

SVMとCARTの組み合わせによるAdaBoostを用いた音声区間検出*

○松田博義, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

雑音下において音声認識を行う際, 非音声の誤認識による認識精度の低下を防ぐため, 音声区間検出 (VAD: Voice Activity Detection) を行う必要がある. 本稿では音声区間検出手法として, SVMとCART(Classification And Regression Trees)を弱識別器としたAdaBoostの二つの識別器を連結した識別器を用いて音声・非音声の識別を行った. 連結はAdaBoostのアルゴリズムを用いることにより行った. 提案手法により, それぞれを個々に用いた場合より, 汎化誤差が小さく, 識別能力の高い分離平面を得ることが出来た. さらに, CENSREC-1-Cの実データを用いた実験により, 提案手法の有効性を検証した.

2 提案手法

2.1 AdaBoost

AdaBoostとは, Boostingと呼ばれる手法の一つで, 精度の低い学習器械 (弱学習器) を組み合わせることで精度の良い学習器械 (強学習器) を構成するための手法である. AdaBoostはBoostingのなかでも顕著な性能を示す手法として注目されている. AdaBoostの特徴は,

- 逐次的に学習器械を構成
- 学習データへの重み付け
- 弱学習器の重みつき結合

が挙げられる. AdaBoostは, X を特徴空間として $x \in X$ から $y \in \{+1, -1\}$ を予測する器械を例題から学習することを目的としている. Schapireらが提案している, Discrete AdaBoostと呼ばれる学習のアルゴリズム [1] を以下に示す.

1. n 個の学習サンプル集合 Z を用意する. $Z = \{(x_1, y_1), \dots, (x_n, y_n)\}$
2. 各学習サンプルに対する重みを $w_{1i} = \frac{1}{n}$ で初期化する.
3. $m = 1, \dots, M$ まで, 以下を繰り返す.
 - 重み付き学習サンプル x_i に対して弱識別器 $f_m(x) \in \{+1, -1\}$ を構成する.
 - $f_m(x)$ を用いて学習誤り err_m , 及び弱識別器への重み α_m を計算する.

$$err_m = \sum_{i=1}^n w_{mi} I\{f_m(x_i) \neq y_i\} \quad (1)$$

$$\alpha_m = \log \frac{1 - err_m}{err_m} \quad (2)$$

- 各データ x_i の重みを更新する.

$$w_{(m+1)i} = \frac{w_{mi} e^{\alpha_m I\{f_m(x_i) \neq y_i\}}}{\sum_{r=1}^n w_{mr} e^{\alpha_m I\{f_m(x_r) \neq y_r\}}} \quad (3)$$

4. 弱識別器の線形結合により, 強識別器 $F(x)$ を得る.

$$F(x) = \sum_{m=1}^M \alpha_m f_m(x) \quad (4)$$

式 (2) は, err_m が小さくなるほど, α_m は大きな値をとることを表している. すなわち, 誤りの少ない, 正確な弱学習器ほど最終的な貢献度が高くなる. さらに式 (3) により, m 段目で誤ったデータに対して α_m に応じて高い重みを与え, $m+1$ 段目で重点的に学習を行なうことにより, 最終的に正しく識別できるようにする. AdaBoostにはここで示した Discrete AdaBoost 以外にも, 弱識別器が実数値を返すものに変更した Real AdaBoost, 他 Gentle AdaBoost, LogitBoost など様々なものが考案されている [2].

一般に弱識別器には CART (Classification and Regression Trees) や Decision Stump が用いられることが多い. CARTとは次元の値の大小に基づき, 決定木を作成し分類を行なう手法である. CARTを弱識別器に用いたAdaBoostでは, データへの重み付けを変えて複数個のCARTの学習を行い, それらを精度に応じた重みを与えて線形結合することで最終的な出力を得る.

2.2 SVMとCART-AdaBoostによる識別器の構成

AdaBoostは, アルゴリズムを組み合わせるためのアルゴリズムであるので, メタアルゴリズムと捉えることが出来る. そこで, AdaBoostのアルゴリズムにおいて一段目の弱識別器に, 高度な識別を行うことが出来る識別器として知られているSVM(Support Vector Machine), 二段目に弱識別器にCARTを用いたAdaBoost(CART-AdaBoost)を配置することにより, それぞれを個々に用いた場合より高度な識別能力・汎化能力を持った識別器を構成することが可能である.

学習はまず通常通りにSVMの学習を行い, 得られたパラメータを用いて学習データを識別し, 式 (3) により, 学習データに対する重みの変更を行う. 次に, 変更された学習データに対する重みを用いてCART-AdaBoostの学習を行う. SVMは,

$$\sum_{i=1}^n y_i \alpha_i \phi(x_i) \cdot \phi(x) \quad (5)$$

(n はサポートベクターの数, $\phi(x_i)$ はサポートベクター) と表され, CART-AdaBoostは式 (4) より,

$$\sum_{m=1}^M \alpha_m f_m(x) \quad (6)$$

(f_m はCART) と表される. 最終的に得られる識別器

*Voice activity detection by AdaBoost adopting SVM & CART as weak learner. by Hiroyoshi Matsuda, Tetsuya Takiguchi, Yasuo Ariki(Kobe University)

$F(x)$ は式 (7) のように表される .

$$F(x) = \alpha_1 \left(\sum_{i=1}^n y_i \alpha_{si} \phi(x_i) \cdot \phi(x) \right) + \alpha_2 \beta \left(\sum_{m=1}^M \alpha_{cm} f_m(x) \right) \quad (7)$$

ここで、 α_1, α_2 は、それぞれ式 (2) により与えられた、SVM へ対する重み、CART-AdaBoost に対する重みである . β はこれら二つの識別器を統合するに当たり、値の正規化を図るために導入した変数である . 具体的には $\beta = \text{SVM のノルム} / \text{CART-AdaBoost のノルム}$ 、で与えられる . 図 1 は、アルゴリズムの概要図である .

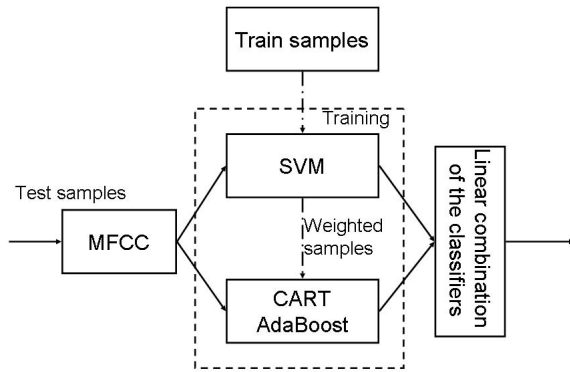


Fig. 1 提案手法の概要

ここでは、音声区間検出のため、音声データを $y = 1$ 、非音声データを $y = -1$ として実験を行った . 式 (7) において、テストデータ x が入力された際の $F(x)$ の値を閾値処理することにより、音声・非音声の判定を行った .

3 評価実験

3.1 実験条件

提案手法の評価は、VAD の評価用に設計されたデータベース CENSREC-1-C[3] を用いて行なった . CENSREC-1-C には人工的に作成したシミュレーションデータと、実環境にて収録されたデータの 2 種類が含まれている . ここでは、実際の環境での VAD 性能を評価するため、実環境にて録音された実データを用いて実験を行なった . 実データは高速道路付近 (STREET) と学生食堂 (RESTAURANT) の 2 種類の環境で、それぞれ高 SNR、低 SNR の状況で収録されたデータセットがある .

学習に用いたデータは、非音声としては、CENSREC-1-C より実験に使われなかったデータの非音声部分を切り出したもの約 20 分を用いた . 音声としては、AURORA-2J の学習用クリーン音声に上記の非音声を重畳させたものを用いた 8440 発話を用いた .

音声特徴量には MFCC と Δ MFCC を用いた . パラメータは窓幅 32 ms、シフト幅 8 ms、 Δ MFCC は 5 フレームから算出した . SVM は、RBF カーネルを用い、 γ の値は、実験より 10 とした . CART-AdaBoost の最大繰り返し数 M は 2500 とした .

実験ではフレーム単位での評価を行なった . 評価尺度は次式の FRR (False Rejection Rate) と FAR

(False Acceptance Rate) である .

$$FRR = N_{FR} / N_s \times 100[\%] \quad (8)$$

$$FAR = N_{FA} / N_{ns} \times 100[\%] \quad (9)$$

N_s は音声フレーム数、 N_{FR} は音声为非音声と検出したフレーム数、 N_{ns} は非音声フレーム数、 N_{FA} は非音声を音声と検出したフレーム数である .

3.2 実験結果及び考察

表 1 ~ 3 に、それぞれ識別器に SVM を用いた場合、CART-AdaBoost を用いた場合、提案手法を用いた場合の実験結果を示す . ここで、 FRR, FAR の値は各環境の高 SNR と低 SNR の結果を平均したものになっている . 各表の、全環境・評価尺度の誤り率平均 (表の右下の数値) を見ると、表 3 の提案手法の値が最も低い値になっていることがわかる . これは、提案手法は、SVM が作る分離平面と、それを受けて CART-AdaBoost が作る分離平面の識別能力・汎化能力を落とすことなく結合できていることを示している .

Table 1 SVM による実験結果

SVM	Restaurant	Street	Average
FRR	10.2 %	23.0 %	16.6 %
FAR	28.2 %	4.3 %	16.2 %
Average	19.2 %	13.7 %	16.4 %

Table 2 CART-AdaBoost による実験結果

CART-AdaBoost	Restaurant	Street	Average
FRR	9.7 %	18.6 %	14.1 %
FAR	25.4 %	3.9 %	14.7 %
Average	17.6 %	11.2 %	14.4 %

Table 3 提案手法による実験結果

Proposed	Restaurant	Street	Average
FRR	11.6 %	15.2 %	13.4 %
FAR	22.0 %	4.9 %	13.5 %
Average	16.8 %	10.1 %	13.4 %

4 まとめ

SVM と CART を弱識別器に持つ AdaBoost を、AdaBoost を用いて連結した識別器を用いて実験を行った . 結果は、それぞれを個別に用いた場合より低い識別誤差を得ることが出来た . 今後は弱識別器に直接 SVM を用いるのではなく、カーネル関数を実装した弱識別器の考案、マルチクラス化等を予定している .

参考文献

- [1] R. Schapire, Y. Freund, *et al.*, Annals of Statistics, pp. 1651-1686, 1998, Oct. vol.26, no.5.
- [2] Jerome Friedman, *et al.*, The Annals of Statistics, pp.337-407, 2000, Vol. 28, No. 2.
- [3] 北岡他, 音講論, 1-P-27, pp. 103-104, Sept. 2006.