

論文 / 著書情報
Article / Book Information

論題(和文)	話者認識におけるモデルとしきい値の更新法の検討
Title(English)	
著者(和文)	松井知子, 西谷隆, 古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会 1996年春季講演論文集, Vol. , No. 1-5-6, pp. 11-12
Citation(English)	, Vol. , No. 1-5-6, pp. 11-12
発行日 / Pub. date	1996, 3

1 まえがき

話者照合では、入力音声と本人のモデルとの類似度(尤度)を調べ、一定のしきい値と比較して話者の判定を行う。実システムでは、話者の発声負担も考慮して、各話者の初期モデルは一時期に発声した少量のデータから作成されることが多いが、高い話者照合性能を得るためには、各話者ごとに、過去のデータを多数蓄積しておく、いろいろな発声変動(主に発声内容や発声時期の違いに起因する)を含んだそのデータを用いて、話者のモデルを再作成することが望ましい。しかしながら、そのような話者のモデルを再作成する方法は効果的である一方、必要とされるメモリ量や計算量は莫大となる。そのため、異なる時期に発声した少量の更新用データを用いて、発声変動に対して頑健になるように、話者のモデルを更新する方法、およびその話者のモデルの頑健性を考慮しながら、話者の判定に用いるしきい値を適切に設定する方法が必要となる。

筆者らは HMM による話者照合において、話者 HMM を効果的に更新する方法、および話者 HMM の更新に合わせて、しきい値を事前に設定する方法について検討し、テキスト独立型話者照合実験においてその有効性を確かめた[1]。本論文では、それらの方法をテキスト指定型話者照合に適用した結果について報告する。

2 モデルの更新

本方法では、HMM のパラメータ(状態 s 、混合分布 m の平均ベクトル μ_{sm} 、重み係数 w_{sm})を、その話者の更新用データを用いて、式(1)、式(2)に従って再推定する。

$$\hat{\mu}_{sm} = \frac{\tau_{sm}\mu_{sm} + \sum_{t=1}^{T_U} c_{smt}x_t}{\tau_{sm} + \sum_{t=1}^{T_U} c_{smt}} \quad (1)$$

$$\hat{w}_{sm} = \frac{\nu_{sm} - 1 + \sum_{t=1}^{T_U} c_{smt}}{\sum_{m=1}^M \nu_{sm} - M + \sum_{m=1}^M \sum_{t=1}^{T_U} c_{smt}} \quad (2)$$

ただし、 T_U は更新用音声の長さを表す。また c_{smt} は、パラメータ θ の HMM において、時刻 t に状態 s 、混合分布 m で x_t が出現する確率を表している。 ν_{sm} は式(3)に従い、 w_{sm} 、 τ_{sm} から計算する。

$$\nu_{sm} = w_{sm} \sum_{m=1}^M \tau_{sm} \quad (3)$$

τ_{sm} は式(4)に従って更新する。なお τ_{sm} の初期値は、初めに話者 HMM を生成する時に用いた学習用音声から計算した $\sum_{t=1}^{T_L} c_{smt}$ に設定する。ここで、 T_L はその学習用音声の長さを表す。

$$\tau'_{sm} = \tau_{sm} + \sum_{t=1}^{T_U} c_{smt} \quad (4)$$

3 事前しきい値の設定

本方法では、しきい値を二種類の誤り率、すなわち本人の棄却率(false rejection rate; FR 率)と詐称者の受率率(false acceptance rate; FA 率)が等しい、等誤り率を与える値の近傍に設定する。具体的には、データ量(つまりは話者 HMM の発声変動に対する頑健性)とオープンデータの FR 率の推定精度との関係性を考慮し、事前しきい値 ϕ を、データ量が増加して話者 HMM が発声変動に対して頑健になるにつれて、その更新用データから計算される、高めの FA 率を与える値から、等誤り率を与える値へ近付くように設定する(式(5))。このように、データ量が十分に大きい時に FR 率と FA 率が等しくなるようにしきい値を設定することは、予め話者照合システムの目的が明確でない場合、ベイズ識別の意味で最適な判別を行うことに相当する。

$$\phi = w\phi_1 + (1-w)\phi_0 \quad (5)$$

$$w = \frac{2}{1 + \exp(a \cdot k)} \quad (6)$$

ここで、 ϕ_0 は更新用データの等誤り率を与える事後しきい値を、 ϕ_1 はそのシステムの性能から推定される、高めの FA 率を与えるしきい値を表す。 k は話者モデルの更新時期数($k = 0, 1, 2, \dots$)を表す。 w はオープンデータののためのしきい値が、更新用データの等誤り率の値へ近付いていく度合を制御するパラメータで、例えば、自由パラメータ a を含む式(6)で表せる。

4 実験

4.1 条件

使用したデータは、男 20 名が約 12 カ月に渡る 5 時期(T1~T5)に発声した文章データである。文章テキストは、ATR 連続音声テキスト 503 文から抜粋した。男 10 名を登録話者、その他の男 10 名を詐称者として用いた。特徴パラメータとして、標本化周波数 12k Hz、フレーム長 32ms、フレーム周期 8ms、LPC 分析次数 16 でケプストラムを抽出した。学習では時期 T1 に発声した 10 文章を用いて、各話者のモデルを生成し、時期 T2、T3 に発声した文章を用いて、そのモデルを更新した。テストでは、2 時期 T4、T5 に発声した各 5 文章を 1 文章ずつ判定に用いた。学習で用いた各時期の 10 文章の内 5 文章は、そのテキストが全話者、全時期に共通で、残りの 5 文章のテキストは各話者、各時期ごとに異なる。テストに用いたテキストは、学習に用いたテキストとは異なるが、全話者、全時期で同じである。実験に用いたテスト音声の総数は 200 文章(20 名 × 2 時期 × 5 文章)である。1 文章長は平均 4.2 秒であった。

*Robust methods of updating model and a priori threshold in speaker recognition.
By Tomoko Matsui[†], Takashi Nishitani[‡], and Sadaaki Furui^{††} (†NTT Human Interface Labs, ‡Tokyo Inst. of Technology.)

表 1: 種々の方法による事前しきい値を用いた場合の FA、FR 率

方法	更新回数 k								
	0			1			2		
	FA 率	FR 率	平均	FA 率	FR 率	平均	FA 率	FR 率	平均
本方法	3.1	4.0	3.6	2.2	1.0	1.6	2.5	3.0	2.8
再設定 FA6%	3.1	4.0	3.6	4.5	0.0	2.3	7.9	0.0	4.0
再設定なし	3.1	4.0	3.6	5.7	1.0	3.4	7.0	1.0	4.0
等誤り率	0.0	67.0	33.5	0.2	31.0	15.6	1.1	17.0	9.1

なお本実験では、事前しきい値の設定において、式 (5) の ϕ_1 は FA 率が等誤り率よりも 6% ほど高い値に、式 (6) の α は 0.55 に設定した。

評価は、テキスト指定型話者照合実験 [2] において、本人が異なるテキストを発声した音声も、棄却すべき音声として含めて認識対象とした場合の話者・テキスト照合率で行った。各話者の声は、各音韻ごとに半連続型 HMM (3 状態、256 混合ガウス分布) を用いてモデル化した。音韻数は 41 である。

4.2 結果

図 1 に、連続する 10 文章を一度に用いて話者 HMM を更新した場合 (バッチ型)、1 文章ずつ用いて更新した場合 (逐次型) の話者・テキスト照合誤り率を示す。ここでは、モデルの更新の効果だけを見るため、FR 率と FA 率が等しくなるように事後的に設定したしきい値を用いた。“再作成”は、学習用データや前回の更新用データを含む全てのデータを用いて、Baum-Welch アルゴリズムによって、話者 HMM の各混合分布の平均、分散、重み係数を推定し、話者 HMM を再作成する方法を表す。いずれの更新の型についても、本方法に必要なメモリ量、計算量は再作成する方法よりも少ないが、本方法は再作成する方法と比べて、同等以上の性能を示した。これらの結果から、本方法による話者モデルの更新は有効であることがわかる。また、特に逐次型の更新において、本方法の方が再作成する方法よりも、更新ごとの照合誤り率のばらつきが小さかった。

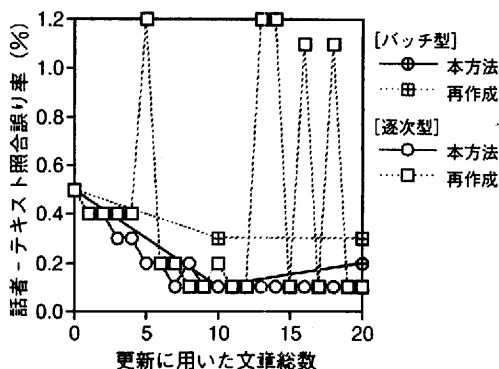


図 1: バッチ型、逐次型で話者 HMM を更新した場合の照合誤り率

表 1 に、連続する 10 文章を用いて、本方法によって話者 HMM を更新した場合 (バッチ型) について、種々の方法による事前しきい値を用いた場合の FA、FR 率、両者の平均誤り率を示す。“再設定 FA6%”は、しきい値を、学習用/更新用データの FA 率が等誤り率よりも 6% ほど高い値に設定/再設定する方法を、“再設定なし”は最初、しきい値を学習用データの FA 率が等誤り率よりも 6% ほど高い値に設定し、その後は再設定しない方法を、“等誤り率”は、学習用/更新用データの等誤り率の値に設定/再設定する方法を表す。本方法によって、事前しきい値の再設定を行った場合の平均誤り率は、学習データの等誤り率の値に設定する場合のほぼ 3 割 (2.8/9.1)、モデルやしきい値を更新しない場合のほぼ 8 割 (2.8/3.6) であった。これらの結果は、事前しきい値を再設定する本方法が有効であることを示している。また、本方法と“再設定 FA6%”を比較すると、本方法の方が FA 率が低かった。これは、モデルの更新が進むにつれて、オープンデータのためのしきい値を、その更新用データから計算される、高めの FA 率を与える値から、等誤り率を与える値へ近づけているためである。“再設定なし”では、平均誤り率が高くなった。このことから、モデルの更新に合わせて、事前しきい値を再設定する必要があることがわかる。

5 むすび

本論文では、少量の更新用データを用いて、各話者のモデルを更新する方法、および話者のモデルの更新に合わせて、事前しきい値を再設定する方法が、テキスト指定型話者照合にも効果的に適用できることを示した。

謝辞

貴重な討論をして頂いた、NTT 基礎研究所白木主任研究員、ヒューマンインタフェース研究所古井特別研究室の方々へ感謝致します。

参考文献

- [1] 松井知子、西谷隆、古井貞照、話者照合におけるモデルとしきい値の更新法、電子情報通信学会技報、SP95-120、1996。
- [2] 松井知子、古井貞照、テキスト指定型話者認識、電子通信学会論文誌、DII (5 月号掲載予定)、1996。