

MKL-SVM を用いた自閉症スペクトラム障害児と定型発達児の音響識別

☆柿原康博, 滝口哲也, 有木康雄 (神戸大), △中井靖 (川崎医療短期大), △高田哲 (神戸大) *

1 はじめに

近年, 自閉症スペクトラム障害の発生頻度の増加が注目されている。自閉症とは, 社会性や他者とのコミュニケーションに困難が生じる発達障害の一種である。また自閉症スペクトラム障害とは, 自閉性障害, アスペルガー障害, 特定不能の広汎性発達障害の総体である。これらの障害は多様な原因に基づいて発症するため根本的な治療は困難とされている。自閉症スペクトラム障害に特化した支援として, 絵カード交換式コミュニケーションシステム (PECS)[1], 応用行動分析 (ABA)[2], ソーシャルスキルトレーニング (SST)[3] が挙げられる。これらの支援は自閉症スペクトラム障害の特性に依拠しており, 早期発見・早期療育による効果が報告されている [4]。

本研究では, 自閉症スペクトラム障害の早期発見を音響的な側面から目指し, MKL-SVM を用いて自閉症スペクトラム障害児と定型発達児の音響識別を行った。Multiple Kernel Learning(MKL) [5] は, 複数のサブカーネルを識別器とみなし, それらを統合することで通常の Support Vector Machine(SVM) の識別能力を向上させることを目的として, 音声の話者認識 [6, 7] や画像認識 [8, 9] の分野等で用いられてきた。近年, 画像の一般物体認識の分野等では, MKL-SVM を用いて特徴選択を行う手法が提案されている [10, 11]。この手法では, 複数の特徴量を用いた画像識別において, サブカーネルを特徴量ごとに定義することで, 識別に適した特徴重みを MKL により学習させている。本研究では, 音響特徴の次元毎にサブカーネルを定義し, MKL により重みを学習させる。これを SVM に適用して特徴統合による音声識別を行った。実験では通常の SVM での識別実験と MKL-SVM による識別実験を行うことで, MKL による特徴選択と重み付けの効果を調べた。また各特徴量次元に対する重みに着目し, 自閉症スペクトラム障害児と定型発達児の識別における各特徴量次元の有効性を評価した。484 次元すべての特徴量を用いて MKL-SVM を行うと, 31 の特徴量が選択され, 精度は 80.2% となった。選択された特徴は, 韻律的特徴よりも線スペクトル周波数 (lspFreq) や MFCC といった音韻的特徴が多く, 自閉症スペクトラム障害児と定型発達児の音響識別を

する上で, 音韻的特徴が重要であると分かった。比較実験として, 484 次元すべての特徴量を用いて通常の SVM を行うと, 精度は 70.7% となった。また MKL で選択された 31 の特徴のみを用いて通常の SVM を行うと, 精度は 79.7% となった。

2 MKL-SVM による音響特徴の重み学習及び識別

本章では, MKL-SVM を用いて, 音響特徴の次元毎に重みを学習し, 自閉症スペクトラム児と定型発達児の音響識別する方法について述べる。音声データから抽出された音響特徴量には, 識別に有効な特徴と有効ではない特徴が含まれていると考えられる。そこで本研究では, MKL により音響特徴の重みを学習し特徴選択を行う。Fig. 1 は, 特徴抽出から識別までの全体像を示す。

MKL は, 複数のサブカーネルを線形結合して新たなカーネルを作成することで, より複雑な非線形空間を作成する手法である。これを用いて, サンプル i, j の音響特徴量に $\mathbf{X}_i, \mathbf{X}_j$ より計算されるカーネル関数は, 以下のように表現される。

$$k(\mathbf{X}_i, \mathbf{X}_j) = \sum_l \beta_l k_l(\mathbf{X}_i, \mathbf{X}_j) \quad (1)$$

β_l は l 番目のサブカーネル k_l の重みである。

MKL-SVM は, 本来それぞれのサブカーネルを識別器とみなし, それらを統合することで通常の SVM の識別能力を向上させることを目的としている。近年, 画像の一般物体認識の分野などで提案されている MKL-SVM を用いた特徴選択の手法 [10, 11] は, 複数の特徴量を用いた画像識別において, サブカーネルを特徴量ごとに定義することで, 識別に適した特徴重みを MKL により学習させている。本研究では, 音響特徴の次元毎にサブカーネルを定義し, MKL により重みを学習させる。

$$k(\mathbf{X}_i, \mathbf{X}_j) = \sum_d \beta_d k_d(X_i(d), X_j(d)) \quad (2)$$

従来の SVM における単一カーネルと, 通常の MKL-SVM におけるマルチカーネル, そして本手法における次元重み学習のためのマルチカーネルを Fig. 2 に示す。従来の単一カーネル SVM(上段) ではある特

* Acoustic Feature Selection Utilizing Multiple Kernel Learning for Classification of Children with Autism and Typically Developing Children. by Yasuhiro Kakiyama, Tetsuya Takiguchi, Yasuo Ariki (Kobe univ.), Yasushi Nakai (Kawasaki College of Allied Health Professions), Satoshi Takada (Kobe univ.)

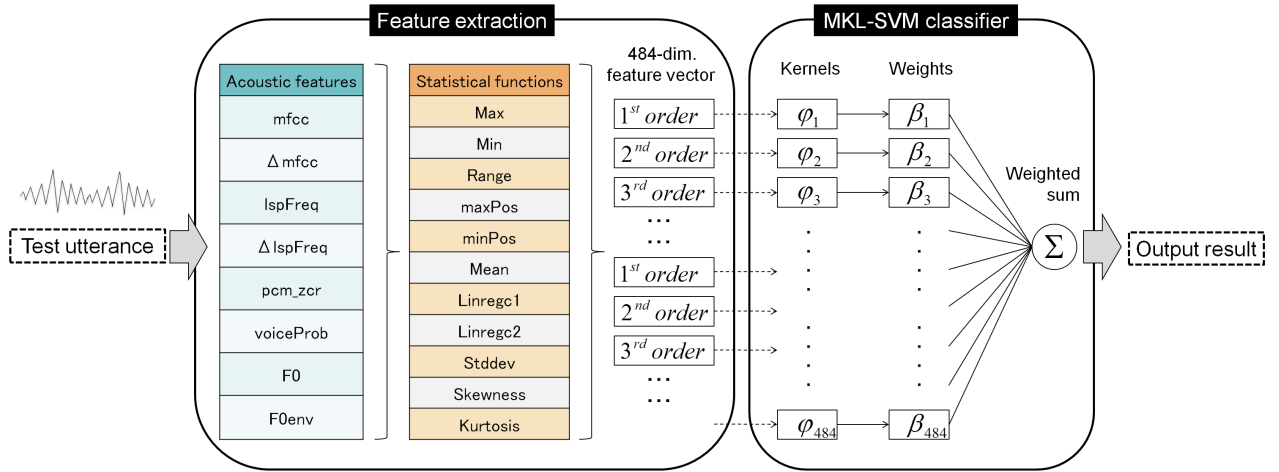


Fig. 1 System outline for MKL-SVM

微量ベクトルに対して一つのカーネル写像関数が適用されているのに対し、MKL-SVM（中段）においては特徴量ベクトルに対して複数のカーネル関数を適用し、その重み付き和によって新たなカーネル関数を定義している。そして提案法におけるマルチカーネル（下段）では、特徴量ベクトルの1次元に対して1つのカーネル関数を適用させ、それらの重み付き和によってカーネル関数を定義している。

MKLの重み β_d の学習は、SVMの枠組み、すなわちマージン最大化の枠組みで解かれるのが一般的である [5]。

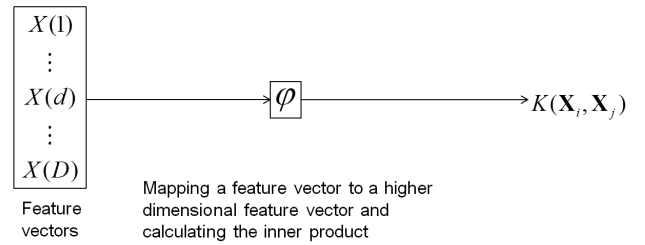
$$\begin{aligned} \min_{\beta, w_d, b, \xi} \quad & \frac{1}{2} \sum_d \frac{1}{\beta_d} \|w_d\|_2^2 + C \sum_i \xi_i \\ \text{s.t.} \quad & \begin{cases} y_i \left(\sum_d w_d^T \phi_d(X_i(d)) + b \right) \geq 1 - \xi_i \quad \forall i \\ \xi_i \geq 0 \quad \forall i, \quad \beta_d \geq 0 \quad \forall d, \quad \sum_d \beta_d = 1 \end{cases} \end{aligned} \quad (3)$$

ここで、 ϕ は高次元空間への写像関数を表し、 y_i はクラスを表す変数 $(-1, 1)$ 、 ξ はスラック変数、 C はマージンと学習データの誤り率とのトレードオフを決定する変数である。(3)式に対する双対問題は、以下のように導出される。

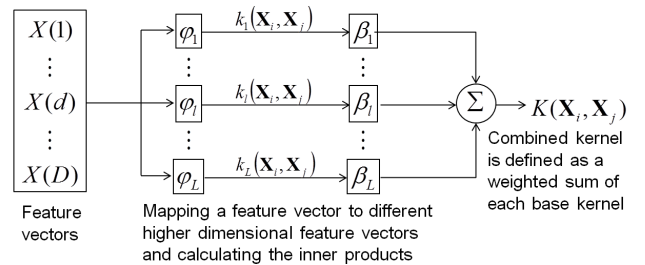
$$\begin{aligned} \max_{\alpha, \beta} \quad & \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \sum_d \beta_d k_d(X_i(d), X_j(d)) \\ \text{s.t.} \quad & \begin{cases} \sum_i y_i \alpha_i = 0, & 0 \leq \alpha_i \leq C \\ \sum_d \beta_d = 1, & \beta_d \geq 0 \end{cases} \end{aligned} \quad (4)$$

α_i はラグランジュ係数である。(4)式を満たす α_i 、 β_d は2ステップの反復による解法を用いて求められる。まず第1ステップでは β_d を固定して α_i を通常のSVMの解法により更新する。そして第2ステップでは α_i を固定して β_d を更新する。本手法では、 α_i の更新には SVM^{light} [12]を用い、 β_d の更新にはprojected-gradient[13]を用いた。これらのステップを繰り返すことにより、特徴次元の重みと識別境界が同時に学習される。

• Conventional single-kernel SVM



• Original Multiple kernel learning



• Multiple kernels for feature weighting

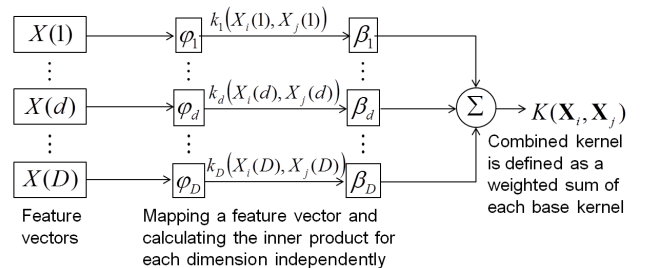


Fig. 2 A conventional SVM with single-kernel, original MKL for SVMs and a new weighting method based on MKL

Acoustic features		dim.	Statistical functions	
mfcc	Mel-Frequency cepstral coefficients.	12	Max	The maximum value of the contour.
Δ mfcc	Delta Mel-Frequency cepstral coefficients.	12	Min	The minimum value of the contour.
lspFreq	Line spectral pair frequencies.	8	Range	max - min.
Δ lspFreq	Delta Line spectral pair frequencies.	8	maxPos	The absolute position of the maximum value.
pcm_zcr	Zero-crossing rate of time signal (frame-based).	1	minPos	The absolute position of the minimum value.
voiceProb	The voicing probability computed from the ACF.	1	Mean	The arithmetic mean of the contour.
F0	The fundamental frequency computed from the Cepstrum.	1	Linregc1	The slope (m) of a linear approximation of the contour.
F0env	The envelope of the smoothed fundamental frequency contour.	1	Linregc2	The offset (t) of a linear approximation of the contour.
			Stddev	The standard deviation of the values in the contour.
			Skewness	The skewness (3rd order moment).
			Kurtosis	The kurtosis (4th order moment).

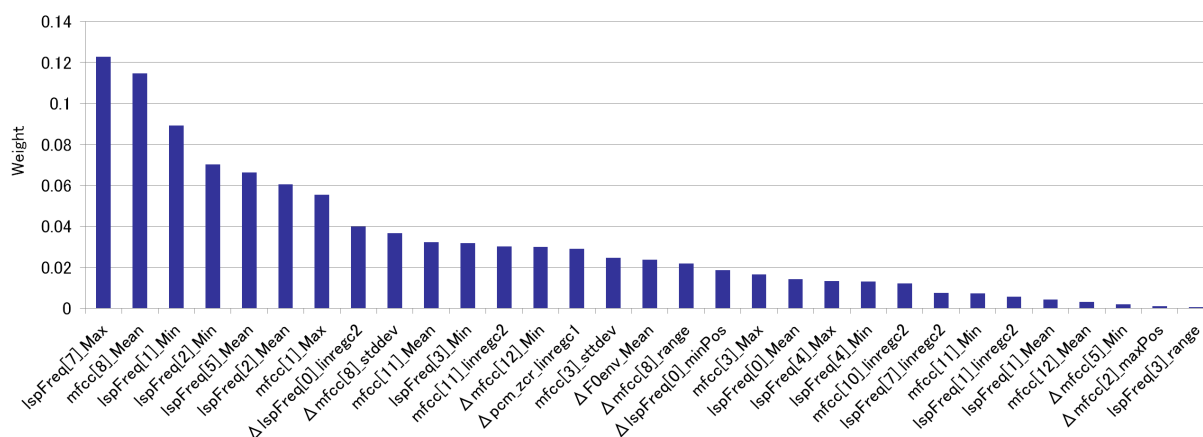


Fig. 3 Upper: The 44 dimensional acoustic features and the 11 statistical functions. Lower: Weight of features estimated by using MKL-SVM

3 識別実験

3.1 特徴抽出

本実験では、感情音声認識の分野で用いられる openEAR によって音声データから特徴量を抽出する。抽出される音響特徴量は、音韻的特徴・韻律的特徴・音の明瞭性といった 44 次元の特徴と、11 の統計量の組み合わせ (Fig. 3 上段) から成り、1 発話につき 484 次元の特徴量を抽出する。

3.2 実験条件

自閉症スペクトラム障害児、定型発達児のそれぞれの音声データから音響特徴量を抽出し、SVM と MKL-SVM を用いて識別実験を行った。音声データは自閉症スペクトラム障害児 20 名の 1560 単語、定型発達児 21 名の 1508 単語である。識別実験はデータを 3 つに分割し、1 つをテストデータ、残る 2 つを学習データとし組み合わせ変更して、3-fold cross-validation を適用した。

3.3 実験結果

Fig. 3 下段は、484 次元の音響特徴量のうち MKL-SVM によって選択・学習された 31 次元に対する重

Table 1 Classification results

	SVM	SVM*	MKL-SVM
Number of dims.	484	31*	484→ 31
Feature selection	—	○	○
Weighting	—	—	○
Accuracy[%]	70.7	79.7*	80.2

みである。選択された特徴は、韻律的特徴よりも線スペクトル周波数 (lspFreq) や MFCC といった音韻的特徴が多く、自閉症スペクトラム障害児と定型発達児の音響識別をうえて、音韻的特徴が重要であると分かる。Table 1 は、通常の SVM と MKL-SVM を用いたときの識別結果である。484 次元すべての特徴量を用いて MKL-SVM を行うと、31 の特徴量が選択され、精度は 80.2% となった。同様に 484 次元すべての特徴量を用いて通常の SVM を行うと、精度は 70.7% となった。このとき精度の差は 9.5% となり、自閉症スペクトラム障害児と定型発達児の音響識別において、MKL-SVM が有効であることが示された。Table 1 中の SVM* は、MKL で選択された 31 の特徴

量のみを用いて通常のSVMを行った結果であり、精度は79.7%となった。この識別結果と、MKL-SVMでの識別結果を比較すると、精度に0.5%の差がある。これはMKL-SVMの重み付けの効果によると考えられる。また、SVMとSVM*の識別結果を比較すると、精度に9.0%の差がある。これはMKL-SVMの特徴選択の効果であると考えられる。

4 おわりに

本稿では、自閉症スペクトラム障害の早期発見を音響的な側面から目指し、MKL-SVMを用いて自閉症スペクトラム障害児と定型発達児の音響識別を行った。本研究では、484次元の音響特徴の次元毎にサブカーネルを定義し、識別に適した特徴重みをMKLにより学習させ、SVMに適用して識別を行った。MKL-SVMを用いることで、484次元の音響特徴から識別に有効な31の特徴次元を選択することができた。選択された特徴は、韻律的特徴よりも線スペクトル周波数 (lspFreq) やMFCCといった音韻的特徴が多く、自閉症スペクトラム障害児と定型発達児の音響識別をする上で、音韻的特徴が重要であると分かった。また本実験では、SVMとMKL-SVMとの比較を行い、MKLによる特徴選択の効果と重み付けの効果について調べた。484次元すべての特徴量を用いてMKL-SVMを行うと、31の特徴量が選択され、精度は80.2%となった。同様に、484次元すべての特徴量を用いて通常のSVMを行うと、精度は70.7%となった。またMKLで選択された31の特徴のみを用いて通常のSVMを行うと、精度は79.7%となり、MKL-SVMでの精度を0.5%下回った。これらの結果から、MKL-SVMの高い精度は、重み付けの効果よりも特徴選択の効果によると考えられる。今後は音声データ数を増やし、単語毎・発話者毎の発話分析を行い、またSVMやMKL-SVMだけでなくBoosting等の識別手法についても検討する。

参考文献

- [1] Andy Bondy and Lori Frost, *A Picture's Worth: PECS and Other Visual Communication Strategies in Autism. Topics in Autism.*, ERIC, 2002.
- [2] Shira Richman, *Raising a child with autism: a guide to applied behavior analysis for parents*, Jessica Kingsley Publishers, 2001.
- [3] Robert Paul Liberman, William J DeRisi, and Kim T Mueser, *Social skills training for psychiatric patients.*, Pergamon Press, 1989.
- [4] Scott M. Myers and Chris Plauché Johnson, "Management of children with autism spectrum disorders," *Pediatrics*, vol. 120, no. 5, pp. 1162–1182, 2007.
- [5] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "More efficiency in multiple kernel learning," in *Proc. ICML*, Corvallis, Oregon, USA, June 2007, pp. 775–782.
- [6] T. Ogawa, H. Hino, N. Reyhani, N. Murata, and T. Kobayashi, "Speaker recognition using multiple kernel learning based on conditional entropy minimization," in *ICASSP*, 2011, pp. 2204–2207.
- [7] C. Longworth and M.J.F. Gales, "Multiple kernel learning for speaker verification," in *Proc. ICASSP08*, Las Vegas, Nevada, USA, March 2008, pp. 1581–1584.
- [8] V. Rapp, T. Senechal, K. Bailly, and L. Prevost, "Multiple kernel learning SVM and statistical validation for facial landmark detection," in *Automatic Face Gesture Recognition and Workshops*, 2011, pp. 265–271.
- [9] P. Liang, G. Teodoro, L. Haibin, E. Blasch, G. Chen, and L. Bai, "Multiple kernel learning for vehicle detection in wide area motion imagery," in *International Conference on Information Fusion*, 2012, pp. 1629–1636.
- [10] M. Varma and D. Ray, "Learning the discriminative power-invariance trade-off," in *Proc. ICCV07*, Rio de Janeiro, Brazil, October 2007, pp. 1–8.
- [11] A. D. Dileep and C. Chandra Sekhar, "Representation and feature selection using multiple kernel learning," in *Proc. International Joint conference on Neural Networks*, 2009, pp. 717–722.
- [12] T. Joachims, *Making large-scale SVM learning practical*, MIT Press, Cambridge, MA, USA, 1999.
- [13] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.