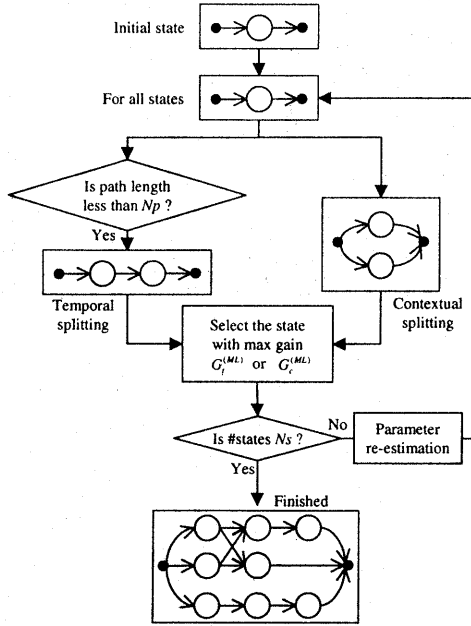


図1 ML-SSS法

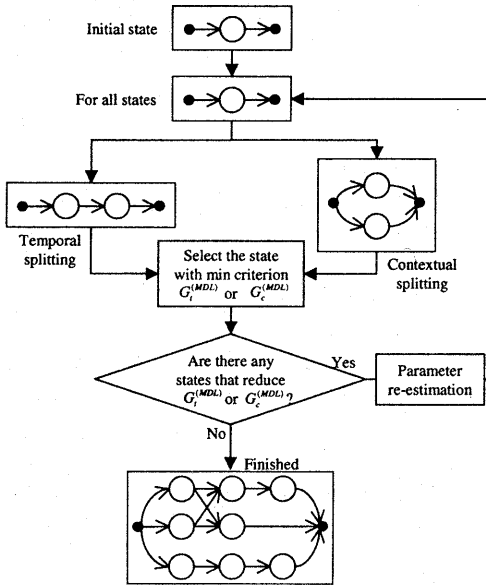


図2 MDL-SSS法

4. MDL-SSS アルゴリズム

4.1 MDL-SSS アルゴリズムでの分割規準

図2に提案法 MDL-SSS の流れ図を示す。ML-SSS での分割規準の代わりに MDL 規準 [6] を用い、MDL 規準値が改善されなくなったら分割停止するという事で自動的に分割を停止できる。以下に、MDL 規準の差分から得られる環境方向分割で

の規準 $G_c^{(MDL)}$ および時間方向分割での規準 $G_t^{(MDL)}$ を示す。

$$G_c^{(MDL)}(S_i) = -G_c^{(ML)}(S_i) + \frac{\alpha'_c - \alpha_c}{2} \log \Gamma(S_{all}) \quad (7)$$

$$G_t^{(MDL)}(S_i) = -G_t^{(ML)}(S_i) + \left\{ \frac{\alpha'_t}{2} \log \Gamma'(S_{all}) - \frac{\alpha_t}{2} \log \Gamma(S_{all}) \right\} \quad (8)$$

右辺第1項目は ML 規準の時と同じ値の反対符号になる。ここで、 $\Gamma(S_{all}) = \sum_i \Gamma(S_i)$ はすべての状態のデータ数期待値を表す。 $\Gamma'(S_{all})$ は時間分割後の値を示す。時間方向分割では状態数よりフレーム数が短くなったセグメントは考慮できなくなるため、学習データ数 $\Gamma(S_{all})$ が減少する。これを補正している。また、 α_c, α'_c はそれぞれ環境方向分割での分割前と後のパラメータ数である。 K 次元特徴量を用い、各状態とも対角共分散行列の単一分布である場合、総状態数 M である時の出力確率分布のパラメータは $2KM$ である。したがって、 $\alpha_c = 2KM, \alpha'_c = 2K(M+1)$ となる。時間方向の場合、遷移確率を考慮し、パラメータは独立であると仮定すると、パラメータ数は $2KM + M$ となる。したがって、時間方向のパラメータ数は、分割前 $\alpha_t = (2K+1)M$ 、分割後 $\alpha'_t = (2K+1)(M+1)$ と表せる。したがって、式(7)、(8)は以下のようになる。

$$G_c^{(MDL)}(S_i) = -G_c^{(ML)}(S_i) + K \log \Gamma(S_{all}) \quad (9)$$

$$G_t^{(MDL)}(S_i) = -G_t^{(ML)}(S_i) + \frac{(2K+1)}{2} \{(M+1) \log \Gamma'(S_{all}) - M \log \Gamma(S_{all})\} \quad (10)$$

MDL-SSS では、各分割時に $G_c^{(MDL)}, G_t^{(MDL)}$ を最小化する状態分割を選択する。これを繰り返し行い、すべての状態で $G_c^{(MDL)} > 0$ かつ $G_t^{(MDL)} > 0$ となった時、分割を停止することにより、自動的に状態分割を停止する。

4.2 MDL-SSS 修正分割規準

式(9)、(10)は MDL 規準から直接導かれるものであるが、予備実験により、このままの条件式では自動で分割が止まらないことが判明した。本手法においては、2. で述べたように従来 ML-SSS と同じく、分割状態選択時の時間方向分割で得られる分布は近似的に計算しているため、ゆう度が大きく見積もられる傾向がある。そのため、第2項の学習データ数とパラメータ数によるペナルティが効果的に効いていないと考えられる。

また、従来手法である文献[3][4]では、環境方向分割において第2項に重み付けすることで、重みなしの時より精度が向上することが報告されている。

そこで、提案手法では式(9)、(10)の右辺第2項に、それぞれ重み係数 C_c, C_t を導入する。

$$G_c^{(MDL)}(S_i) = -G_c^{(ML)}(S_i) + C_c K \log \Gamma(S_{all}) \quad (11)$$

$$G_t^{(MDL)}(S_i) = -G_t^{(ML)}(S_i) + C_t \frac{(2K+1)}{2} \{(M+1) \log \Gamma'(S_{all}) - M \log \Gamma(S_{all})\} \quad (12)$$

上式を用いて係数 C_c, C_t により、第1項と第2項との関連度

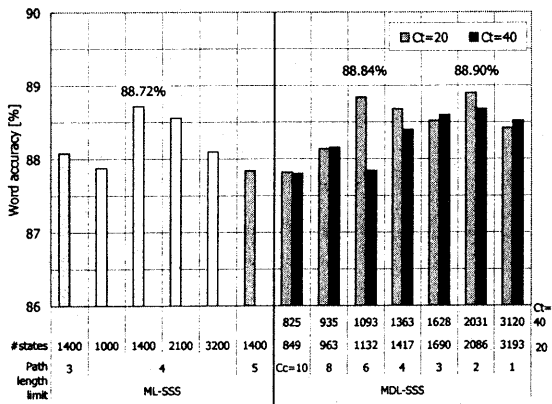


図3 ML-SSSとMDL-SSSとの単語認識率による比較

合を調整し、自動的に停止できるようにする。実際にはこれらの係数の最適な値は実験的に求める必要がある。次節で最適な値が存在するかどうか評価を行う。

5. ML-SSSとMDL-SSSとの比較実験

5.1 実験条件

ATR 旅行会話データベース[9]から“The Travel Arrangement Task (TRA)”を用いて評価実験を行った。音響モデル学習データとして、TRA データベース、話者 407 人、および音素バランス 503 文、話者 402 人による音声データを用いた。それぞれ約 5 時間、約 25 時間の音声データであった。評価データとしては、TRA から学習データに含まれない話者である 42 人による 42 片対話を用いた。

分析条件としては、16kHz サンプリング周波数、フレーム長 20msec、フレーム周期 10msec を用いた。特徴量には 12 次元 MFCC、12 次元 Δ MFCC、 Δ 対数パワーを用いている。Cepstrum Mean Subtraction は発声単位で行った。音素は 26 種類とした。各音素モデルごとに音響モデル構造を作成するとし、初期モデルは各音素とも 3 状態 3 ループのものを用いた。音響モデル構造を各状態単一分布として、従来法 (ML-SSS)、提案手法 (MDL-SSS) で作成したのち、各状態を 5 混合分布とした。男女別に学習データを分け、性別依存モデルとした。

言語モデルの学習データとしては TRA を含む 7,195 片対話 (約 160 万語) を用いた。実験に用いた語彙数は 27,398 単語であり、評価データの単語はすべて含んでいる。言語モデルとしては、多重クラス複合 bigram [10] を用いた。

5.2 ML-SSSとMDL-SSSの比較

図3にML-SSSとMDL-SSSの単語認識率を示す。ML-SSSでは状態数とパス長制限(時間方向の最大状態数)のいくつかの値に対し、評価を行った。MDL-SSSでは環境方向の係数 C_c と時間方向の係数 C_t のいくつかの値に対し評価した。式7, 8の右辺第2項は対数ゆう度差に対するペナルティになるため、 C_c, C_t の値が大きいくほど状態分割は抑制される。

実際には、時間方向は C_t の値を適切に設定することで、自動で分割停止できることがわかった。 $C_c = 2, C_t = 20$ 付近で

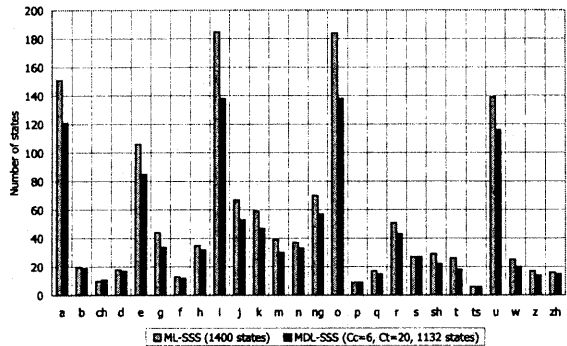


図4 各音素モデル内の総状態数

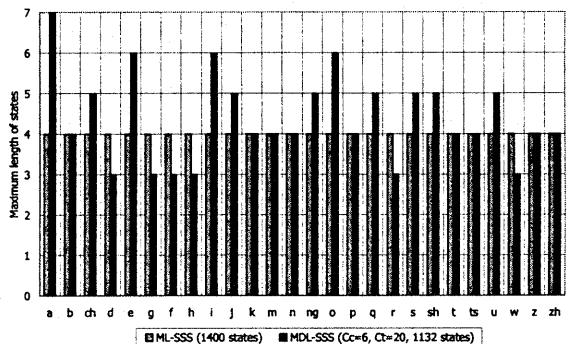


図5 各音素モデル内での最大パス長

最も高い性能得られた。ML-SSSと比較しても、適切な係数を用いれば、同等か若干よい性能が得られる場合も存在した。

また、時間方向の係数が環境方向の値に比べ極端に大きいのは、分割状態選択時の時間方向の対数ゆう度が近似値であることが理由として考えられる。

5.3 モデル構造の比較

得られたモデルの違いを詳細に調べた。図4に各音素モデルごとに見た総状態数を、図5に各音素モデルでの時間方向に対する最大状態長を示す。パス長制限 $N_p = 4$ 状態の“ML-SSS (1400 states)”と“MDL-SSS ($C_c = 6, C_t = 20, 1132$ states)”を比較した。総状態数はそれぞれ 1,400, 1,132, 単語認識率はそれぞれ 88.72%, 88.84%である。音素ごとに見た場合でも状態数は“ML-SSS (1400 states)”の方が多いが、最大状態長はすべて制限の 4 状態になっている。これに対し、“MDL-SSS ($C_c = 6, C_t = 20, 1132$ states)”では最大状態長は音素ごとに異なっている。母音は比較的長い $g/y/r$ といった子音は短い。MDL-SSSでは音素の特徴により合ったモデルができており、このため、従来法より小さい状態数でも同様の性能を出せると考えられる。また、従来法で N_p を音素ごとに設定することは可能であるが、最適値の組を見つけるのは大変困難である。

5.4 少量学習データでの比較

MDL-SSSでの係数の妥当性を見るため、学習データを少量にしたときの評価を行った。学習データにはTRAデータ、約5時間分のみを用いた。それ以外の条件は前節での実験と同じで

表 1 音素環境方向のクラスタリング手法

手法名	{ML,MDL}-SSS	{ML,MDL}-PDT-SSS	{ML,MDL}-PDT+Chou-SSS
初期分布作成	Chou アルゴリズム	音素決定木	音素決定木
繰返し	Chou アルゴリズム	-	Chou アルゴリズム

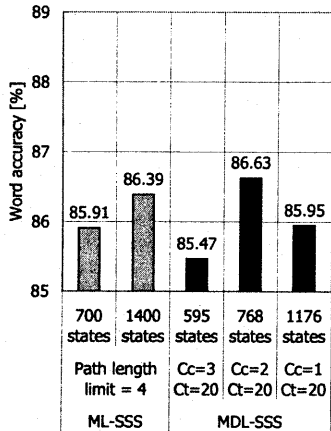


図 6 TRA データのみで学習したモデルでの単語認識率

ある。図 6 に単語認識率を示す。前節と同様、 $C_c = 2, C_t = 20$ 付近で最もよいモデルが得られた。そのときの総状態数は 768 であり、従来法の最もよいモデル (総状態数 1,400) と比較して総状態数が小さい。これは MDL 規準を用いたことで、学習データ量に応じたモデル構造が適切に得られたと考えられる。

これらの実験から、学習データ量によらず、係数を $C_c = 2, C_t = 20$ と設定できることがわかった。これらの値を用いれば自動的に N_s や N_p のほぼ最適値を見つけることができる。

6. 音素決定木クラスタリングを用いた MDL-SSS アルゴリズム

ML-SSS 法では、環境方向分割で学習データによるクラスタリングに依存しているため、学習データにない音素環境を扱えないという問題があった。文献 [8] では、未知環境を扱えるようにするため、環境方向分割に音素決定木クラスタリングを導入している。ここでは、提案手法の発展として音素決定木クラスタリングを用いた ML-SSS 法に MDL 規準を取り入れた検討を行う。

6.1 提案手法

文献 [8] において、ML-SSS 法を元に、環境方向の分割時にデータのみに基づく分布推定およびクラスタリングだけでなく、音声学の知識による音素カテゴリを利用した音素決定木クラスタリングも利用した手法が提案されている。ML-SSS 法での環境方向分割処理は大きく 2 つに分けられる。

(1) 初期分布作成

各状態 1 つの分布から 2 つの初期分布を作成する。学習データをゆう度の値に基づき、どちらの分布に近いかでクラスタリングする。

(2) 繰返し

2 つの分布の平均、分散をクラスタから再計算し、再度クラスタリングする。これを繰返す。

これで得られた分割のうち、分割前より分割規準において改善が最も大きいものを選択する。

ここで、Chou のアルゴリズムの場合、各状態 1 つの分布から 2 つの初期分布を平均値をずらすことで作成する。triphone データの集合ごとにゆう度を計算し、ゆう度の高い分布にクラスタリングする ({ML, MDL}-SSS)。

音素決定木の場合は、音素カテゴリに関する質問により、カテゴリに入るサンプルとそうでないものを分け、分布を求める ({ML, MDL}-PDT-SSS)。このとき、分布の再推定、クラスタリングの繰返しは行わない。

上記の {ML, MDL}-PDT-SSS に Chou アルゴリズムを適用した手法が組合せ手法になる ({ML, MDL}-PDT+Chou-SSS)。音素環境方向のクラスタリング手法を表 1 にまとめる。次節でこれらの手法について、前節までと同様な評価実験を行う。

6.2 音素決定木クラスタリングを用いた MDL-SSS の評価

音素カテゴリの数は無音も入れて 47 種類用意し、これを質問として用いた。そのほかの実験条件は 5.1, 5.2 と同じである。また、評価データに対する未知音素環境は 1.1% であった。表 1 に示した ML-PDT-SSS, MDL-PDT-SSS, ML-PDT+Chou-SSS, MDL-PDT+Chou-SSS 各手法を比較した。

図 7 に環境方向分割に音素決定木のみを用いた SSS 法の比較、図 8 に音素決定木と Chou アルゴリズムの組合せを用いた SSS 法の比較を示す。また、ML-SSS と MDL-SSS とは図 3 を用いて比較できる。

図 7 に示した音素決定木のみを用いた SSS 法での比較では、MDL-PDT-SSS において最も良い精度 89.00% が得られた。これは ML-PDT-SSS、さらには 5.2 の ML-SSS, MDL-SSS に比べ、改善の度合いがやや大きいモデルであった。しかし、 C_c, C_t の値による性能の変動は MDL-SSS の時に比べ、若干大きかった。MDL-SSS では、ある程度、極大点になるような点が存在したが、MDL-PDT-SSS ではそれがなく、最適な係数を決めるのがやや困難である。分割方法がデータそのものではなく、知識を用いているため、できあがった分布は学習サンプル数に重み付けしている係数と関連性が小さいと考えられる。

図 8 に示した音素決定木と Chou アルゴリズムの組合せによる SSS 法での比較では、ML-PDT+Chou-SSS が最もよい精度を出しているが、MDL-PDT+Chou-SSS においても近い性能を出した。MDL-SSS での最適値であった $C_c = 2, C_t = 20$ とは若干異なる値で最適値になっていた。また、ML-PDT+Chou-SSS との比較では、状態数が小さいモデルで同等かそれ以上の性能が出ることがわかった。

音素決定木を利用した SSS 法において MDL 規準を用いる

謝 辞

本研究は通信・放送機構の研究委託「大規模コーパスベース音声対話翻訳技術の研究開発」により実施したものである。

文 献

- [1] S. J. Young, J. J. Odell, and P. C. Woodland, "Tree-based state tying for high accuracy acoustic modeling," Proc. of the ARPA Workshop on Human Language Technology, pp. 307-312, 1994.
- [2] J. Takami, S. Sagayama, "A Successive State Splitting Algorithm for Efficient Allophone Modeling," Proc. ICASSP'92, vol. 1, pp. 573-576, 1992.
- [3] Koichi Shinoda, Takao Watanabe, "MDL-based context-dependent subword modeling for speech recognition," The Journal of the Acoustical Society of Japan (E), vol. 21, no. 2, pp. 79-86, 2000.
- [4] Wu Chou, W. Reichl, "Decision Tree State Tying Based on Penalized Bayesian Information Criterion," Proc. of ICASSP'99, vol. I, pp. 345-348, 1999.
- [5] M. Ostendorf, H. Singer, "HMM topology design using maximum likelihood successive state splitting," Computer Speech and Language, vol. 11, pp. 17-41, 1997.
- [6] J. Rissanen, "Universal Coding, Information, Prediction, and Estimation," IEEE Trans. on IT, vol. IT-30, no. 4, pp. 629-636, 1984.
- [7] P. A. Chou, "Optimal Partitioning for Classification and Regression Trees," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, pp. 340-354, 1991.
- [8] H. Singer, A. Nakamura, "Unified Framework for Acoustic Topology Modelling: ML-SSS and Question-Based Decision Trees," Proc. of EUROSPEECH'99, vol. 3, pp. 1355-1358, 1999.
- [9] T. Takezawa, T. Morimoto, Y. Sagisaka, "Speech and Language Databases for Speech Translation Research in ATR," Proc. of the 1st International Workshop on East-Asian Language Resources and Evaluation (EALREW'98), 1998.
- [10] H. Yamamoto, Y. Sagisaka, "Multi-Class Composite N-gram Based on Connection Direction," Proc. of ICASSP'99, vol. 1, pp. 533-536, 1999.

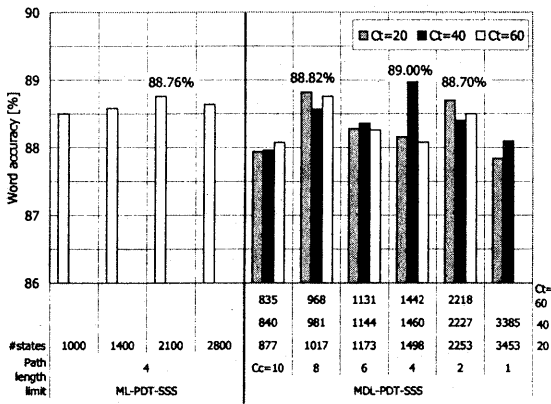


図7 音素決定木を用いた SSS での単語認識率による比較

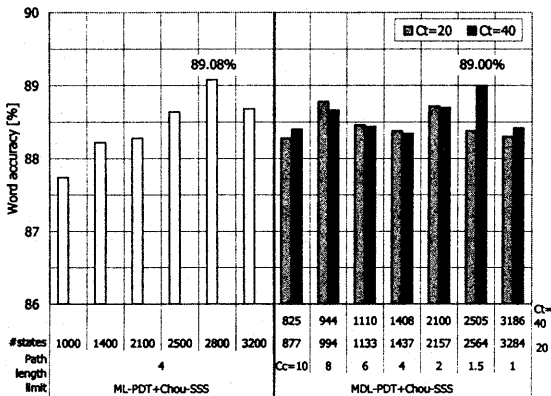


図8 音素決定木と Chou アルゴリズムを組合せた SSS での単語認識率による比較

ことで、精度的に大きな改善は得られなかったが、MDL-SSSと同様に、自動的に従来法と同程度の精度を得られることがわかった。 C_c , C_t の 2 つの係数については、手法が異なるため、最適値は異なることがわかった。

7. ま と め

最ゆう逐次状態分割法 (ML-SSS) において MDL 規準を分割および停止条件として導入した。これにより、2 つ係数は必要とするものの、自動でほぼ最適なモデルを得られることがわかった。また、音素モデルごとに最大パス長を調べると従来法より、時間方向のモデル化が精度よくなっていることがわかった。さらに、2 つの係数は学習データ量が変化した場合でも同じような最適値をとることがわかり、それらの係数を 1 度使うだけで、ほぼ最適なモデルが得られることがわかった。さらに、音素決定木クラスタリングを用いた SSS 法においても MDL 規準を取り入れ、検討を行った。音素決定木と Chou アルゴリズムの組合せで、自動的に同等の性能が得られることがわかった。2 つの係数に対する最適値は手法それぞれに依存することがわかった。