

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のためのブースティングによる重み最適化
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 2004年秋季講演論文集, Vol. , No. 2-4-8, pp. 101-102
Citation(English)	, Vol. , No. 2-4-8, pp. 101-102
発行日 / Pub. date	2004, 9

## 1 はじめに

話者照合の耐雑音性を高める手法の一つとして、マルチストリーム HMM によって、耐雑音性に優れた情報をスペクトル情報と融合させて用いる方法が注目されている。我々は、ハフ変換によって抽出した基本周波数情報を韻律情報としてスペクトル(音韻)情報と融合させることによって、話者照合の雑音に対する頑健性が向上することを確認している [1]。しかしこれまでは、マルチストリーム HMM のストリーム重みは事後的に手動で最適値に設定していた。実用性を考えると、ストリーム重みは、システムが使用される環境における最適な値にあらかじめ自動的に設定しておく必要がある。そこで本研究では、ディベロップメントセットを用いて素早くストリーム重みを最適化する方法として、線形判別分析 (LDA) と Adaboost [2] を組み合わせたストリーム重みの適応化手法を提案する。

以下では、まず、LDA で線形判別式を求めることによってマルチストリーム HMM のストリーム重みを推定できることを示す。次に、Adaboost を用いてより精度の高い判別式を求めることによって重み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 によるストリーム重み推定

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 [1]。

与えられた音韻・韻律融合特徴量  $O_{sp}$  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)  $b_c(O_{sp})$ ,  $b_g(O_{sp})$  は次のようになる。

$$b_c(O_{sp}) = \lambda_s b_c(O_s) + \lambda_p b_c(O_p) \quad (1)$$

$$b_g(O_{sp}) = \lambda_s b_g(O_s) + \lambda_p b_g(O_p) \quad (2)$$

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

$$p(O_{sp}) = b_c(O_{sp}) - b_g(O_{sp}) \quad (3)$$

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、

$$p(O_{sp}) = \lambda_s p(O_s) + \lambda_p p(O_p) \quad (4)$$

となる。これが閾値  $\theta$  以上であれば本人であるとして受理、 $\theta$  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、

$$z = \lambda_s p(O_s) + \lambda_p p(O_p) - \theta \quad (5)$$

の値の正負によって行うことになる。

このように、照合は  $p(O_s)$  と  $p(O_p)$  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  $p(O_s)$  と  $p(O_p)$  を計算し、LDA を行うことによって判別式  $\lambda_s p(O_s) + \lambda_p p(O_p) - \theta$  を得る。この  $\lambda_s, \lambda_p, \theta$  を  $\lambda_s + \lambda_p = 1$  となるように正規化したものをスト

リーム重みと照合の閾値の推定値とする。

## 3 Adaboost を用いたストリーム重み最適化法

Adaboost [2] は、単純な識別機を複数組み合わせることによって精度の高い識別機を構成する Boosting 法の中でも顕著な性能を示す手法である。以下、LDA で得られる線形判別式を識別機として用いた Adaboost のアルゴリズムと、Adaboost を用いたストリーム重み推定法について説明する。

### 3.1 Adaboost アルゴリズム

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機を学習し、得られた識別機の精度によって識別機に重みを与え、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。最終的な識別機は、それまでに得られた識別機の重み付き多数決となる。

学習データ数  $n$ , 繰り返し回数  $T$  のときの Adaboost のアルゴリズムは以下ようになる。学習データを  $\{x_i\}$  ( $i = 1, \dots, n$ ), 各データの重みを  $\{w_i\}$  ( $i = 1, \dots, n$ ) とする。

(1) 各データの重みを  $w_i := 1/n$  で初期化する。

(2)  $t = 1, \dots, T$  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

$$z_t = \lambda_s^{(t)} p(O_s) + \lambda_p^{(t)} p(O_p) - \theta^{(t)}$$

を得る。

iii)  $z_t$  を使って全学習データ  $\{x_i\}$  に対して照合を行い、次の重み付き識別誤差  $\epsilon_t$  を計算する。

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

ただし、 $0 < \epsilon_t \leq 1/2$  とする。 $\epsilon_t > 1/2$  のときは  $z_t$  の判断を逆にし、 $\epsilon_t := 1 - \epsilon_t$  とする。 $\epsilon_t = 0$  ならば  $w_i := 1/n$  として i) からやり直す。

iv)  $z_t$  の重みを  $c_t = \frac{1}{2} \log \frac{1-\epsilon_t}{\epsilon_t}$  とする。

v) 次の式によって  $w_i$  を更新する。

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

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

(3) 最終的な識別機を  $z_t$  の重み付き多数決、

$$z = \sum_{t=1}^T \{c_t \times \text{sign}(z_t)\}$$

とし、 $z$  の正負によって識別を行う。

\* A stream-weight optimization method using boosting for multi-stream speaker verification

### 3.2 Adaboost を用いたストリーム重み推定法

前節で述べた Adaboost アルゴリズムの出力は、線形判別式の形をしていないので、ストリーム重みの推定に直接用いることができない。そこで今回は近似を行い、

$$z = \sum_{t=1}^T (c_t z_t) \quad (6)$$

を Adaboost の結果として用いた。これにより  $z$  は、

$$\lambda_s^{(boost)} = \sum_{t=1}^T (c_t \lambda_s^{(t)}) \quad (7)$$

$$\lambda_p^{(boost)} = \sum_{t=1}^T (c_t \lambda_p^{(t)}) \quad (8)$$

$$\theta^{(boost)} = \sum_{t=1}^T (c_t \theta^{(t)}) \quad (9)$$

を用いて、

$$z = \lambda_s^{(boost)} p(\mathbf{O}_s) + \lambda_p^{(boost)} p(\mathbf{O}_p) - \theta^{(boost)} \quad (10)$$

と線形判別式の形に表すことができる。

この式を  $\lambda_s^{(boost)} + \lambda_p^{(boost)} = 1$  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、推定されたストリーム重みと照合の閾値とする。照合の閾値については、システムの用途によって最適な値が変化するため、今回はストリーム重みのみをこの方法で推定した。

## 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 ヶ月毎に 5 時期に渡って収録を行っている。男性話者 37 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 で標準化・量子化した。

1 ~ 3 時期目のデータをマルチストリーム HMM の学習データとした。4, 5 時期目のデータは 18 名と 19 名の 2 グループに分け、第 2 グループをストリーム重み推定のためのディベロップメントセットとし、第 1 グループを評価データとした。

学習データには 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データは 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、まず学習データを用いて各話者の個人モデルと不特定話者モデルを学習し、次に評価データと同じ 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を推定する。そして推定したストリーム重みを用いて評価データからスコアを計算し、照合を行う。

ストリーム重みを推定する際、Adaboost の繰り返し回数を 1 ~ 5 まで変化させ、それぞれの場合について照合実験を行った。

### 4.3 実験結果

SN 比 30dB において、Adaboost の繰り返し回数を変化させ、それぞれの場合で推定されたストリーム重みを用いて照合を行ったときの Detection Error Tradeoff (DET) カーブを図 1 に示す。図中の Baseline は、マルチストリーム HMM を用いず、音韻情報のみによって照合を行ったときの結果である。LDA のみで推定した重みを用いた場合 (Adaboost  $i=1$ ) から、Adaboost の繰り返し回数を増やすと、カーブが図の左下方向に移動し、繰り返し回数が 2

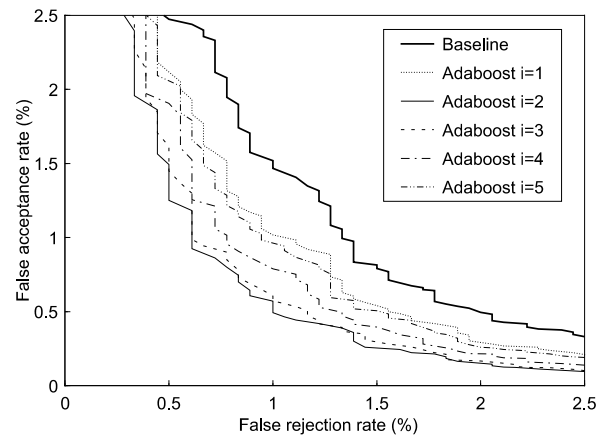


図 1. SN 比 30dB において Adaboost の繰り返し回数を変化させたときの DET カーブの推移

表 1. 各 SN 比での重み推定による EER(%) の改善

SNR(dB)	Baseline	LDA	Adaboost
30	1.24	1.01	0.79
20	5.28	4.10	3.33
15	15.92	9.49	8.86
10	27.68	16.27	16.17
5	35.73	23.63	23.63

回のときに最も良い重みを得られていることが分かる。さらに繰り返し回数を増やしていくと、カーブは右上に移動していくが、このときにも極端な性能劣化はなく、ベースラインよりも良い性能を保っている。この結果から、Adaboost を用いたストリーム重み最適化法によって、LDA のみで推定するよりも良い重みを得られることが分かる。

各 SN 比において、LDA のみで推定したストリーム重みを使用した場合と、Adaboost を用いて推定した重みを使用した場合の、照合における等誤り率 (EER) は表 1 のようになった。Adaboost の繰り返し回数は、SN 比 30dB において最も性能が良くなった 2 回とした。SN 比 30dB ~ 10dB において、LDA のみで推定した重みを用いるよりも、LDA に Adaboost を適用して推定した重みを用いた方が EER が低くなっており、5dB においても Adaboost を用いたことによる性能の劣化は見られなかった。これより、Adaboost によるストリーム重み最適化法は雑音環境においても効果があることが確認された。

## 5 まとめ

マルチストリーム HMM のストリーム重みを、LDA に Adaboost を適用して推定する手法を提案し、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、全ての SN 比において妥当なストリーム重みを得られた。

今後の課題としては、照合スコアの閾値を推定する方法の検討、3.2 節で行った近似を用いない Adaboost による重み推定法の検討などが挙げられる。

### 参考文献

- [1] 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).
- [2] 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).