

# AdaBoost を用いた雑音の検出と識別\*

三宅信之, 滝口哲也, 有木康雄 (神戸大)

## 1 はじめに

雑音が音声に混入することで音声認識率は低下する。そのため雑音の多い実環境下で音声認識システムを使用するためには雑音抑圧が重要になる。カーナビを音声で使用することを考えた場合、車内には走行音といった比較的定常な音だけでなく、雨の音や工事現場の音など非定常な雑音も存在し、また携帯電話などがなることもある。そのような雑音が発話中に突発的に発生し音声に重畳する場合、その重畳している区間の推定が困難であるため、雑音を抑圧することは難しい。本稿ではそのように突発的に発生した音を AdaBoost を用いて検出、識別する方法について検討する。

## 2 AdaBoost

多数の単純な弱識別器を積み付け加算していき、その結果に応じて出力を決定する強識別器を作成する手法は Boosting と呼ばれる。AdaBoost はその弱識別器と重みを逐次的に決定していく学習アルゴリズムである [1]。二値判別に対し高速、高精度であることから顔などのオブジェクト検出などに利用されている。また音声区間検出に用いられている例もある [2][3]。そのアルゴリズムを図 1 に示す。

AdaBoost では、多次元の特徴ベクトル  $x_i$  とそのラベル  $y_i = \{-1, 1\}$  の組である学習データ  $z_i = (x_i, y_i)$  に  $w_t(z_i)$  の重みを付け、誤識別するデータ重みの和を最小にする弱識別器  $h_t(x)$  を学習する。 $h_t(x)$  の識別能力に応じ、識別器重み  $\alpha_t$  を決める。識別に誤ったデータの次の重み  $w_{t+1}(z_i)$  を大きくし、次の識別器  $h_{t+1}(x)$  を学習する。このステップを繰り返す。学習が進むほど、識別しづらいデータの重みが大きくなり、それらを識別できるような弱識別器が構成できる。弱識別器を統合することで構成される最終的な識別器は、非常に複雑な識別関数となる。

## 3 提案手法

### 3.1 雑音の検出

AdaBoost を用いて音声中に重畳した雑音の検出を行う。学習では、雑音が重畳した音声と、雑音が重畳していないクリーンな音声を使用し、これらを分離できる二値判別の識別器を AdaBoost で学習する。学習した識別器を用いて、雑音重畳音声であるか音声

**Input:**  $n$  examples  $Z = \{(x_1, y_1), \dots, (x_n, y_n)\}$

**Initialize:**  $w_1(z_i) = 1/n$  for all  $i = 1 \dots n$

**Do for**  $t = 1, \dots, T$ ,

1. Train a base learner with respect to weighted example distribution  $w_t$  and obtain hypothesis

$h_t : \mathbf{x} \mapsto \{-1, 1\}$

2. Calculate the training error  $\epsilon_t$  of  $h_t$ :

$$\epsilon_t = \sum_{i=1}^n w_t(z_i) \frac{I(h_t(\mathbf{x}_i) \neq y_i) + 1}{2}.$$

3. Set

$$\alpha_t = \log \frac{1 - \epsilon_t}{\epsilon_t}$$

4. Update example distribution  $w_t$ :

$$w_{t+1}(z_i) = \frac{w_t(z_i) \exp\{\alpha_t I(h_t(\mathbf{x}_i) \neq y_i)\}}{\sum_{j=1}^n w_t(z_j) \exp\{\alpha_t I(h_t(\mathbf{x}_j) \neq y_j)\}}.$$

**Output:** final hypothesis:

$$f(\mathbf{x}) = \frac{1}{\|\alpha\|_1} \sum_t \alpha_t h_t(\mathbf{x}).$$

Fig. 1 Adaboost アルゴリズム

のみであるかをフレーム単位で識別する。

### 3.2 雑音の識別

前述の AdaBoost は二値判別であるため、雑音の検出は可能であるが、雑音の種類が複数ある場合、それらを識別することはできない。そのため、AdaBoost を二値判別できるように拡張する。そのアルゴリズムを図 2 に示す。これは one-vs-rest の AdaBoost を雑音の種類だけ作成し、最もスコアの良かった雑音の結果として出力する。

それぞれの種類の雑音に別々のラベルを振り、それらの雑音を重畳させた音声を学習データとして用い、音声中でも雑音の種類を判別できるような識別器を作成する。

3.1 において、雑音と検出されたフレームをこの拡張した AdaBoost を用いて雑音の識別を行う。

## 4 評価実験

### 4.1 実験条件

学習発話データは ASJ 研究用データベースから 150 発話 (男性話者 15 人) を用い、学習雑音データとし

\*Noise detection and discrimination using AdaBoost, by MIYAKE Nobuyuki, TAKIGUCHI Tetsuya, ARIKI Yasuo (Kobe University)

(1) 学習  
 $K$  クラスの学習データ  $(x_1, y_1), \dots, (x_n, y_n)$   
 $y_i = \{1, \dots, K\}$   
for  $k = 1, \dots, K$   
(1.1) データのラベル付け  
if  $k = y_i$   $y_i^k \leftarrow +1$   
else  $y_i^k \leftarrow -1$   
(1.2)  $(x_1, y_1^k), \dots, (x_n, y_n^k)$  を用い AdaBoost を学習  

$$H^k(x) = \left( \frac{1}{\sum_t \alpha_t^k} \sum_t \alpha_t^k h_t^k(x) \right)$$
  
(2) クラス識別  

$$C(x) = \operatorname{argmax}_k H^k(x)$$

Fig. 2 Adaboost による多クラス判別

Table 1 AdaBoost による実験結果

	Test SNR				
	-5dB	0dB	5dB	10dB	15dB
再現率	99.75	98.89	96.34	87.44	63.61
適合率	89.22	89.72	90.06	89.73	87.08
ノイズ識別率	98.85	97.99	95.83	92.06	87.12

て RWCP 非音声ドライソース [4] から紙を破る音, 電話音, スプレー音の 3 種類を用いる. この雑音データは各種類 50 サンプルの雑音を用意されており, そのうち 30 サンプルを学習に, 残りをテストに用いる.

学習時には音声データに各々の雑音を SNR を -5 ~ 5 dB の範囲でランダムで調整して重畳させ, クリーンな発話と雑音が重畳した発話を使用して検出器と識別器をひとつずつ作成する.

評価用データには同データベースから 1204 発話 (男性話者 5 人) を使用した. 各発話に, 継続時間 200 ms 程度に調整した雑音を区間が被らないように 1~3 つ重畳し, フレーム毎に雑音を検出, 識別していく. フレーム幅 20 ms, フレームシフト幅 10 ms, 特徴量は 24 次元対数メルスペクトルと 12 次元 MFCC を連結したもので,  $1 - 0.97z^{-1}$  のプリエンファシスを用いている.

評価法としてフレーム単位による再現率, 適合率, ノイズ識別率を算出した. ノイズ識別率は検出に成功したフレームのうち, ノイズ識別が正しかったフレームの割合である.

#### 4.2 実験結果

AdaBoost による雑音の検出と識別の実験結果を表 1 に示す. 15 dB で再現率が 63 % となるが, その他は基本的に高い値を示している.

Table 2 AdaBoost と平滑化による実験結果

	Test SNR				
	-5dB	0dB	5dB	10dB	15dB
再現率	99.54	99.01	97.50	89.99	65.20
適合率	92.29	92.88	93.35	93.49	91.95
ノイズ識別率	99.70	99.23	97.75	94.61	91.19

Table 3 GMM と平滑化による実験結果

	Test SNR				
	-5dB	0dB	5dB	10dB	15dB
再現率	98.22	93.88	83.93	60.49	28.54
適合率	96.67	97.34	97.70	97.64	95.79
ノイズ識別率	98.64	95.59	93.04	91.64	91.98

さらにフレーム単位での識別を行っているため, ごく短時間の湧き出しや誤識別を平滑化し消すことで, 精度が改善すると期待できる. そこで, 自身と前後 3 フレームを見て, その中で最も数の多いものを自身のフレームの出力とする. その結果を表 2 に示す. 結果から, 平滑化することで精度が改善していることがわかる.

また, 比較手法として GMM による識別を行った. これは 64 混合 GMM を音声と, 各雑音重畳音声について作成し, 最も尤度の高かったものを出力し, AdaBoost と同様の平滑化を行う. その結果を表 3 に示す. 比較してみると, GMM による検出では適合率が高い. しかし, 雑音が弱くなるにつれ (SNR が高くなる), 再現率が大幅に下がる問題がある. 一方, AdaBoost は適合率では劣るものの, 再現率は SNR 10 dB で 90 %, 15 dB で 65 % と, 比較して高い値を保っているのがわかる. つまり AdaBoost は GMM に比べ, 誤検出が多くなるものの, SNR が高い状況で未検出の量が大幅に減少するという結果になった.

#### 5 おわりに

AdaBoost を使用して, 音声中に突発的に重畳するような雑音の検出と識別を試みた. 背景雑音のない状況において, 多少の誤検出を含むものの, ほぼ正しく検出できることを示した. 今後これらの情報を使った雑音除去を行っていく予定である. また, 実環境においては数多くの雑音が存在するため, それらをいくつかに分類し, 実験を行う予定である.

#### 参考文献

- [1] Y.Freund *et al.*, JCSS, 55(1), pp.119-139, Aug, 1997.
- [2] O.Kwon *et al.*, IEEE, ICASSP.,SP-P2.12,2003.
- [3] 松田, 滝口, 有木, 音講論 (秋), pp.117-118, 2006.
- [4] 中村ら, 音講論 (秋), pp.137-138, 1998.