論文 / 著書情報
Article / Book Information

論題(和文)	半音節HMMを用いた音声認識のための教師なし適応化
Title(English)	
著者(和文)	篠田浩一, 渡辺隆夫
Authors(English)	Koichi Shinoda
出典(和文)	日本音響学会平成6年度春季研究発表会講演論文集, Vol. 3-7-8, No. , pp.
Citation(English)	, Vol. 3-7-8, No. , pp.
発行日 / Pub. date	1994,

3—7—8　半音節 HMM を用いた音声認識のための教師なし適応化 *

⊙篠田 浩一　渡辺 隆夫
(NEC 情報メディア研究所)

1. はじめに

教師なし話者適応化は、教師あり適応化に比べ一般に性能は低い。しかし、使用者の側から見た場合、教師なし適応化のほうが適応化の際の拘束が少なく使いやすい。教師なし適応化で教師あり適応化にほぼ匹敵する性能が得られれば、教師なし適応化の方が好ましいと言える。

教師なし適応化の方法は、音韻表記を全く用いない方法 [1, 2] と、入力音声を認識し認識結果を利用して適応化する方法 [3, 4, 5, 6, 7, 8] の 2 つに分けられる。前者の方法では、音声の特徴量に対しベクトル量子化を行ない、標準話者のベクトル量子化コードブックと未知話者の入力発声から作成されたベクトル量子化コードブックとの間の写像を作成する手法が主である。一定の効果は望めるものの、改善の度合はさほど大きくない。後者では、認識結果を教師とした教師あり適応化を行なう。認識誤りが不可避で安定に動作しない可能性があるが、ベースとなる認識システムの性能が高い場合には、適応化の効果は大きい。後者の方法は、認識対象を制限する必要があるので、任意の発声を受け付けるわけではなく厳密な意味での教師なし適応化ではない。しかしながら、実用においては、認識装置使用時の発声を用いて適応化するという使い方が想定されるので、認識時の認識対象を適応化のときにも用いることにすれば支障はない。また、認識対象として連続音節を用いれば、事実上認識対象を限定しない場合と等価になる。今回、教師あり適応化手法としてスペクトル内挿話者適応化 [9, 10] を用い、また、認識システムとして半音節を認識単位とした混合連続分布 HMM を用いて、認識結果を用いる教師なし適応化法の評価実験を行なった。

2. スペクトル内挿話者適応化 [9,10]

混合分布連続 HMM を用いた認識システムを考える。本適応化では、各分布 j の平均ベクトル μ_j の移動量を求める。この移動量を適応化ベクトルと呼び、Δ_j と書く。

まず、適応化用データを用い、各分布の適応化ベクトル Δ_j を求める。他のパラメータを固定して平均ベクトルの Viterbi 学習を行う。Viterbi Alignment の過程では、同一状態内の各分布のうち出現確率の最も大きい分布が選ばれる。適応化用データが極めて少量の場合、Viterbi 学習で学習されない、すなわち、対応する適応化用データのない、分布が存在する。これらの分布の適応化ベクトルを以下のスペクトル内挿により求める。

$$\Delta_i^B = \sum_j w_{ij}\Delta_j^A \quad (1)$$

$$w_{ij} = \frac{\| \mu_i^B - \mu_j^A \|^{-m}}{\sum_{j'} \| \mu_i^B - \mu_{j'}^A \|^{-m}} \quad (2)$$

ここで A は適応化用データありの分布を示す添字、B は適応化データなしの分布を示す添字である。m は非負の実数パラメータである。

3. 教師なし話者適応化

認識結果を用いた教師なし適応化法は、認識システムを用いて得られた認識結果を適応化にフィードバックし、それを用いて適応化を行なう方式である。認識結果として得られる情報は、認識結果単語、正解単語の尤度、その他の単語の尤度などである。今回は、そのうち、正解単語の音韻表記のみを教師として用いることとする。方式は以下の 2 つのステップから構成される。

1.初期モデルを用いて入力パターンの認識を行ない、認識結果を出力する。これを適応化に用いるすべての入力パターンについて行なう。
2.認識結果を教師として教師あり適応化を行なう。

4. 実験

4.1. 実験条件

評価実験は半音節を認識単位とした混合ガウス分布 HMM を用い、5000 単語大語彙離散単語認識をシミュレートした類似 100 単語認識 [11] を行なった。混合ガウス分布数は 2 とした。多数話者のデータとして、男性 46 名女性 39 名計 85 名の音素バランスを考慮した 250 単語 1 回発声を用いた。また、評価話者として上の 85 名に含まれない男性 3 名 (M1~M3) 女性 4 名 (F1~F4) 計 7 名を用い、適応化用データ、および、評価用データとしてそれぞれ、学習時とは異なる語彙 250 単語 1 回発声を用いた。適応化用、評価用のデータの語彙はお互いに異なっている。

分析条件は、サンプリング周波数 16 kHz、帯域 0.1–7.2 kHz、フレーム間隔 10 ms で、メルケ

* Unsupervised Speaker Adaptation for Speech Recognition Using Demi-syllable HMM,
by Koichi SHINODA and Takao WATANABE
(NEC Corporation)

表 1: 5000 単語認識 (5000 単語認識による適応化)(%)

	M1	M2	M3	F1	F2	F3	F4	平均
不特定	78.8	89.6	88.4	79.1	86.4	82.8	86.7	84.5
特定	86.0	94.0	92.4	90.8	88.8	90.4	90.0	90.3
適応化 50 単語	84.0	86.0	90.0	81.5	84.4	79.5	85.2	84.4
	(82.4)	(90.4)	(90.0)	(83.9)	(84.4)	(83.9)	(84.8)	(85.7)
適応化 100 単語	82.8	91.2	90.8	83.1	88.0	83.1	83.6	86.1
	(87.6)	(90.0)	(90.4)	(86.3)	(87.6)	(87.1)	(86.8)	(88.0)
適応化 150 単語	87.6	91.6	92.8	87.1	88.8	85.1	88.4	88.8
	(92.4)	(92.8)	(92.4)	(89.2)	(89.2)	(90.0)	(90.0)	(90.9)
適応化 250 単語	90.8	94.4	94.4	89.2	90.0	89.2	90.8	91.3
	(92.0)	(95.2)	(94.0)	(92.0)	(90.4)	(93.6)	(91.2)	(92.6)

() 内は教師あり

表 2: 5000 単語認識 (連続音節認識による適応化)(%)

単語数	M1	M2	M3	F1	F2	F3	F4	平均
適応化 50 単語	77.2	87.6	85.2	77.9	83.2	79.5	77.6	81.2
適応化 100 単語	78.4	83.6	86.4	79.1	82.8	79.5	76.8	80.9
適応化 150 単語	78.8	87.6	87.2	79.5	85.6	81.9	80.8	83.1
適応化 250 単語	81.6	90.4	90.0	80.7	91.2	87.1	86.0	86.7

プストラム分析を用いた。特徴ベクトルは正規化パワー差分、メルケプストラム 10 次元、メルケプストラムの変化量 10 次元の計 21 次元である。

適応化の初期モデルは話者 85 名の発声データを用いて Baum-Welch アルゴリズムで学習した不特定話者モデルを用いた。(2) 式の μ_i^B と μ_j^A の間の距離はユークリッド 2 乗距離、パラメータ m は 1.0 とした。

4.2. 実験結果

まず、離散 5000 単語を認識対象とした場合について教師なし適応化の評価実験を行なった。適応化用単語数は適応化用データセットの単語番号 1 から 50 の 50 単語、1 から 100 の 100 単語、1 から 150 までの 150 単語、1 から 250 までの 250 単語の 4 通りである。話者 7 名についての実験結果を表 1に示す。表 1には教師あり適応化の結果もあわせて示す。表 1で「特定」とあるのは、単語番号 1 から 250 までの 250 単語で Baum-Welch アルゴリズムで学習した HMM を用いた特定話者認識の認識率である。教師なし適応化を行なうことにより、性能が大幅に向上した。話者 7 名平均で不特定話者認識率 84.5% のところ、適応化単語数 250 単語で教師なし適応化後の認識率 91.3% と誤りが半分近く減少している。また、教師あり適応化と比べても、各々の適応化用単語数において、1~2% 低いに過ぎない。

次に、連続音節認識を用いた認識対象を限定しない適応化の評価実験結果について述べる。日本語の音節の連鎖を有限状態オートマトンで表現したタスクで認識し、認識結果として出力された音節列を教師として適応化を行なう。5000 単語認識実験結果を表 2に示す。適応化用単語数 250 単語の場合、認識率は 250 単語で 84.5% から 86.7% へと 2.2% 向上した。しかし、離散 5000 単語を対象とした適応化の場合に比べ改善の幅は小さい。

5. おわりに

教師なし適応化の一手法として、「認識結果を教師とした教師あり適応化」を検討・評価した。離散 5000 単語を認識対象とした場合と、連続音節認識を認識対象とした場合 (事実上、認識対象を限定しない場合と同じ) の 2 つの場合について調べ、両者で効果のあることが確認できた。特に離散 5000 単語の場合には、教師あり適応化との認識性能差はわずかである。今後は、尤度差などの情報を用いた誤認識の可能性の高い単語の検出法の検討、また、連続音節認識を用いた適応化の性能向上のための連続音節認識方式の改良を行ないたい。

謝辞

日頃熱心にご討論いただく音声言語研の諸氏に感謝致します。

参考文献

[1]S.Furui: Proc.ICASSP-89, pp.286-289(1989).
[2]山下, 松本: 音響学会音声研資,SP87-118(1988).
[3]中川, 坂井: 信学論 (A),J61-D,6(1978).
[4]溝口, 木下, 角所: 信学論 (A),J67-A,6(1984).
[5]杉山, 好田: 音響学会音声研資,S85-18(1985).
[6]宮沢, 大倉, 嵯峨山: 信学技報,SP92-75(1992).
[7]松岡,C.H.Lee: 音講論,2-7-13(1993-10).
[8]鶴見, 中川: 信学技報,SP93-104(1993-12).
[9]篠田, 磯, 渡辺: 音講論,1-8-12(1990-9).
[10]K.Shinoda et al.: ICASSP-91,pp.857-860(1991).
[11]渡辺, 磯谷, 塚田: 信学論 (D-II),J75-D-II,8(1992).