論文 / 著書情報
Article / Book Information

論題(和文)	マルチストリーム話者照合におけるブースティングを用いた閾値の最適化
Title(English)	
著者(和文)	浅見太一, 岩野公司, 古井貞熙
Authors(English)	Koji Iwano, Sadaoki Furui
出典(和文)	日本音響学会2005年秋季講演論文集, Vol. , No. 3-7-13, pp. 127-128
Citation(English)	, Vol. , No. 3-7-13, pp. 127-128
発行日 / Pub. date	2005, 9

マルチストリーム話者照合におけるブースティングを用いた閾値の最適化*

◎浅見太一, 岩野公司, 古井貞熙 (東工大)

1 はじめに

マルチストリーム HMM を用いた話者照合では，ストリーム重みと照合スコアの閾値を設定する必要がある．これらのパラメータの最適値はシステムが使用される環境によって異なる上，特に多数のストリームを用いる場合にはパラメータ数が増大するため，手動で最適値に設定することが非常に困難となる．実用性を考えると，各パラメータは自動的に最適化されることが望ましい．我々はこれまで，基本周波数 (韻律) 情報とスペクトル (音韻) 情報をマルチストリーム HMM によって融合させて用いる話者照合において，ストリーム重みを自動的に最適化する手法を提案している [1]．本稿では，このストリーム重み最適化の枠組みを用いて，同時に照合スコアの閾値を最適化する手法について述べる．詐称者受理誤り率 (FAR: False Acceptance Rate) と本人拒否誤り率 (FRR: False Rejection Rate) のバランスを決める照合スコアの閾値はシステムの用途によって最適値が異なるが，本研究では FAR = FRR となる閾値に設定することを目標とした．

以下では，まず，線形判別分析 (LDA) で判別式を求めることによってストリーム重みと同時に照合の閾値を推定する方法を示す．次に，Adaboost を用いて複数の線形判別式を組み合わせることによってストリーム重みと閾値を最適化する手法について説明する．最後に提案手法の有効性を確認する話者照合実験について述べる．

2 LDA による閾値の推定

音韻情報と韻律情報を融合したマルチストリーム HMM による話者照合は，次のように行われる [2]．

与えられた音韻・韻律融合特徴量 $\boldsymbol{O}_{sp}$ に対し，マルチストリーム HMM で構成された申告話者モデルと不特定話者モデルによって得られる対数尤度 (フレーム平均) $b_c(\boldsymbol{O}_{sp})$，$b_g(\boldsymbol{O}_{sp})$ は次のようになる．

$$b_c(\boldsymbol{O}_{sp}) = \lambda_s b_c(\boldsymbol{O}_s) + \lambda_p b_c(\boldsymbol{O}_p) \quad (1)$$
$$b_g(\boldsymbol{O}_{sp}) = \lambda_s b_g(\boldsymbol{O}_s) + \lambda_p b_g(\boldsymbol{O}_p) \quad (2)$$

ただし，音韻特徴量 $\boldsymbol{O}_s$ と韻律特徴量 $\boldsymbol{O}_p$ から，申告話者モデルによって出力される対数尤度を $b_c(\boldsymbol{O}_s)$，$b_c(\boldsymbol{O}_p)$，不特定話者モデルによって出力される対数尤度を $b_g(\boldsymbol{O}_s)$，$b_g(\boldsymbol{O}_p)$ とする．λ_s，λ_p は音韻・韻律ストリーム重みであり，$\lambda_s+\lambda_p = 1$ $(0 \le \lambda_s, \lambda_p \le 1)$ とする．話者照合スコア $p(\boldsymbol{O}_{sp})$ は，$b_c(\boldsymbol{O}_{sp})$ と $b_g(\boldsymbol{O}_{sp})$ を用いて，

$$p(\boldsymbol{O}_{sp}) = b_c(\boldsymbol{O}_{sp}) - b_g(\boldsymbol{O}_{sp}) \quad (3)$$

のように表せる．これは，申告話者モデルから得られた尤度を，不特定話者から得られた尤度で正規化することを意味している．これを書き換えると，

$$p(\boldsymbol{O}_{sp}) = \lambda_s p(\boldsymbol{O}_s) + \lambda_p p(\boldsymbol{O}_p) \quad (4)$$

となる．これが閾値 θ 以上であれば本人であるとして受理，θ よりも小さければ詐称者であるとして棄却するので，照合は，

$$z = \lambda_s p(\boldsymbol{O}_s) + \lambda_p p(\boldsymbol{O}_p) - \theta \quad (5)$$

の値の正負によって行うことになる．

このように，照合は $p(\boldsymbol{O}_s)$ と $p(\boldsymbol{O}_p)$ についての線形判別式によって行われる．そこで，本人であるか詐称者であるか分かっているデータについて $p(\boldsymbol{O}_s)$ と $p(\boldsymbol{O}_p)$ を計算し，LDA を行うことによって判別式 $\lambda_s p(\boldsymbol{O}_s) + \lambda_p p(\boldsymbol{O}_p) - \theta$ を求める．この λ_s，λ_p，θ を $\lambda_s + \lambda_p = 1$ となるように正規化したものをストリーム重みと照合スコアの閾値の推定値とする．

3 Adaboost によるパラメータ最適化法

Adaboost [3] は，単純な識別器を複数組み合わせることによって精度の高い識別器を構成する Boosting 法の中でも優良な性能を示す手法である．

Adaboost では，毎回の繰り返しごとにデータに付けられた重みにしたがって学習データのリサンプリングを行う．リサンプリングした学習データを使って識別器を学習し，得られた識別器の精度によって識別器に信頼度を与え，各データの重みを変更する．変更された重みを用いて，再びデータのリサンプリングから繰り返す．

ここでは，識別器として LDA によって求められる線形判別式を用いる．ただし，各回の繰り返しで得られた線形判別式の FAR と FRR が要求される条件 (本研究では FAR = FRR) からどの程度離れているかを調べ，次の繰り返しで得られた線形判別式を条件に近付くようにシフトさせる操作を加える．Adaboost で得られる最終的な識別器は，それまでに得られた識別器の信頼度による重み付き多数決となる．しかし，本研究では Adaboost の結果をマルチストリーム HMM による (5) 式の枠組みでの判別に利用するため，ブースティングの各繰り返しで得られた線形判別式の重み付き和を最終的な照合に利用する．

学習データ数 n，繰り返し回数 T のときのパラメータ最適化のアルゴリズムは以下のようになる．学習データを $\{x_i\}$ $(i = 1, \cdots, n)$，各データの重みを $\{w_i\}$ $(i = 1, \cdots, n)$ とする．$\{x_i\}$ としては，2 節の手法で推定したストリーム重みを用いたマルチストリーム HMM から計算したスコアの分布を用いる．

(1) 各データの重みを $w_i := 1/n$ で初期化する．

(2) $t = 1, \cdots, T$ で以下を実行する．

i) $\{w_i\}$ を確率分布として，$\{x_i\}$ から重複を許して n 個，重み付きリサンプリングしたものを $\{x'_i\}$ とする．

ii) $\{x'_i\}$ に対して LDA を行い，線形判別式

$$z_t = \lambda_s^{(t)} p(\boldsymbol{O}_s) + \lambda_p^{(t)} p(\boldsymbol{O}_p) - \theta^{(t)} + \delta_{t-1}$$

を得る．δ_{t-1} は前回 (z_{t-1}) の判別結果から得られるオフセット量である．ただし，$\delta_0 = 0$ とする．

iii) z_t を使って全学習データ $\{x_i\}$ に対して照合を行い，次の ϵ_t，ϵ_{FA}，ϵ_{FR} を計算し，重み付きコ

---

* A threshold optimization method using Boosting for multi-stream speaker verification. by ASAMI, Taichi, IWANO, Koji, FURUI, Sadaoki (Tokyo Institute of Technology)

スト関数 $cost_t = \omega \cdot \epsilon_{FA} + (1-\omega) \cdot \epsilon_{FR}$ (ここでは FAR = FRR を目標とするので $\omega = 0.5$) を求める.

$$\epsilon_t = \sum_{i:x_i\text{を誤識別}} w_i$$

$$\epsilon_{FA} = \frac{\sum_{i:x_i\text{を } FA} w_i}{\sum_{i:x_i\text{が詐称者}} w_i}$$

$$\epsilon_{FR} = \frac{\sum_{i:x_i\text{を } FR} w_i}{\sum_{i:x_i\text{が申告話者}} w_i}$$

iv) z_t を使って $\{x_i\}$ の照合をしたときの FAR と FRR から, $\alpha = FAR/FRR$, オフセット量 $\delta_t = 0.05 \cdot \frac{1-\alpha}{1+\alpha}$, $d_t = |FAR - FRR|$ を計算する.

v) $c_{\epsilon_t} = \frac{1}{2}\log\frac{1-\epsilon_t}{\epsilon_t}$ と $c_{d_t} = \frac{1}{2}\log\frac{1-d_t}{d_t}$ を計算し, z_t の信頼度を $c_t = c_{\epsilon_t} \cdot c_{d_t}$ とする.

vi) 次の式によって w_i を更新する.

$$w_i := \begin{cases} w_i \times e^{-c_{cost_t}} & (i:\ x_i \text{ を正しく識別}) \\ w_i \times e^{c_{cost_t}} & (i:\ x_i \text{ を誤識別}) \end{cases}$$

ここで, $c_{cost_t} = \frac{1}{2}\log\frac{1-cost_t}{cost_t}$ である.

vii) $\sum_{i=1}^{n} w_i = 1$ となるように w_i を正規化する.

(3) 最終的な識別器を z_t の重み付き和,

$$z = \sum_{t=1}^{T} (c_t \cdot z_t)$$

とする.

(4) $z = \lambda_s^{(boost)} p(\boldsymbol{O}_s) + \lambda_p^{(boost)} p(\boldsymbol{O}_p) - \theta^{(boost)}$ の係数を, $\lambda_s^{(boost)} + \lambda_p^{(boost)} = 1$ となるように正規化したときの $\lambda_s^{(boost)}$, $\lambda_p^{(boost)}$, $\theta^{(boost)}$ の値を, 最適化されたストリーム重みと照合スコアの閾値とする.

4 話者照合実験

4.1 音声データ

音声データは時期差による変化を考慮し, 1 ヵ月毎に 5 時期に渡って収録を行っている. 男性話者 36 名が 1 時期に 50 個の 4 桁連続数字を発声しており, 音声は 16kHz, 16bit で標本化・量子化した.

1 ～ 3 時期目のデータをマルチストリーム HMM の学習セットとし, 4, 5 時期目のデータを閾値の推定に使うディベロップメントセットと評価セットとして用いる. データは 12 名ずつ 3 グループに分け, 各グループを不特定話者の学習セット, ディベロップメントセット, 評価セットとして用いる. 学習セットとディベロップメントセット, 評価セットの 3 グループの組み合わせの計 6 通りについて実験を行い, 得られた結果の平均によって評価を行う.

学習セットには SN 比 30dB の白色雑音を付加させ, ディベロップメントセットと評価セットは SN 比 5, 10, 15, 20, 30dB の白色雑音を付加させたものを用いる.

4.2 実験方法

照合を行う際は, まず学習セットを用いて各話者のモデルと不特定話者モデルを学習する. このとき不特定話者モデルの学習セットに申告話者が含まれないようにする. 次に評価セットと同じ SN 比の雑音が重畳したディベロップメントセットを使ってストリーム重みと照合の閾値を推定する. そして推定されたパラメータを用いて評価データからスコアを計算し, 照合を行う. このとき, 評価セット中の申告話者以外の全話者を詐称者とする.

なお, パラメータを推定する際, Adaboost の繰り返し回数は 100 回 とした.

Table 1 各 SN 比における, LDA のみで推定したパラメータを使用した場合と Adaboost によって最適化したパラメータを使用した場合の FAR (上段) と FRR (下段) の比較.

SNR (dB)	LDA only	Adaboost
30	1.22	1.69
	0.51	0.67
20	4.82	3.56
	2.68	3.89
15	8.15	9.42
	11.40	8.70
10	9.55	17.03
	27.60	15.89
5	10.90	25.86
	50.89	23.54

4.3 実験結果

各 SN 比において, LDA のみで推定したストリーム重みと閾値を用いて照合を行った場合と, Adaboost によって最適化したパラメータを用いた場合の FAR と FRR を表 1 に示す. Adaboost でパラメータを最適化することによって 30 dB 以外の全ての SN 比において FAR と FRR が近付いていることが分かる. SN 比 30 dB において FAR と FRR の差が広がったことと, それ以外の SN 比においても完全に FAR = FRR とならなかった原因は, パラメータの推定に用いたディベロップメントセットと評価セット間にスコア分布のずれがあるためと考えられる.

5 まとめ

話者照合の閾値を, ストリーム重みと同時に Adaboost によって最適化する手法を提案し, 4 桁連続数字音声を用いた実験において本手法の有効性を確認した. 提案手法を用いることによって, SN 比 30 dB 以外の全ての SN 比において FAR = FRR に近いバランスとなる閾値が得られた.

今後の課題としては, FAR と FRR のバランスを自由に調整できる閾値最適化法の検討, より多数のストリームを用いた話者照合における本手法の有効性の確認, 本人であるか詐称者であるかが分かっているディベロップメントセットではなく, 正解ラベルの付いていない評価セットのみでパラメータを最適化する手法の検討などが上げられる.

参考文献

[1] 浅見太一, 岩野公司, 古井貞熙, "マルチストリーム話者照合におけるブースティングに基づく重み最適化法の検討," 信学技報, vol.104, no.542, pp.85-90 (2004-12).

[2] 浅見太一, 岩野公司, 古井貞熙, "雑音に頑健な話者照合のための基本周波数情報の利用," 信学技報, vol.104, no.87, pp.1-6 (2004-5).

[3] Y. Freund and R.E. Schapire, "A decision theoretic generalization of on-line learning and an application to boosting," Journal of Computer and System Science, 55(1), pp.119-139 (1997).