

マルチカーネル学習を用いた話者認識における最適化の検討

小川 哲 司^{†1} 日野 英 逸^{†2} レイハニ ニマ^{†3}
村田 昇^{†2} 小林 哲 則^{†4}

本稿では、マルチカーネル学習を話者認識システムに適用した場合における、最適化アルゴリズムと認識性能の関係について調査を行った。話者認識システムにカーネル法を適用する場合、与えられたデータに対して適切なカーネル関数やパラメータを決定する必要がある。マルチカーネル学習は、複数のカーネル関数を凸結合することで、カーネル関数やパラメータを厳密に決定する必要性を減じることができる。本稿では、このマルチカーネル学習で用いる最適化基準および最適化アルゴリズムに焦点を当て、条件付きエントロピー最小化に基づくアルゴリズムと従来多く用いられているマージン最大化に基づくアルゴリズムを話者認識において比較した。その結果、条件付きエントロピー最小化に基づくシステムは、マージン最大化に基づくシステムの誤りを削減することがわかった。

Investigation on optimization in speaker recognition using multiple kernel learning

TETSUJI OGAWA,^{†1} HIDEITSU HINO,^{†2} NIMA REYHANI,^{†3}
NOBORU MURATA^{†2} and TETSUNORI KOBAYASHI^{†4}

We investigated the relation between the optimization algorithm for multiple kernel learning (MKL) and the speaker recognition performance. Most of the kernel methods applied to speaker recognition systems require a suitable kernel function and its parameters to be determined for a given data set. In contrast, MKL eliminates the need for strict determination of the kernel function and parameters by using a convex combination of element kernels. In the present paper, we focused on the optimization criterion and algorithm applied to MKL. We compared an MKL algorithm based on conditional entropy minimization (MCEM) with a conventional maximum-margin-based MKL algorithm in terms of speaker recognition accuracy; the MCEM-based system reduced the speaker error rate as compared to the maximum-margin-based system.

1. はじめに

近年、話者認識システムにおいて、サポートベクタマシン (Support Vector Machine; SVM) に代表されるカーネル法に基づく識別器が採用されている^{1),2)}。しかしながら、カーネル法に基づくシステムでは、扱うデータに応じてカーネル関数やそのパラメータを適切に選択しなければ良好な性能が得られないという問題がある。与えられたデータに応じてカーネル関数を最適に設計する手法は数多く提案されているが、複数のカーネル関数を凸結合するマルチカーネル学習 (Multiple Kernel Learning; MKL)^{3),4)} はその代表的なアプローチである。近年では、相補的なカーネルとして、derivative カーネルとパラメトリックカーネルを結合する MKL が、話者照合に適用されている⁵⁾。

多くの MKL では、線形識別関数のパラメータとカーネル関数の結合係数の双方を学習するのに、マージン最大化に基づくアルゴリズムを採用している³⁾。一方、我々は、上述のパラメータに関してクラス条件付きエントロピーを最小化する MKL アルゴリズム (Multiple kernel learning algorithm based on Conditional Entropy Minimization; MCEM) を提案してきた^{6),7)}。MCEM は、人工データを用いた実験において従来のマージン最大化に基づく MKL を上回る性能を与えた。

本稿では、情報論的な最適化に基づく MKL アルゴリズムである MCEM の有効性を話者認識において実証することを試みる。そのために、同一のカーネル関数とパラメータを用いて、MCEM に基づいて学習したシステムの性能を、従来のマージン最大化に基づいて学習したシステムの性能と比較する。本研究で検討を行う MCEM は、識別境界の構築に焦点を当てた方式ではないものの、データが各クラスごとに密集するように抽出された部分空間においてデータを識別しようとするものであり、識別平面周辺のノイズに対し頑健であると考えられる。また、MCEM のような情報論的な最適化は、収束性、頑健性、効率性、一貫性のような統計的性質の解析との親和性が高いと考えられる。

^{†1} 早稲田大学 高等研究所

Waseda Institute for Advanced Study

^{†2} 早稲田大学 先進理工学部 電気・情報生命工学科

Department of Electrical Engineering and Bioscience, Waseda University

^{†3} Information and Computer Science Lab., Aalto University

^{†4} 早稲田大学 基幹理工学部 情報理工学科

Department of Computer Science, Waseda University

本稿の構成は、以下の通りである。2では、本研究の基礎である MKL について述べる。3では、条件付きエントロピー最小化を MKL に適用することの妥当性と、MCEM アルゴリズムについて述べる。4では、話者識別実験を行い、MCEM の有効性を実験的に検証する。最後に 5 において、まとめを述べる。

2. マルチカーネル学習

本章では、本研究の基礎となる MKL の基本方式と、従来研究について概観する。本研究では、観測データ集合 $D = \{\mathbf{x}_i\}_{i=1}^N$ 、そのクラスラベル $\{y_i\}_{i=1}^N$ 、 $y_i \in \{\pm 1\}$ のもと、以下の判別関数を考える。

$$f(\mathbf{x}) = \sum_{i=1}^N \alpha_i k(\mathbf{x}, \mathbf{x}_i) \quad (1)$$

ここで、 $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$ は、特徴空間におけるデータ $\Phi(\mathbf{x}_i)$ 、 $\Phi(\mathbf{x}_j)$ の内積として定義されるカーネル関数である。このような判別関数を用いて、典型的なカーネル法に基づく学習は以下のように表される。

$$\min_{\alpha} \left[\sum_{i=1}^N \mathcal{L}(y_i, \sum_{j=1}^N \alpha_j k(\mathbf{x}_i, \mathbf{x}_j)) + \|\alpha\|_K^2 \right] \quad (2)$$

ここで、 \mathcal{L} は損失関数であり、SVM では hinge 損失 $\mathcal{L}(y, t) = \max(0, 1 - yt)$ が用いられる。カーネル関数はデータ点における値が評価できれば十分であり、以下では $[K]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ のようにカーネル行列を用いて表現する。MKL では、カーネル行列 K を S 個のカーネル行列の凸結合 $\sum_{s=1}^S \beta_s K_s$ (このとき、 $\sum_{s=1}^S \beta_s = 1$ 、 $\beta_s \geq 0$ を満たす) で置き換え、識別器のパラメータ $\alpha = \{\alpha_j\}_{j=1}^N$ とカーネル関数の結合係数 $\beta = \{\beta_s\}_{s=1}^S$ の双方に関して以下の最適化問題を解く。

$$\min_{\alpha, \beta} \left[\sum_{i=1}^N \mathcal{L}(y_i, \sum_{j=1}^N \alpha_j \sum_{s=1}^S \beta_s [K_s]_{ij}) + \|\alpha\|_K^2 \right] \quad (3)$$

従来の MKL の多くは、識別器として SVM を採用し、マージン最大化基準を用いて α と β を最適化するアプローチを採っており、情報論的な解釈は陽に与えられていない。Lanckriet らは、様々な種類の損失関数を用いた SVM に対して、複数のカーネル関数を結合する枠組みを提案し、半正定値計画問題 (Semi-Definite Programming; SDP)⁸⁾ とし

て、SVM のマージンを α と β に関して同時に最大化するという定式化を行った³⁾。しかし、この方式は計算コストが著しく高く、話者認識に直接適用するのは困難である。一方、SimpleMKL⁹⁾ のように効率的な MKL も提案されているが、その精度は上述の SDP に基づく方式と同等である。また、Do らは、SVM の汎化誤差がマージンのみではなく特徴空間においてデータを包含する最小の球のラディウス (半径) にも依存すること、およびカーネル関数の組み合わせによって特徴空間におけるデータ分布が変化するという事実に基づき、マージンとラディウスの両方を最適化する MKL の枠組みとして RMKL を提案した⁴⁾。RMKL は SimpleMKL と同等の計算コストで、SDP に基づく MKL を上回る性能を与えることが示されている。したがって本研究では、MCEM と RMKL のみ比較を行う。

3. 条件付きエントロピー最小化に基づくマルチカーネル学習

本章では、条件付きエントロピー最小化基準に基づく MKL である MCEM⁶⁾ について、この基準を MKL に導入することの妥当性と、最適化アルゴリズムの概要を述べる。

3.1 条件付きエントロピー最小化基準の妥当性

$f: \mathbf{x} \mapsto z$ 、 $\mathbf{x} \in \mathbb{R}^n$ 、 $z \in \mathbb{R}$ なる識別関数 f は、1 次元の判別軸上への射影関数とみなせることから、(特に二値判別のような) 識別問題は教師あり次元圧縮問題と解釈できる⁶⁾。教師あり次元圧縮では、次元圧縮後のデータは各クラスごとに密集していることが望ましい。情報論的な観点から捉えなおせば、各クラスのデータが密集していることは、クラス条件付きエントロピーが小さいことに相当する¹⁰⁾。以上より、識別関数 f を教師ありの枠組みで学習する場合、条件付きエントロピー最小化基準を目的関数とすることは妥当と言える。

3.2 MCEM アルゴリズム

MCEM は、上述の議論から、クラス条件付きエントロピーを α と β に関して最小化する問題として、以下のように定式化できる。

$$\min_{\alpha, \beta} \left[H(f(\mathbf{X}; \alpha, \beta) | Y) \right] \quad (4)$$

$$\text{s. t. } H(f(\mathbf{X}; \alpha, \beta)) = \text{const.}, \quad \sum_{s=1}^S \beta_s = 1, \quad \beta_s \geq 0$$

ここで、 $f(\mathbf{x}; \alpha, \beta)$ は識別関数であり、 α と β の依存性を陽に示している。このとき、自明な解や過学習を防ぐため、 $H(f(\mathbf{X}; \alpha, \beta) | Y)$ の最小化に対する正則化項として $H(f(\mathbf{X}; \alpha, \beta))$ を用いる。MCEM では、MKL の最適化において、カーネルフィッシャー判別分析 (Kernel

Fisher Discriminant Analysis; KFDA)¹¹⁾ を採用している。

3.2.1 KFDA とその MCEM への適用

KFDA は、教師ありの枠組みで観測データの非線形な射影関数を見つける問題である。観測データ $\mathbf{x} \in \mathbb{R}^n$ は、 $\Phi: \mathbb{R}^n \rightarrow \mathbb{R}^{n'}$ によって n' 次元の特徴空間 $\mathbb{R}^{n'}$ に写像されるとする。このとき、射影関数として特徴ベクトルの線型和で定義される関数 $f(\mathbf{x}) = \mathbf{a}^T \Phi(\mathbf{x})$, $\mathbf{a} \in \mathbb{R}^{n'}$ を考える。カーネル法では、判別のコストに正則化項 $\lambda \|\mathbf{a}\|^2$ を加えることでリプレゼンター定理が成立し、上式における係数ベクトル \mathbf{a} は実数の重みパラメータ $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$ を用いて $\mathbf{a} = \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_i)$ と書ける。このとき、射影関数 $f(\mathbf{x})$ は式 (1) の形式で表される。与えられたデータセット $D = \{\mathbf{x}_i\}_{i=1}^N$ のカーネル行列を $K \in \mathbb{R}^{N \times N}$, \mathbf{k}_i を K の i 番目の列ベクトルとすると、各クラスに属するデータに対する $f(\mathbf{x})$ の平均ベクトル \bar{k}^y , 全データに対する $f(\mathbf{x})$ の平均ベクトル \bar{k} は、 $\bar{k}^y = \frac{1}{N_y} \sum_{i \in D_y} \mathbf{k}_i$, $\bar{k} = \frac{1}{N} \sum_{i \in D} \mathbf{k}_i$ のように計算できる。ここで、 D_y はクラス y に属するデータの集合、 $N_y = |D_y|$ はクラス y に属するデータの数、 $N = \sum_y N_y$ は全データ数である。このとき、特徴空間におけるクラス内分散行列 \mathbf{V}_w , クラス間分散行列 \mathbf{V}_b は、以下のように計算される。

$$\mathbf{V}_w = \frac{1}{N} \sum_{y \in \{\pm 1\}} \sum_{i \in D_y} (\mathbf{k}_i - \bar{k}^y)(\mathbf{k}_i - \bar{k}^y)^T \quad (5)$$

$$\mathbf{V}_b = \frac{1}{N} \sum_{y \in \{\pm 1\}} N_y (\bar{k}^y - \bar{k})(\bar{k}^y - \bar{k})^T \quad (6)$$

KFDA では、同じクラスに属するデータの散らばり ($\alpha^T \mathbf{V}_w \alpha$) を小さく、クラスどうしのデータの散らばり ($\alpha^T \mathbf{V}_b \alpha$) を大きくするように次元圧縮を行う。したがって、KFDA は、 $\alpha^T \mathbf{V}_b \alpha$ が定数という制約条件下において $\alpha^T \mathbf{V}_w \alpha$ を最小化する問題に帰着される。つまり、KFDA は以下のように定式化できる。

$$\min_{\alpha} \left[\alpha^T (\mathbf{V}_w + \zeta K) \alpha \right] \quad \text{s. t.} \quad \alpha^T \mathbf{V}_b \alpha = \text{const.} \quad (7)$$

このとき、過学習を防ぐために、 \mathbf{V}_w は $\mathbf{V}_w + \zeta K$, $\zeta \geq 0$ のように正則化される。

次に、KFDA を MCEM に適用することの妥当性について述べる。クラス条件付きエントロピー $H(f(\mathbf{X})|Y)$ と KFDA の目的関数について、以下の不等式が成立する⁶⁾。

$$H(f(\mathbf{X})|Y) \leq \log(2\pi)^{\frac{1}{2}} e + \frac{1}{2} \log(\alpha^T \mathbf{V}_w \alpha) \quad (8)$$

ここで、KFDA の目的関数は、式 (8) の右辺と本質的に同一である。したがって、KFDA

は、次元圧縮後の値を用いて計算したクラス条件付きエントロピーの上界の最小化とみなすことができる。これは、MCEM がクラス条件付きエントロピー $H(f(\mathbf{X})|Y)$ を最小化する射影 $f: \mathbf{x} \mapsto z$ を教師ありの枠組みで推定する、教師あり次元圧縮の問題として定式化できることを意味する。

3.2.2 α と β の最適化

MCEM アルゴリズムを図 1 にまとめる。式 (4) で表される最適化問題を α と β に対して同時に解くことは困難であるため、 α と β に関して以下の繰り返し最適化を行う。

- (1) β を固定した上で、上界による近似と KFDA により α に関する最適化を行う。
- (2) 前ステップで推定された α (固定) を用いて β に関する最適化を行う。

この繰り返しアルゴリズムの挙動に関する概念図を図 2 に示す。このとき、 t 回目の繰り返し後に得られる α と β の推定値を $\alpha^{(t)}$, $\beta^{(t)}$ と書く。

α の最適化に関しては、KFDA を適用した。式 (8) で示した通り、KFDA はクラス条件付きエントロピーの上界を最小化する。したがって、 β を固定したときの最適な α は、KFDA により式 (9) に基づいて求めることができる。

β の最適化については、ランダムサーチに基づく方法を採用した⁶⁾。このとき、 $\eta > 0$ を導入し、正則化項 $H(f(\mathbf{X}; \alpha, \beta))$ を用いて、式 (10) で表される最小化を行う。このアルゴリズムでは、 $\beta^{(t-1)}$ を平均ベクトル、単位行列を共分散行列とするガウス分布から P 個の β 候補をサンプリングする。得られた $\{\beta_p\}_{p=1}^P$ を用いて式 (10) の目的関数を計算し、その最小値を与える β を最適化の結果として採用する。この手法は極めて単純であるが、予備実験において良好な結果を与えたため、本実験においてもこのアルゴリズムを採用した。本研究では、エントロピーおよびクラス条件付きエントロピーは、Faivishevsky らが提案した Mean Nearest Neighbor (MNN) により推定した Shannon エントロピーを意味する¹²⁾。

この β に関する最適化の結果、更新された結合係数 β を用いた新たなカーネル行列を得る。この行列を用いて共分散行列 $\mathbf{V}_w(\beta)$ と $\mathbf{V}_b(\beta)$ を計算する。その後、更新された KFDA の目的関数を α に関して最小化する。これら 2 つのステップを、 α と β が収束するか、条件付きエントロピーの値が収束するまで繰り返す。

4. 話者認識実験

マルチカーネル学習における最適化アルゴリズムと認識性能の関係を調査するために、テキスト独立型話者識別実験を行った。本稿では、以下の 2 つの実験を行った。

実験 1 識別器の持つ本質的な識別能力を調査するために、男性話者 20 名の各組み合わせ

MCEM: Multiple kernel learning algorithm based on conditional entropy minimization

Input: A set of kernel functions $\{k_s(\mathbf{x}_i, \mathbf{x})\}_{s=1}^S$, training data $D = \{\mathbf{x}_i\}_{i=1}^N$, $\mathbf{x}_i \in \mathbb{R}^n$, class labels $\{y_i\}_{i=1}^N$, $y_i \in \{\pm 1\}$, and a regularization parameter ζ for KFDA.

Initialization: Calculate kernel matrices $\{K_s\}_{s=1}^S$ using D . Initialize the combination coefficients $\beta^{(0)} = \{\beta_s^{(0)}\}_{s=1}^S$ of element kernels by random values such that $\sum_{s=1}^S \beta_s^{(0)} = 1$ and $\beta_s^{(0)} \geq 0$.

Repetition: Until convergence, from $t = 1$:

α optimization step: Solve a KFDA minimization problem for fixed $\beta^{(t-1)}$ to obtain $\alpha^{(t)}$:

$$\min_{\alpha} \left[\alpha^T (\mathbf{V}_w(\beta^{(t-1)}) + \zeta K) \alpha \right] \quad \text{s. t.} \quad \alpha^T \mathbf{V}_b \alpha = \text{const.} \quad (9)$$

β optimization step: Minimize the conditional entropy of the classification function $f(\mathbf{x}; \alpha^{(t)}, \beta)$ for fixed $\alpha^{(t)}$ to obtain $\beta^{(t)}$:

$$\min_{\beta} \left[H(f(\mathbf{X}; \alpha^{(t)}, \beta) | Y) - \eta H(f(\mathbf{X}; \alpha^{(t)}, \beta)) \right] \quad \text{s. t.} \quad \sum_{s=1}^S \beta_s = 1, \quad \beta_s \geq 0. \quad (10)$$

Output: Converged parameter α and β , used to construct the classification function as follows:

$$f(\mathbf{x}; \alpha, \beta) = \sum_{i=1}^N \alpha_i \sum_{s=1}^S \beta_s k_s(\mathbf{x}_i, \mathbf{x}). \quad (11)$$

図1 クラス条件付きエントロピーを最小化するための繰り返しアルゴリズム
Fig.1 An iterative algorithm for minimizing class-conditional entropy.

に対して、一対比較を行った。このとき、組み合わせの総数は190組である。

実験2 男性話者20名の識別を行った。

4.1 実験条件

4.1.1 評価項目

表2に比較対象となる識別器を示す。本実験では、マルチカーネルを用いたSVMを、

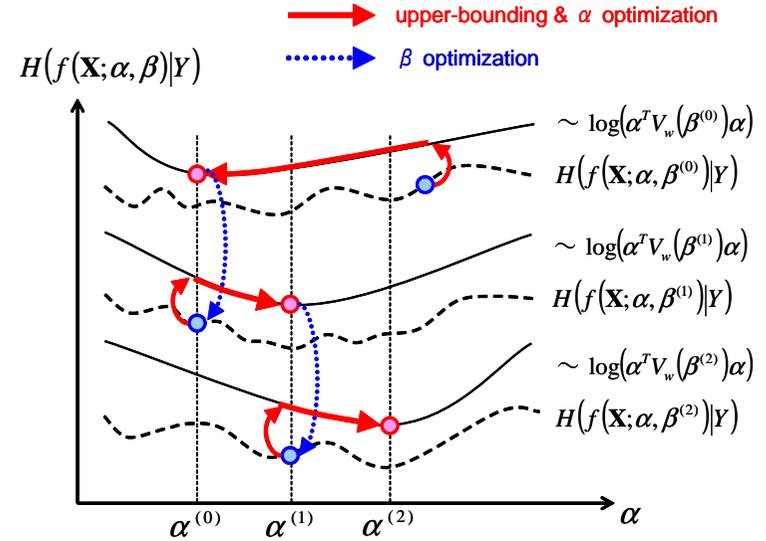


図2 MCEM アルゴリズムにおける繰り返し最適化の概念図。点線は条件付きエントロピー、実線は条件付きエントロピーの上界 (KFDA の目的関数) を表す。

Fig.2 Conceptual image of iterative optimization in MCEM algorithm. Dotted lines express the class-conditional entropy, and solid lines express the upper bounds of the class-conditional entropy (i.e., the objective function in KFDA).

RMKL⁴⁾, MCEM⁶⁾ といった MKL のアルゴリズムを用いて構築した。このとき、結合要素となるカーネル関数は RBF カーネル $\exp(-\sigma \|\mathbf{x} - \mathbf{x}'\|^2)$ 3 種類 ($\sigma = 0.01, 0.05, 0.1$)、多項式カーネル $(\mathbf{x}^T \mathbf{x}' + 1)^d$ 2 種類 ($d = 2, 3$)、線形カーネル $\mathbf{x}^T \mathbf{x}'$ の計 6 種類である。本実験では、結合要素となるカーネル関数のうち 1 種類のみ (シングルカーネル) を用いて構築した SVM、および GMM についても評価を行った。

4.1.2 音声データ

本実験では、音素バランス単語 4320 発話を用いた。これは、20 人の男性話者各々が 216 種類の孤立単語音声を発話したものを、近接マイクロホンを通して収録することで得た。音声は 16 kHz でサンプリング、16 ビットで量子化されている。

各話者ごとに 108 単語を用いて話者モデルを学習し、残りの 108 単語を用いて評価を行った。このような評価を組み合わせを変えて行い (2-fold cross validation)、その性能の平均値により比較を行った。

表 1 特徴抽出に関する実験条件

Table 1 Experimental conditions for feature extraction.

sampling frequency	16 kHz
frame length	25 ms
frame shift	10 ms
analysis window	Hamming window
pre-emphasis	$1 - 0.97z^{-1}$

4.1.3 音響特徴抽出

音響特徴量として、MFCC 12 次元、 Δ MFCC 12 次元の計 24 次元のパラメータを用いた。音響特徴抽出の条件は表 1 に示した通りである。

4.1.4 SVM に基づく話者認識システム

本研究では、異なるフレーム数を持つ音声を扱うために、カーネル関数として系列カーネル¹³⁾を用いた。いま、 $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_T) \in \mathbb{R}^{n \times T}$ および $\mathbf{x}' = (\mathbf{x}'_1, \dots, \mathbf{x}'_{T'}) \in \mathbb{R}^{n \times T'}$ ($n=24$) の 2 発話に対して、系列カーネルは、以下のように定義される。

$$\mathcal{K}(\mathbf{x}, \mathbf{x}') = \frac{1}{T \cdot T'} \sum_{t=1}^T \sum_{t'=1}^{T'} k(\mathbf{x}_t, \mathbf{x}'_{t'}). \quad (12)$$

本実験で用いたカーネル関数 $k(\mathbf{x}_t, \mathbf{x}'_{t'})$ は表 2 で示した 6 種類である。式 (9) の正則化パラメータ ζ は、全ての実験において $\zeta = 0.001$ とした。ソフトマージン SVM に基づく RMKL では、ソフトマージンパラメータを 1.0 とした。MCEM では、トレードオフパラメータとして $\eta = 0.5$ を用い、 β の最適化処理においては、500 個のサンプルを生成した。これらのパラメータは、予備実験において最良の性能を与えた値を用いた。 α と β の最適化における繰り返し処理は、条件付きエントロピーの減少率が 0.0001 を下回った場合、もしくは条件付きエントロピーが上昇に転じた場合に打ち切った。

変換されたデータを用いて識別を行う方法は様々なものが考えられるが、本実験では one-versus-one SVM (ある 1 話者と他の 1 話者を識別する SVM) を用いた。このときの識別関数は、式 (11) で表される 1 次元の判別軸上への射影関数である。本実験では男性話者 20 名のデータを用いているので、このとき構築された識別器の総数は 190 である。実験 2 では、winner-takes-all strategy を用いて、これら one-versus-one SVM の結果を統合した。

4.2 実験結果

表 2 に、本実験における評価項目となる識別器と、その話者識別誤り率を示す。これよ

表 2 GMM, シングルカーネルを用いた SVM, マルチカーネルを用いた SVM の話者識別誤り率 (%)。実験 1 は 20 話者の一対比較に対する平均誤り率, 実験 2 は 20 話者の識別誤り率を表す。

Table 2 Speaker error rates (%) for GMM, SVMs with a single kernel, and SVMs with multiple kernels. In Exp. 1, each pair of 20 speakers was classified. In Exp. 2, 20 speakers were classified.

System			Exp. 1	Exp. 2
GMM (UBM-MAP)			3.0	4.2
SVM (single kernel)	RBF	$k(\mathbf{x}, \mathbf{x}') = \exp(-0.01 \cdot \ \mathbf{x} - \mathbf{x}'\ ^2)$	1.0	0.046
	RBF	$k(\mathbf{x}, \mathbf{x}') = \exp(-0.05 \cdot \ \mathbf{x} - \mathbf{x}'\ ^2)$	12.0	12.6
	RBF	$k(\mathbf{x}, \mathbf{x}') = \exp(-0.1 \cdot \ \mathbf{x} - \mathbf{x}'\ ^2)$	27.1	4.4
	Polynomial	$k(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + 1)^2$	2.5	0
	Polynomial	$k(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + 1)^3$	1.9	0
	Linear	$k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'$	5.7	0
SVM (multiple kernel)	RMKL		1.6	0
	MCEM		1.4	0

り、条件付きエントロピー最小化に基づく MKL (MCEM) は高い話者識別性能を与えた。実験 1 においては、MCEM に基づくシステムの誤り率は 1.4% であり、マージンとラディウスの最適化に基づく MKL (RMKL) を用いたシステムの誤り (1.6%) を 12.5% 削減した。また、MCEM に基づくシステムは、GMM に基づくシステムの誤り (3.0%) を 53.3% 削減した。実験 2 においては、GMM に基づくシステムでは 4.2% の誤りが生じたが、MCEM および RMKL に基づくシステムでは誤りは生じなかった。

また、MKL に基づくシステムは、シングルカーネルを用いた SVM に基づくシステムと比較して、同等もしくはそれ以上の性能を与えた。実験 1 では、RBF カーネル ($\sigma=0.01$) を用いた SVM に基づくシステムは MKL に基づくシステムの性能を上回った。しかしながら、シングルカーネルを用いた SVM では、選択されたカーネル関数やそのパラメータに応じて話者識別性能が著しく異なる (例えば、最良の誤り率は 1.0% であり、最悪の誤り率は 27.1% である) ことから、扱うデータに応じてそれらの厳密なチューニングが必要であることがわかる。一方、MKL に基づくシステムは、カーネル関数やそのパラメータの厳密な決定を必要とせず、高い話者識別性能を達成した。これは、異なるカーネル関数やパラメータを持つ SVM は誤り傾向も異なっており、そのような相補的な識別器を統合することによって、シングルカーネル SVM に基づくシステムの誤りを削減することに成功したものと考えられる。実験 2 では、マルチカーネル学習に基づくシステムと幾つかのシングルカーネル SVM に基づくシステムでは、誤りが生じなかった。この場合、winner-takes-all strategy による識別器の統合が、誤り訂正効果に寄与したものと考えられる。

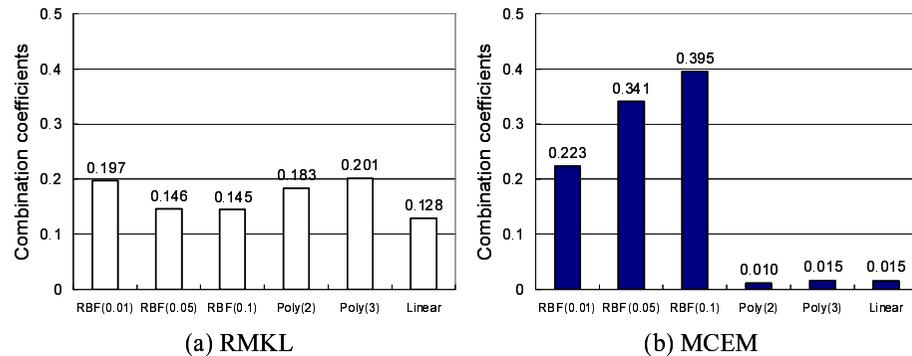


図3 (a) RMKL と (b) MCEM によって学習した結合係数 β の平均値。結合要素となるカーネルは、RBF カーネル ($\sigma = 0.01, 0.05, 0.1$)、多項式カーネル ($d = 2, 3$)、線型カーネルである。
Fig. 3 Mean values of learned combination coefficients for element kernels such as RBF kernels with $\sigma = 0.01, 0.05$, and 0.1 , polynomial kernels with $d = 2$ and 3 , and a linear kernel in the R-MKL-based system (a) and MCEM-based system (b).

図3は、表2に示された6種類のカーネルに対する結合係数の推定値 $\hat{\beta}$ の平均を表す。MCEMに基づくシステムとRMKLに基づくシステムは同等の性能を与えるが、両者は $\hat{\beta}$ に関して異なる傾向を示した。MCEMに基づくシステムは、多項式カーネルと線型カーネルに対して $\hat{\beta}$ がゼロに近い値となっているが、RMKLに基づくシステムでは、そのような明確な傾向は見られない。

このとき、モデル選択や特徴選択の観点からは、 β の推定値のスパース性、つまり、 $\hat{\beta}$ のうち幾つかがゼロ（もしくは近似的にゼロ）という性質は重要である。図3の結果は、モデル選択や特徴選択に対して、MCEMはRMKLと比較してより有効であることを示唆している。

以上より、MKLにおける最適化アルゴリズムとして、MCEMが話者識別に対して有効であることが示された。

5. まとめ

MKLを用いた話者認識システムにおける最適化アルゴリズムに着目し、条件付きエントロピー最小化に基づくアルゴリズムであるMCEMと、従来のマージン最大化に基づくア

ルゴリズムの性能を調査した。MCEMは、最適なカーネル関数とそのパラメータの厳密な決定を必要とせず、情報論的な次元圧縮によって得られた識別的な部分空間においてデータを識別するものであり、GMMに基づくシステム、およびマージンやラディウスの最適化に基づくMKL(RMKL)に基づくシステムと比較して、各々53%、13%の誤りを削減した。
謝辞 本研究の一部は文部科学省の科研費(22800067)の助成を受けたものである。

参考文献

- 1) W. M. Campbell et al., "Support vector machines using GMM supervectors for speaker verification," IEEE Sign. Process. Lett., vol.13, no.5, pp.308-311, May 2006.
- 2) A. Stolcke et al., "MLLR transforms as features in speaker recognition," Proc. INTERSPEECH, pp.2425-2428, Sept. 2005.
- 3) G. R. G. Lanckriet et al., "Learning the kernel matrix with semidefinite programming," JMLR, vol.5, pp.27-72, 2004.
- 4) H. Do et al., "Margin and radius based multiple kernel learning," Proc. ECML, pp.330-343, Sept. 2009.
- 5) C. Longworth et al., "Multiple kernel learning for speaker verification," Proc. ICASSP, pp.1581-1584, March 2008.
- 6) H. Hino and N. Murata, "Conditional entropy minimization criterion for dimensionality reduction and multiple kernel learning," Neural Computation, vol.22, no.11, pp.2887-2923, 2010.
- 7) H. Hino et al., "Multiple kernel learning by conditional entropy minimization," Proc. ICMLA, Dec. 2010.
- 8) Y. Nesterov and A. Nemirovskii, "Interior point polynomial methods in convex programming: Theory and applications," Society for Industrial and Applied Mathematics, 1994.
- 9) A. Rakotomamonjy et al., "SimpleMKL," JMLR, vol.9, pp.2491-2521, 2008.
- 10) T. M. Cover and J. A. Thomas, "Elements of information theory," John Wiley and Sons, Inc., 1991.
- 11) S. Mika et al., "Fisher discriminant analysis with kernels," Proc. NNSP, pp.41-48, Aug. 1999.
- 12) L. Faivishevsky and J. Goldberger, "ICA based on a smooth estimation of the differential entropy," Proc. NIPS, pp.433-440, Dec. 2009.
- 13) J. Mariethoz and S. Bengio, "A kernel trick for sequences applied to text-independent speaker verification systems," Pattern Recognition, vol.40, no.8, pp.2315-2324, 2007.