

Kernel PCA を用いた残響下音声認識の検討*

○ 滝口 哲也, 有木 康雄 (神戸大・工)

1 はじめに

本稿では、Kernel PCA (Principal Component Analysis) を用いた音声強調手法について検討する。画像分野では、Kernel PCA はノイズ除去などに優れた性能を発揮することが知られており、[1] では、画質の粗い画像を高画質な画像へと変換を行う (Super-Resolution) 手法が提案されている。一方、これまで Kernel PCA を音声特徴量抽出において適用した研究も行われてきており、例えば [2] では、クリーン音声に対して MFCC 計算後、Kernel PCA を適用した結果が報告されている。本研究では、特に Kernel PCA による雑音抑圧に注目し、その有効性について検討を行う。

2 Kernel PCA による音声強調

Kernel PCA とは、カーネル関数 Φ を用いて、高次元の特徴空間で PCA を行う手法である [3]。この Kernel 法の特徴は、対象に対する事前知識をカーネル関数の形で表現することにある。クリーン音声の構造を Kernel 関数で表現する事が出来れば、雑音音声を特徴空間へ写像した際に、クリーン音声の構造のみを反映した特徴量が得られると期待出来る。

ここで、 d 次元観測ベクトルを $\mathbf{x}_j$ (j はフレーム番号) とすると、共分散行列 C は、

$$C = \frac{1}{N} \sum_{j=1}^N \bar{\Phi}(\mathbf{x}_j) \bar{\Phi}(\mathbf{x}_j)^T \quad (1)$$

$$\bar{\Phi}(\mathbf{x}_j) = \Phi(\mathbf{x}_j) - \frac{1}{N} \sum_{j=1}^N \Phi(\mathbf{x}_j) \quad (2)$$

となる (N は全フレーム数)。 C の固有値を λ 、固有ベクトルを $\mathbf{v}$ と置くと、

$$\lambda \mathbf{v} = C \mathbf{v} \quad (3)$$

$$\lambda (\bar{\Phi}(\mathbf{x}_k) \cdot \mathbf{v}) = (\bar{\Phi}(\mathbf{x}_k) \cdot C \mathbf{v}), \quad k = 1, \dots, N \quad (4)$$

が得られる。また $\mathbf{v}$ は以下のように表現出来るので、

$$\mathbf{v} = \sum_{i=1}^N \alpha_i \bar{\Phi}(\mathbf{x}_i) \quad (5)$$

式 (1) と (5) を (4) に代入して、

$$N \lambda \alpha = \bar{\mathbf{K}} \alpha \quad (6)$$

となり、最終的に次式で表される $\bar{K}$ の固有値問題に帰着する事になる。

$$\begin{aligned} \bar{K}_{ij} &= \bar{K}(\mathbf{x}_i, \mathbf{x}_j) = (\bar{\Phi}(\mathbf{x}_i) \cdot \bar{\Phi}(\mathbf{x}_j)) \\ &= K_{ij} - \frac{1}{N} \sum_{m=1}^N 1_{im} K_{mj} - \sum_{n=1}^N K_{in} 1_{nj} \\ &\quad + \frac{1}{N^2} \sum_{m,n=1}^N 1_{im} K_{mn} 1_{nj} \end{aligned} \quad (7)$$

$$1_{ij} = 1 \quad \text{for all } i, j \quad (8)$$

ただし、式 (7) に示されるように $\bar{K}_{ij}$ は $K_{ij} = (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j))$ から計算することが可能である。 $\bar{K}_{ij}$ の行列表現は次式で与えられる。

$$\bar{\mathbf{K}} = \mathbf{K} - \mathbf{1}_N \mathbf{K} - \mathbf{K} \mathbf{1}_N + \mathbf{1}_N \mathbf{K} \mathbf{1}_N \quad (9)$$

ここで、 $\mathbf{1}_N$ は全ての要素が $1/N$ である $N \times N$ 行列である。

このカーネル行列 $\bar{\mathbf{K}}$ の固有値問題を解き、固有ベクトルを $\mathbf{1} = \lambda^{(i)} (\alpha^{(i)} \cdot \alpha^{(i)})$ が満たされるように正規化を行うことにより、テストデータ $\mathbf{y}$ に対する第 i 主成分が次式により求まる。

$$\begin{aligned} (\mathbf{v}^{(i)} \cdot \bar{\Phi}(\mathbf{y})) &= \sum_{j=1}^N \alpha_j^{(i)} (\bar{\Phi}(\mathbf{x}_j) \cdot \bar{\Phi}(\mathbf{y})) \\ &= \sum_{j=1}^N \alpha_j^{(i)} \bar{K}^{test}(\mathbf{x}_j, \mathbf{y}) \end{aligned} \quad (10)$$

ここで、同様に $\bar{K}^{test}$ も K^{test} から求めることが出来る。

$$\bar{\mathbf{K}}^{test} = \mathbf{K}^{test} - \mathbf{1}'_N \mathbf{K} - \mathbf{K}^{test} \mathbf{1}_N + \mathbf{1}'_N \mathbf{K} \mathbf{1}_N \quad (11)$$

* A Study on Reverberant Speech Recognition Using Kernel PCA. by Tetsuya Takiguchi and Yasuo Ariki (Kobe University)

テストデータ y のフレーム数が L の場合、 $1'_N$ は要素が全て $1/N$ の $L \times N$ 行列となる。

本稿では、Kernel PCA を対数パワースペクトルに対して適用する (x_i を d 次元対数パワースペクトルとする)。すなわち、従来のケプストラム分析における離散コサイン変換のかわりに、Kernel PCA を適用する事により、雑音等に頑健な特徴量抽出を行う。また、全音素共通の Kernel 行列を求めた場合と、各音素毎に Kernel 行列を求めた場合の比較も行う。

3 認識実験

3.1 実験条件

評価用データとして残響音声を使用し、提案手法の有効性を検討する。残響音声の作成には、RWCP 実環境音声・音響データベースより残響時間 300ms のインパルス応答を使用した。カーネル行列の計算に使用したクリーン音声データのフレーム数は、 $N = 3500$ とした。これは音響モデルの学習データ (2620 単語) からランダムに選択した。本実験では、カーネル行列の計算の際に次式の多項式カーネルを使用した。

$$K(x, y) = (x \cdot y + 1)^p \quad (12)$$

タスクは語彙 1000 単語として、テストデータは男性話者一人が対象語彙を一回発声したものである。音響モデルは特定話者 HMM を使用した。MFCC16 次元での評価データに対する認識率は 78.1% (ベースライン) である。

3.2 実験結果

表 1 に多項式カーネル次数 $p = 1$ の認識結果を示す (式 (12) において $p = 1$)。対数パワースペクトルに対して Kernel PCA を適用した場合 (KPCA+log_spec.) 80 次元で 82.2% となった。一方、MFCC に対して Kernel PCA を適用した場合 (KPCA+MFCC)、16 次元で 79.9% となった。KPCA+MFCC では正の固有値が 16 次までしか存在しなかったため、16 次より大きい次元に対するパラメータ抽出は行われていない。多項式カーネル次数を $p = 2$ とした場合、KPCA+log_spec. では 80 次元で 65.5%、KPCA+MFCC では 64 次元で 74.7% となった。次数 $p = 2$ の場合は、どちらもベースラインの認識率 (78.1%) を改善出来なかった。データ毎に最適な次数の調整を行う必要があるといえる。

Table 1 単語認識率 (多項式カーネル次数 $p = 1$)

dim.	KPCA+log_spec.	KPCA+MFCC
12	*	73.2%
14	*	75.0%
16	73.5%	79.9%
32	78.5%	-
64	82.0%	-
80	82.2%	-
100	79.8%	-

また、各音素毎にカーネル関数を求めた場合 (各音素 $N = 1500$ フレーム使用) 多項式カーネル次数 $p = 1$ の時は、16 次元で 76.1% となった。本実験では、多項式カーネルしか検討しておらず、他のカーネル (例えばガウスカーネル) を使用した場合の結果についても今後検討していく必要がある。

4 おわりに

残響下における Kernel PCA を用いた音声特徴量抽出について検討を行った。Kernel 法では、選択されたカーネル、その次数の値によって、得られる結果が変わってくる。今後は様々なカーネルを使用した際の実験を行い、考察を続けていく。

参考文献

- [1] K. I. Kim, M. O. Franz, and B. Schölkopf, "Kernel Hebbian Algorithm for Single-Frame Super-Resolution," Statistical Learning in Computer Vision (SLCV 2004), pp. 135-149, 2004.
- [2] A. Lima, H. Zen, Y. Nankaku, C. Miyajima, K. Tokuda, and T. Kitamura, "On the Use of Kernel PCA for Feature Extraction in Speech Recognition," IEICE Trans. Inf. & Syst., Vol. E87-D, No.12, pp. 2802-2811, 2004.
- [3] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," Neural Computation, Vol.10, pp. 1299-1319, 1998.