

## 特徴抽出過程におけるブースティングの適用による 相補的な識別器の生成およびその統合

加藤 健一<sup>†</sup> 小川 哲司<sup>†</sup> 小林 哲則<sup>†</sup>

<sup>†</sup> 早稲田大学 理工学部

〒 169-8555 東京都新宿区大久保 3-4-1

あらまし 本稿では、特徴変換にブースティングの枠組を適用した識別器統合手法を提案する。一般的に、複数の識別器を統合するとき、識別性能は向上することが期待できる。しかし、識別器の統合にあたって、二つの重要な課題がある。一点目は、統合する識別器各々の誤り傾向が異なっていなければ(相補性がなければ)、わずかな性能の改善しか得られない点、二点目は、相補的な識別器が生成されたとしても、各々の識別器が与える情報の統合手段が適切でない場合、やはりわずかな性能の向上しか得られないという点である。そこで本稿では、上述した二点を考慮した上で、相補的な識別器の生成手法と、その統合手法について検討を行う。相補的な識別器を生成するにあたっては、Heteroscedastic linear discriminant analysis (HLDA) に基づく特徴変換の過程でブースティングの枠組を適用した。また、統合においては、各々の識別器から出力される尤度の情報を特徴ベクトルとし、このベクトルが張る空間上で Support vector machine (SVM) に基づくパターン認識を行った。提案手法により識別器を統合することで、孤立単語音声認識実験において、統合前と比較し 74%の誤りが削減されることがわかった。

キーワード システム統合, 相補性, ブースティング, HLDA, SVM

## Combining Complementary Classifiers generated by Boosting in Feature Transformation

Ken-ichi KATO<sup>†</sup>, Tetsuji OGAWA<sup>†</sup>, and Tetsunori KOBAYASHI<sup>†</sup>

<sup>†</sup> Dept. of Computer Science, Waseda University

3-4-1 Okubo, Shinjuku-ku, Tokyo, JAPAN 169-8555

**Abstract** A framework of system combination using boosting in a feature transformation is proposed. In general, the combination of multiple classifiers improves the classification performance of each classifier. However, there are two important issues in such a system combination. First, the classification performance is not necessarily improved if the classifiers are not complementary. Second, an inappropriate combination makes the performance worse even if the complementary classifiers can be obtained. In this paper, we attempt to solve how to generate and how to combine the complementary classifiers. Aiming at generating the complementary classifiers, the boosting was applied in HLDA based feature transformation. At the combination stage, a pattern recognition using support vector machine was performed, in which a pair of the likelihoods emitted by the classifiers of the first stage was used as a feature parameter. Experimental results showed the effectiveness of proposed method: it reduced the errors by 74% compared to the case without any system combination.

**Key words** System combination, Complementary classifier, Boosting, HLDA, SVM

### 1. はじめに

近年、識別器の性能を向上させるために、確率モデルや探索アルゴリズムの精密化に注力するのではなく、複数の識別器から得られる情報を統合する枠組が検討されている。このとき、

複数の識別器を用いたとしてもそれらの誤り傾向が異なっていなければ、情報を統合することによる識別性能の向上は見込めない。また、誤り傾向の異なる識別器を生成したとしても、各識別器からどのような情報を得て、それらをどのように用いることで識別器を統合するかは、識別性能の向上に直結する重要

な課題である。そこで本稿では、音声認識の性能向上を目的として複数システムの統合を行うにあたり、誤り傾向の異なる相補的な識別器の生成手法、および生成された識別器から得られる情報を効果的に用いた識別器の統合手法について検討を行う。

複数の識別器を用いてパターン認識を行う枠組としては、これまでに様々な手法が検討されてきた。発話音声と口蓋の動画像といった異なるモダリティを用いて独立に識別器を構築する手法 [1]、異なる特徴パラメータを用いることで複数の識別器を構築する手法 [2]、異なる構造を持つ確率モデルを用いることで複数の識別器を構築する手法 [3] などが、その例として挙げられる。しかし、これらの手法は、複数の識別器が独立に作成されるため、識別器の相補性を理論的に保証するものではなく、また、統合した際の識別性能の向上はわずかである。それに対し、生成される識別器の相補性を理論的に保証する枠組として、識別器の構築においてブースティングを適用する手法があり、有効性が示されている [4] [5]。

一方、識別器の統合に関しては、性質の異なる特徴ベクトルを結合し新たな特徴ベクトルとして識別を行う枠組 (early integration) や、複数の識別器が出力する尤度の重みつき平均を新たな尤度として識別を行う枠組 (late integration) が一般的に用いられている。ブースティングにおいても、複数の識別器を線形に結合することで、識別器の統合を行っている。また、連続音声認識に限れば、単語レベルで統合を行う ROVER [6] や、Confusion Network Combination (CNC) [7] などの有効性が見出されている。ただし、これらの方法では、統合対象となる識別器のうち一つあるいは全てが、識別において不適切な特性 (例えば、特定のクラスに対して極端に低い識別性能しか得られないなど) を持つ場合、統合後もその影響が残り、識別性能が向上しない可能性がある。

以上を考慮して、我々は、Heteroscedastic linear discriminant analysis (HLDA) [8] に基づく特徴変換にブースティングを適用することで相補的な識別器を生成し、生成された複数の識別器から得られるクラスに対する尤度の情報を特徴量としてパターン認識を行うことで、識別器を統合することを試みる。本手法では、特徴変換にブースティングを適用するにあたり、基になる識別器で誤ったサンプルの尤度が、新たに生成する識別器では上昇するように変換行列を推定する。通常、ブースティングは識別器の設計時 (モデルのパラメータ推定時) に適用されるため、学習において重要視されるサンプルと軽視されるサンプルが生じる。したがって、新たに生成される識別器は相補性は得られているものの、識別性能は基の識別器よりも劣化する。それに対し本手法は、特徴変換の過程において相補性を得ようとしているため、学習の過程で軽視するサンプルは生じず、結果的に識別性能劣化の可能性を低く抑えた上で、誤り傾向の異なる識別器が生成されることが期待できる。得られた相補的な識別器を統合するにあたり、複数の識別器から得られる識別クラスの尤度情報を結合して特徴ベクトルとし、このベクトルが張る空間上で support vector machine (SVM) に基づくパターン認識を行う。このとき特徴量として用いる、全ての識別クラスに対する尤度の出力パターンは、統合対象となる各々の識別

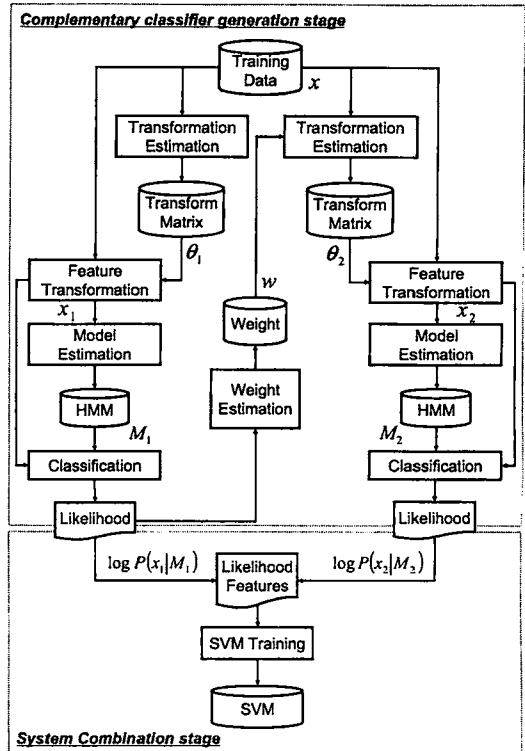


図 1 提案手法による学習の手順  
Fig. 1 The training procedure of the proposed system.

器が有する誤り傾向 (相補性) そのものと言え、識別器の相補性の情報を有効に使うことが出来る。このように、全ての識別クラスに対する尤度の出力パターンを特徴量とすることで、各識別クラスの情報のみならず、識別器の特性も含めて学習することができるため、識別器が持つ特性に対して頑健な識別を行うことが可能となる。

本論文の構成は以下の通りである。2. において、提案手法である、特徴変換においてブースティングを適用した相補的な識別器の生成方法と、尤度空間での SVM を用いた識別器の統合方法について述べる。3. において、孤立単語音声認識実験によって本手法の評価を行い、4. においてまとめを述べる。

## 2. 相補的な識別器の生成および統合

### 2.1 提案システムの概要

提案する相補的な識別器の生成とそれらの統合に関する手順を図 1 に示す。本手法では、HLDA に基づく特徴変換に対してブースティングの枠組を導入することで、相補的な識別器を生成し、得られた複数の識別器が出力する識別クラスに対する尤度の情報を特徴量としたパターン認識を行うことで、識別器を統合する。

まず、最尤基準で HLDA の変換行列  $\theta_1$  を推定する [8]。この  $\theta_1$  を用いて学習サンプル  $x$  を変換することで得られる特徴

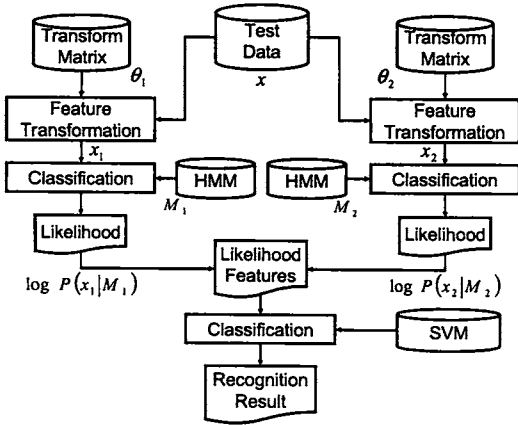


図2 提案手法による認識の手順

Fig. 2 The recognition procedure of the proposed system.

量  $x_1$  から音響モデル  $M_1$  を構築する。このようにして得られた識別器を C-1 と呼ぶ。つづいて、識別器 C-1 に対して誤り傾向の異なる識別器を生成する。識別器 C-1 を用いて学習サンプル  $x_1$  の識別を行い、各サンプルに対して誤り易さに応じた重み  $w$  を計算する。通常のブースティングでは、識別器を設計する際、具体的にはモデルのパラメータを推定する際に、誤りに応じた重みを導入するが、本手法では、この枠組を特徴変換において適用する。このとき得られる新たな変換行列は、基になる識別器 C-1 で誤りが生じ易いデータの尤度が上昇するように（誤りが生じ易いデータに重きを置いて）推定される。この、HLDA における変換行列の推定にデータ毎の重みを導入した手法を、weighted-HLDA (WHLDA) と呼ぶ。WHLDA を用いることで、識別器 C-1 で誤認識されたサンプルの識別性能を向上させることが期待できる。WHLDA で得られた新たな変換行列  $\theta_2$  を用いて特徴変換を行い、得られた学習サンプル  $x_2$  を用いて新たな音響モデル  $M_2$  を構成する。このように、WHLDA を利用して生成された識別器を C-2 と呼ぶことにする。この識別器 C-2 は、特徴変換においてブースティングの枠組を適用することで得られたものであり、識別器 C-1 に対して相補的であることが期待できる。WHLDA については 2.2 において詳細を述べる。

続いて、識別器 C-1 および C-2 を生成した後、各々の識別器が出力する尤度  $\log P(x_1|M_1)$ ,  $\log P(x_2|M_2)$  を結合したベクトルが張る空間上で SVM に基づく識別器を構築する。このようにして得られた SVM を用いてパターン認識を行うことで、識別器の統合を行う。統合に関する詳細は、2.3 において述べる。

一方、生成された複数の識別器を用いて識別を行う手順を図 2 に示す。ここでは、孤立単語認識を行うことを想定する。事前に推定しておいた変換行列  $\theta_1$ ,  $\theta_2$  を用いて評価サンプルの特徴変換を行い、対応する音響モデル  $M_1$ ,  $M_2$  を用いて各単語クラスの尤度を計算する。この、各識別器から出力される全

ての単語クラスに対する尤度を結合して特徴量とし、SVM に基づく識別を行う。

## 2.2 相補的な識別器の生成

相補的な識別器を生成するために、ブースティングの枠組を HLDA による特徴変換に適用する。ここでは、基になる識別器である C-1 と、WHLDA に基づいて生成される識別器である C-2 を生成する方法、および特徴変換に対してではなく識別器の生成においてブースティングの枠組を適用する手法との相異点について述べる。

### 2.2.1 HLDA に基づき生成する識別器

まず、HLDA に基づく特徴変換を施した特徴量を用いて、基になる識別器 C-1 を生成する。HLDA は  $d$  次元の特徴ベクトルを、識別に寄与する  $p$  次元と、識別に寄与しない  $d-p$  次元に分割する枠組である。このとき、HLDA の変換行列は以下のように表される。

$$\theta = [\theta_1 \dots \theta_d] = [\theta_p \theta_{d-p}] \quad (1)$$

ここで、 $\theta_1, \dots, \theta_d$  は各々  $d$  次元の列ベクトル、 $\theta_p$  と  $\theta_{d-p}$  は各々、 $d \times p$  行列、 $d \times (d-p)$  行列である。変換行列は、以下の式 (2) で表される目的関数を二次計画法により最大化することで得られる  $\theta$  を用いる [8]。

$$L(\theta) = -\frac{N}{2} \log |\theta_{d-p}^T T \theta_{d-p}| - \sum_{j=1}^J \log |\theta_p^T W_j \theta_p| + N \log |\theta| - \frac{Nd}{2} \log(1 + 2\pi) \quad (2)$$

ここで、 $N$  は全データの個数、 $T$  は変換前の全データに対する分散、 $W_j$  は変換前の音素クラス  $j$  に帰属するデータの分散を表す。この分散  $T$ ,  $W_j$  の計算は変換行列推定用の学習データを用いて行う。このとき、分散  $W_j$  の計算は、ピタビアルゴリズムにより各サンプルに対して帰属する音素クラスのアライメントをとったうえで、音素クラス  $j$  に帰属するデータを用いて行う。

以上の基に得られた変換行列  $\theta$  を用いて特徴変換を行った、音響モデルの学習サンプルを用いて HMM を構築する。これを、識別器 C-1 とする。

### 2.2.2 WHLDA に基づき生成する識別器

次に、識別器 C-1 に対して相補的な識別器 C-2 を、WHLDA によって変換された特徴量を用いて生成することを試みる。WHLDA では、識別器 C-1 で誤りが生じ易いデータの尤度が高くなるように、変換行列を推定する。そのため、サンプルに付与する重みは誤りの度合いが大きいほど大きくなるように設計する。重みを計算するにあたり、ここでは、各サンプル (1 フレームのデータ) に対して重みを計算する枠組と、本稿 3. で述べる孤立単語音声認識において用いた、単語発話に対して重みを計算する枠組を、各々述べる。

#### a) 各サンプルに対して重みを計算する枠組

誤り易さの尺度として、以下の式で表される誤分類測度  $d_i(x_n)$  を定義することができる。

$$d_i(x_n) = \log P_i(x_n|M) - \max_{j \neq i} \log P_j(x_n|M) \quad (3)$$

ここで、 $\mathbf{x}_n$  は  $K$  サンプルからなる発話  $\mathbf{x} = \{x_1, x_2, \dots, x_K\}$  ( $1 \leq K \leq N_u$ )、( $N_u$  は発話数) を構成するサンプルである。また、 $\log P_i(\mathbf{x}_n|M)$  は正解の音素クラス  $i$  のモデルから得られる尤度、 $\max_{j \neq i} \log P_j(\mathbf{x}_n|M)$  は不正解の音素クラスのうち、最大の尤度を与える音素クラスの尤度である。また、 $M$  は全ての音素クラスを含めた音響モデルを表す。次に、重み関数  $w(\mathbf{x}_n)$  を以下のように定義する。

$$w(\mathbf{x}_n) = \frac{1}{1 + \exp(\alpha \cdot d_i(\mathbf{x}_n))} \quad (4)$$

ここで、 $\alpha$  はシグモイド関数の傾きを制御するパラメータである。WHLDA の変換行列を推定するために式 (2) の目的関数を最適化するにあたり必要となる、全サンプルの平均、分散、およびある音素クラス  $j$  に所属するサンプルの平均、分散は、変換行列の学習データの各サンプルに式 (4) の重みを付与した上で計算される。そのため、新たに生成される識別器 C-2 では、WHLDA の変換行列を推定する際に大きな重みが付与されたサンプルの尤度が高くなることが予想される。このようにして、基の識別器 C-1 で誤り易いデータに付与した重みの情報が WHLDA における変換行列推定のための目的関数に導入される。

b) 単語発話に対して重みを計算する枠組

3. で述べる孤立単語音声認識実験では、変換行列の学習データとして用いる各単語発話に対して重みを計算する。このとき、誤分類割合と重み関数は単語発話  $\mathbf{x}$  を用いて以下のように計算される。

$$d_i(\mathbf{x}) = \log P_i(\mathbf{x}|M) - \max_{j \neq i} \log P_j(\mathbf{x}|M) \quad (5)$$

$$w(\mathbf{x}) = \frac{1}{1 + \exp(\alpha \cdot d_i(\mathbf{x}))} \quad (6)$$

ここで、 $\log P_i(\mathbf{x}|M)$  は正解単語クラス  $i$  のモデルから得られる尤度、 $\max_{j \neq i} \log P_j(\mathbf{x}|M)$  は不正解の単語クラスのうち、最大の尤度を与える単語クラスの尤度である。このように、1 つの単語に対して 1 つの重みが決まるが、各音素クラスに対する重み付きの平均、分散を計算するためには、各サンプルに対して重みを付与する必要がある。したがって、各単語発話を構成するサンプルには共通の重みが付与される。

### 2.2.3 識別器の設計時にブースティングを用いる枠組との相異点

通常、ブースティングを用いた手法は、識別器の設計時にデータに対して重み付けを行なう。したがって、学習において重要視されるサンプルと、軽視されるサンプルが生じることになる。この場合、ブースティングによって新たに生成される識別器は、基の識別器に対して相補性が与えられるものの、基の識別器と比較して識別性能は劣化する可能性がある。一方、提案手法は特徴変換の段階で相補性を得ようとするものであり、識別器の学習において偏りに軽視されるようなサンプルは生じない。そのため、モデルパラメータの推定にブースティングを適用する枠組とは異なり、新たに生成された識別器の性能は基の識別器と同等の性能となることが期待できる。

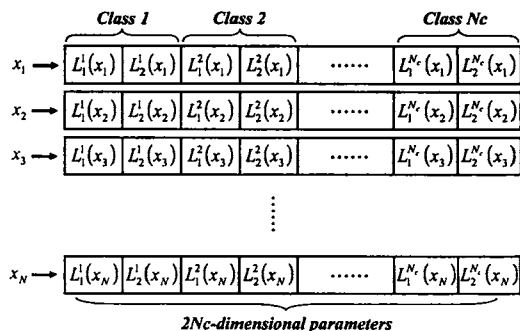


図 3 統合に用いる特徴ベクトル ( $N_c$  は識別クラス数を表す)

Fig. 3 Feature vector used in the combination stage.

( $N_c$  denotes the number of classes.)

### 2.3 相補的な識別器の統合

ここでは、2.2 で生成された相補的な識別器の統合手法について述べる。識別器を統合する手段としては、異なる性質を有する特徴ベクトルを単純に結合したものを新たな特徴ベクトルとする方法 (early integration) や、複数の識別器から得られる尤度の線形和を新たな尤度とする方法 (late integration) などが用いられることが多い。しかし、これらの方法は、統合対象となる識別器が識別器として望ましくない特性 (例えば、ある特定のクラスに対して常に高い尤度を出力する) を有していた場合、その特性が識別器を統合しても解消されずに、そのまま残る可能性がある。

一方、本手法では、複数の識別器から得られる尤度の情報を結合することで特徴ベクトルを構成し、それらが作る空間において SVM に基づくパターン認識を行うことで、識別器の統合を行う。ここでは、孤立単語音声認識を想定して、統合法の説明を行う。まず、入力される音声データに対し、全ての識別クラス (単語) に対する尤度を、識別器 C-1 と C-2 各々について計算する。そして、得られる尤度の組 ( $L_1^i(\mathbf{x}_n), L_2^i(\mathbf{x}_n)$ ) を、識別クラスの個数 (単語数) 分結合する。ここで、 $i$  は単語クラスの ID を表す。本手法では、この尤度の集合を特徴として扱う (図 3)。図 3 において、 $N_c$  は識別クラス (単語) の個数を表すものとする。この特徴ベクトルを用いて、識別クラス (単語) 数分の SVM を構築する (1 vs other)。このように、尤度の線形結合などを用いて情報を統合するのではなく、複数の識別器により与えられる尤度の空間で新たに識別器を構成する手法の利点は、全ての識別クラス (単語) に対する尤度の出力パターンを特徴ベクトルとすることで、識別クラス (単語) に関する情報のみならず、識別器そのものの特性も含めて学習することが出来る点にある。これにより、統合対象の識別器のうちひとつ、あるいは全てにおいて、ある特定のカテゴリに対して常に高い尤度を与えるというような不適切な特性が見られたとしても、その不適切な特性を吸収してモデルが学習されることになる。また、識別器の各識別クラスに対する尤度の出力パターンは、識別器の誤り傾向そのもの (相補性) の情報を保持している

と言え、識別器を生成する際に用いた相補性の情報を有効に使用することができる。そのため、ブースティングの枠組との整合性も良いことが期待できる。

### 3. 孤立単語音声認識実験

WHLDA に基づく相補的な識別器の生成とその統合の有効性を示すため、孤立単語音声認識実験を行なった。本実験では、提案手法によって生成された識別器が相補性を有することを確認するとともに、様々な統合手法と比較を行なうことで、提案した統合法の有効性を示す。

#### 3.1 実験条件

実験に使用する音声データは、16 kHz でサンプリング、16 bit で量子化されている。音響モデルの学習、および評価に用いる音響特徴パラメータとしては、MFCC 12 次元とパワー、およびそれらの  $\Delta$  パラメータを付加した計 26 次元の特徴量を用いた。このとき、フレーム長は 25ms、フレーム周期は 10ms である。音響モデルは、ASJ-JNAS、ASJ-PB [9] より、男性 133 話者が発話した音声 (計 20406 文) から学習し、状態数 5、ループ数 3 の left-to-right 型のモノフォン HMM を用いた。混合正規分布の混合数は 16、分散行列は対角共分散とした。基になる識別器 C-1 および、特徴変換にブースティングの枠組を用いて生成した識別器 C-2 で用いる音響モデルの学習および評価には、上述した MFCC を各々 HLDA、WHLDA によって特徴変換を施すことで得られる特徴量を用いる。特徴変換は静的特徴である MFCC 12 次元に対してのみ行ない、その後パワーを加え、それらの  $\Delta$  パラメータを付加した。このとき、座標変換のみを行い、次元圧縮は行っていない。すなわち、変換後の次元数も 26 次元である。HLDA および WHLDA の変換行列の推定には、ATR 音素連鎖バランス単語 216 単語を、男性話者 12 名が各単語につき 5 回発話した音声 (計 12960 単語) を用いる。評価には、ATR 音素連鎖バランス単語より、変換行列を推定するために用いたデータとは異なる男性 8 話者が各単語につき 5 回発話した計 8640 単語に対し、同一の計算機雑音を 20dB で重畳して作成した音声を用いた。

#### 3.2 識別器の相補性

統合を行わない識別器単体での認識率は、識別器 C-1 で 95.0%、識別器 C-2 で 95.0% であった。また、識別器の相補性を示す尺度の一つとして、一方の識別器で不正解であったデータがもう一方の識別器において正解となる割合を考える。C-1 で誤識別されたデータのうち、C-2 で正しく識別されるものの割合は 34.1%、C-2 で誤識別されたデータのうち、C-1 で正しく識別されるものの割合は 33.8% となり、これより、C-1 と C-2 は異なる誤り傾向を持つと言える。さらに、基の識別器 C-1 と比較して、ブースティングに基づいて生成された識別器 C-2 の性能が劣化していない。以上より、特徴変換においてブースティングの枠組みを適用することで、識別器単体でも高い識別性能を維持したまま、相補性を有する識別器を生成できることを確認した。

#### 3.3 識別器統合の効果

表 1 に識別器の統合に関する比較項目を示す。H および wH

表 1 評価項目 ("C-1" は基になる識別器を、"C-2" はブースティングの枠組を適用した相補的な識別器を生成するための特徴変換を表す)

Table 1 Evaluation items. ("C-1" and "C-2" represent the feature transformations for generating first classifier and second classifier.)

classifier	C-1	C-2	combination
H	HLDA	—	—
wH	—	WHLDA	—
H-wH-EI	HLDA	WHLDA	early integration
H-wH-LI	HLDA	WHLDA	late integration
H-wH-SVM	HLDA	WHLDA	SVM

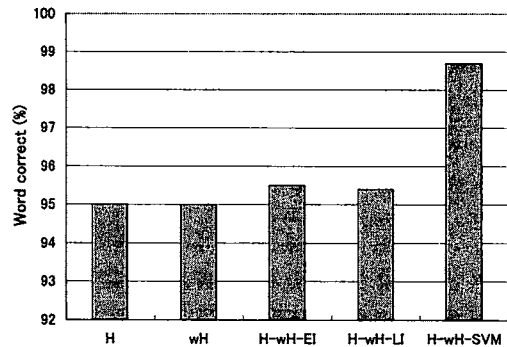


図 4 各評価項目による単語認識率

Fig. 4 Word correct for evaluation items.

は統合処理を行っていない単体の識別器を用いたときの識別性能であり、H は HLDA で変換した特徴量、wH は WHLDA で変換した特徴量を用いて構成されている。H-wH-EI、H-wH-LI、H-wH-SVM は各々統合処理を行なった識別器である。H-wH-EI は HLDA および WHLDA によって変換された特徴ベクトルを結合することで作成した新たな特徴ベクトルを用いて HMM の学習、および識別を行う枠組 (early integration) である。H-wH-LI は各々の識別器から得られる尤度の線形和を最終的な尤度として識別を行う枠組 (late integration) である。提案手法である H-wH-SVM は、各々の識別器から出力される尤度を特徴量として SVM に基づくパターン認識を行う枠組である。このとき、評価すべき単語の種類 (識別クラス数) は 216 であり、統合の段階で SVM は 432 次元の特徴量を扱うことになる。

図 4 に表 1 に示した評価項目に対する単語認識率を示す。early integration での認識率は 95.5%、late integration での認識率は 95.4% であり、単純な統合の結果では識別器単体での認識率 95.0% と比べてわずかに性能が向上していない。一方、提案手法では、98.7% の認識率が得られ、識別器単体での結果だけでなく他の統合手法の結果と比較しても、多くの誤りの削減していることがわかる。これは、識別器単体の結果と比較して 74%、early integration の結果と比較して 71% の誤り削減率となる。以上より、相補性のある複数の識別器から得ら

れる識別クラスの尤度情報の特徴量とした, SVM に基づくパターン認識を用いて識別器を統合することの効果が確認できた.

#### 4. ま と め

本稿では, 特徴変換にプースティングを適用した相補的な識別器の生成と, 尤度空間における SVM を用いた統合手法を提案し, 孤立単語音声認識実験によって評価を行った. 提案手法によって識別器を統合することで, 識別器統合前と比較して 74%, early integration などの単純な統合法と比較して 71% の誤り削減率が得られた. 今後は, 本手法を拡張し, 大語彙連続音声認識に適用する予定である.

#### 文 献

- [1] S. Gurbuz *et al.*, "Independent Information from Visual Features for Multimodal Speech Recognition," Southeast-Con 2001, no.5, pp.221-228, 2001.
- [2] T. Kubo *et al.*, "Manifold HLDA and its application to robust speech recognition," Proc. Interspeech, pp.1551-1554, 2006.
- [3] K. Kirchhoff "Combining Articulatory and Acoustic Information for Speech Recognition in Noisy and Reverberant Environments," Proc. ICSLP, pp.891-894, 1998.
- [4] R. E. Schapire *et al.*, "Improved Boosting Algorithms Using Confidence-rated Predictions," Kluwer Academic Publishers, pp.297-336, 1999.
- [5] R. Zhang *et al.*, "Investigations of Issues for Using Multiple Acoustic Models to Improve Continuous Speech Recognition," Proc. Interspeech, pp.529-532, 2006.
- [6] J. G. Fiscus, "A Post-processing system to yield reduced word error rates: Recognizer output voting error reduction (ROVER)," Proc. ASRU, pp.347-354, 1997.
- [7] L.Mangu *et al.*, "Finding Consensus in Speech Recognition: Word Error Minimization and Other Applications of Confusion Networks," Computer Speech and Language, vol.14, no.4, pp.373-400, 2000.
- [8] N. Kumar, "Investigation of silicon-auditory models and generation of linear discriminant analysis for improved speech recognition," Ph.D. dissertation, Johns Hopkins Univ., Baltimore, MD, 1997.
- [9] K.Itou *et al.*, "The design of the newspaper-based japanese large vocabulary continuous speech recognition corpus," Proc. ICSLP, pp.722-725, 1998.