

論文 / 著書情報
Article / Book Information

論題(和文)	MLLRにおける適応データの量に応じたクラスタ数の選択法
Title(English)	A Method of deciding the number of MLLR adaptation clusters according to the amount of data
著者(和文)	張 志鵬, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会 2000年秋季講演論文集, Vol. , No. 1-5-8, pp. 15-16
Citation(English)	, Vol. , No. 1-5-8, pp. 15-16
発行日 / Pub. date	2000, 9

MLLRにおける適応データの量に応じたクラスタ数の選択法*

◎張 志鵬 古井 貞熙 (東工大)

1. はじめに

我々はこれまでに、話者交代を混合ガウス分布モデル(GMM)の尤度比較によって自動的に検出しながら、オンライン逐次話者適応を行う手法[1]を提案し、ニュース音声認識におけるその有効性を確認した。この適応法では、MLLR,MAP及びVFS法を併用した。MLLRにおいて、適応に一文を用いた場合でも音素クラスタごとに変換を行う方が単一変換行列より性能が良い。一文ずつ用いる教師なし逐次オンライン適応の場合、7クラスタの条件で一番良い結果が得られた。適応データが更に増加すれば最適クラスタ数も変動すると考えられる。一般に、モデルの自由パラメータ数の多い複雑なモデルを用いると、データ量が少ない時に性能悪化になるが、またモデルの自由パラメータ数の少ない簡単なモデルを用いると、データ量が多い時に性能が低く抑えられる。本稿ではまず、適応データ量と適応による改善効果を考察し、MDL基準により適応データ量に応じてクラスタ数を決定する手法を提案する。

2. 音響モデルのクラスタ化とMDL基準

2.1 音響モデルのクラスタ化

音響モデルのクラスタ化は、音声認識でよく用いられる方法の一つである。この手法は声質の近いモデル(音素、状態あるいは混合)が同じグループになるようにクラスタリングを行う。例えば、Triphone音響モデルの構築において音素クラスタリングし、同じクラスタに属するTriphoneがパラメータを共有するようにモデル化が行われる[2]。またMLLR適応では前述したように1クラスタよりも音素クラスタリングを行う方が有効である。

2.2 MDL基準

MDL(minimum description length)基準[3]はモデルの最適化によく利用される基準である。この基準はデータに対し、最適な確率モデルを選択する方法として有効である。MDL基準では、モデル

$i=1,\dots,M$ のうち、データ $X^N=\{x_1,\dots,x_N\}$ の記述長を最小にするモデルが最適なモデルであるとする。記述長は以下の式で表される。

$$l^{(i)} = -\log P_{\theta(i)}(X^N) + \frac{k_i}{2} \log N + \log M$$

ここで、 k_i はモデル i の次数、 $P_{\theta(i)}(X^N)$ はデータ X^N に対するモデル i のパラメータ $\theta(i)$ の最尤推定量である。第一項はデータに対する対数尤度を記号反転させたもの、第二項はモデルの複雑さを表す量である。モデルが複雑になるほど、第一項は小さくなり、第二項は大きくなる。このように両者の間にトレードオフがあり、ある適当な複雑さを持ったモデルが最適なモデルとして選択される。

2.3 MDL基準を用いたクラスタ数の選択法

ここではMLLRにおける適応データの量に応じたクラスタ数の決定を対象とする。適応データ量に応じてMLLRのクラスタ数を決定するために、モデルの次数 k_i としてクラスタ数の対数 $\log C$ を用い、パラメータ α を加えてMDLの式を次のように修正する。

$$l^{(i)} = -\log P_{\theta(i)}(X^N) + \alpha \times \frac{\log C}{2} \log N + \log M$$

ある N で α の値を決めれば、 $l^{(i)}$ の最小化によって最適なクラスタ数 C が決定される。

3. 認識実験

3.1 音響モデル

今回の実験で用いたHMMは、tree-based clusteringによって状態共有化を行なった文脈依存音素HMMである。音響特徴量としては16次のLPCケプストラムと正規化対数パワー、及びそれらの一次微分の計34次元を用いた。モデルの総状態数は2083である。各状態のガウス分布の混合数はすべて4である。

3.2 言語モデル

言語モデルの学習に用いたデータは放送ニュー

*A method of deciding the number of MLLR adaptation clusters according to the amount of data

By Zhipeng Zhang and Sadaoki Furui (Tokyo Institute of Technology)

ス原稿テキスト5年分、約50万文である。単語出現頻度上位2万語を認識語彙とし、間投詞と読みを考慮した言語モデル[4]を用いた。

3.3 評価用データ

実際に放送されたニュース音声から、スタジオで収録された女性のクリーンな発話18文を抽出した。適応データは同じ話者の1000文余りを用いた。

4. 認識実験結果

4.1 1クラスタMLLRのときの適応文数の効果

1クラスタの場合について適応文の数を変えて、MLLRの性能を考察した。結果は図1に示すとおりである。図にはSD(特定話者)モデルの結果を合わせて示す。適応文数が100の場合に一番良い結果が得られることが分かる。100文を超えると適応データを増やしても性能が上がらず、かえって下がる。このことは、クラスタ数が少ないときのMLLRのような変換手法の限界を示していると考えられる。従って適応データ量によって最適なクラスタ数を決める必要がある。

4.2 MDLによる最適クラスタ数の決定

適応文数を変えた時に最適なクラスタが達成できる α の値を実験的に求めた。その結果、各適応文数の条件で最適なクラスタ数を求めると、図2の中の星印に示す結果が得られる。図2には各適応文数の条件で各クラスタ数における単語誤り率の二次近似曲線を表す。MDL基準によって最適条件に近いクラスタ数が得られることがわかる。

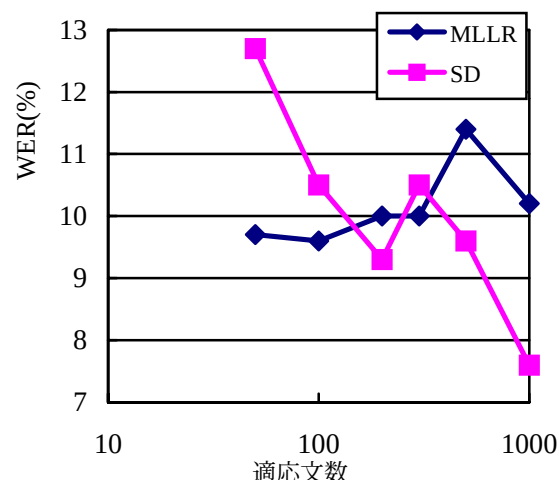


図1. 1クラスタの場合MLLRとSDモデルの適応文数と単語誤り率の関係

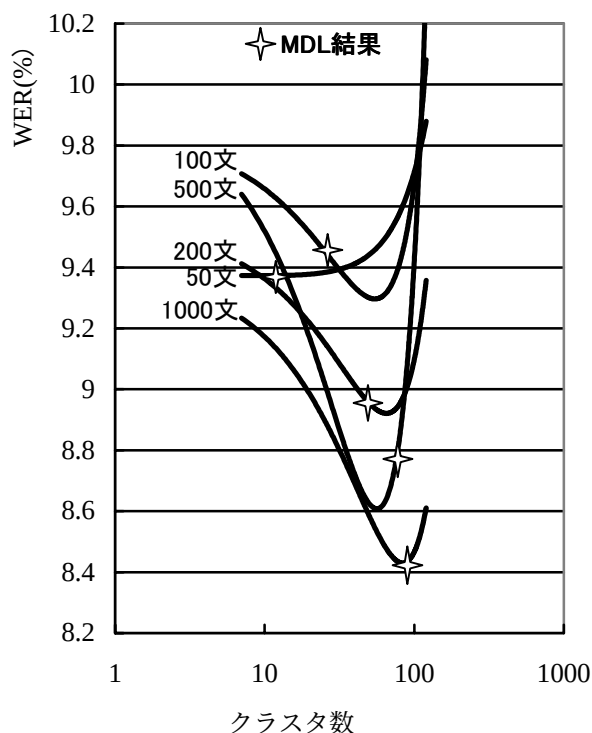


図2. MDLによるクラスタの決定結果

5. まとめ

MLLRにおける適応データ量と改善効果の関係を考察した。MDL基準によって適応データ量に応じたクラスタ数を決定する手法を提案した。MLLRのクラスタに適用するためMDL基準を修正し、パラメータを追加した。実験によって提案した手法の有効性が確認された。今後は他の音声変動を複合的に含む場合への対処法について研究を進める予定である。

謝辞

ニュース原稿及び音声データを提供して頂いたNHK放送技術研究所に感謝します。日頃討論頂く東工大の研究室の方々に感謝します。

参考文献

- [1] 張,古井, 音講論, pp.45-46, 1999-9.
- [2] S. Young et al., HTK BOOK, Cambridge Univ, 1995
- [3] Koichi Shinoda and Takao Watanabe, Journal of ASJ (E), Vol.19, No.2, pp.79-86, 2000.
- [4] 桜井 他, 音講論, pp.57-58, 1999-3.