0. 論文

A REGULARIZED KERNEL-BASED APPROACH TO UNSUPERVISED AUDIO SEGMENTATION

Zaïd Harchaoui[1], *Félicien Vallet*[1,2], *Alexandre Lung-Yut-Fong*[1], *Olivier Cappé*[1]

[1]LTCI, TELECOM ParisTech & CNRS
46 rue Barrault
75634 Paris cedex 13, France

[2]Institut National de l'Audiovisuel
4 avenue de l'Europe
94366 Bry-sur-Marne cedex, France

タイトル：A regularized kernel-based approach to unsupervised audio segmentation
著者：Zaid Harchaoui, Felicien Vallet, Alexandre Lung-Yut-Fong, Olivier Cappe

arXiv投稿日：
学会/ジャーナル：ICASSP 2009

1. どんなもの？

- カーネルフィッシャー判別比のより単純なバージョンに基づいた教師無し変化検出手法
- 正則化されたカーネルベースのルールを導入
- 計算が容易かつ，既知の漸近分布を持つ

2. 先行研究

- ガウス混合モデルに基づくパラメトリックモデリング
- MMDを用いたカーネル法
  - 以下のカーネルを用いる

$$\mathsf{MMD} = \frac{n_1 n_2}{n_1 + n_2} \langle \hat{\mu}_2 - \hat{\mu}_1, \hat{\mu}_2 - \hat{\mu}_1 \rangle_{\mathcal{H}} \ ,$$

  - ただし，n_1, n_2はサンプル数
  - $\hat{\mu}_1, \hat{\mu}_2$はデータにカーネルを適用した後の標本平均
- KCDを用いたカーネル法

$$\mathsf{KCD} = \frac{\alpha_1^T \mathbf{K}_{12} \alpha_2^T}{\sqrt{\alpha_1^T \mathbf{K}_{11} \alpha_1^T} \sqrt{\alpha_2^T \mathbf{K}_{22} \alpha_2^T}} \ .$$

When $\nu = 1$ for both 1-class SVMs, then KCD simply writes as

$$\mathsf{KCD} = \frac{\langle \hat{\mu}_1, \hat{\mu}_2 \rangle_{\mathcal{H}}}{\sqrt{\langle \hat{\mu}_1, \hat{\mu}_1 \rangle_{\mathcal{H}}} \sqrt{\langle \hat{\mu}_2, \hat{\mu}_2 \rangle_{\mathcal{H}}}} \ .$$

  - 1クラスSVMにするとシンプルな設定になる
- 今回紹介，提案する手法はスライディングウィンドウアプローチを取っている

    - ウィンドウ内で単一の変化点の発生する可能性がある
    - 変化点数を推論する手法もあるが計算が複雑かつ統計的有意性の評価を行うことができない

3. コアアイディア

- 以下を変化点検出のカーネル？(スコア？)として計算する

$$\mathrm{KFDR}(X_1^{(1)},\ldots,X_{n_1}^{(1)};X_1^{(2)},\ldots,X_{n_2}^{(2)}) = \frac{n_1 n_2}{n_1+n_2}\left\langle \hat{\mu}_2-\hat{\mu}_1, \mathcal{I}(\hat{\Sigma}_W;\gamma)(\hat{\mu}_2-\hat{\mu}_1)\right\rangle_{\mathcal{H}} \quad (2)$$

where

$$\hat{\Sigma}_W = \frac{n_1}{n_1+n_2}\hat{\Sigma}_1 + \frac{n_2}{n_1+n_2}\hat{\Sigma}_2$$

    - ただし，$\mathcal{I}(\hat{\Sigma}*W;1/d) = *sum*{p=1}^d\lambda_p^{-1}(e_p\otimes e_p)$
    - λ, eはそれぞれ$\hat{\Sigma}_W$の固有値，固有ベクトルとする
    - KFDRはサンプル数を大きくするとχ^2分布に近づくことが確認できる

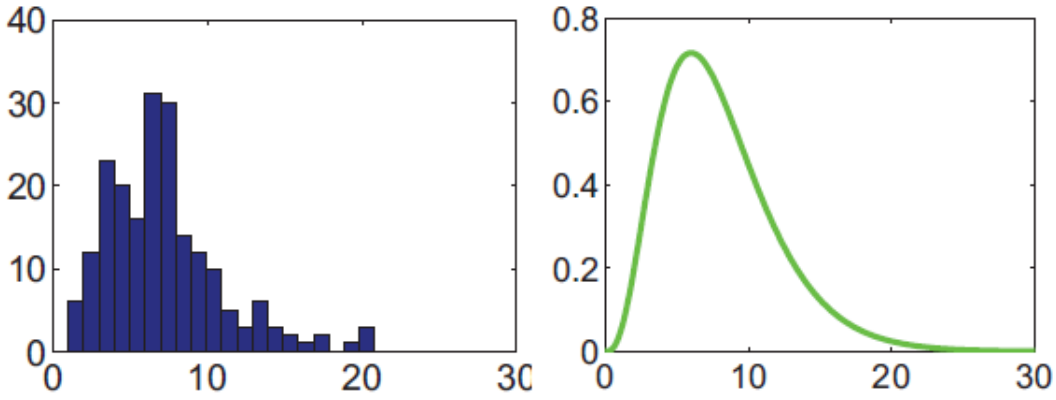

Fig. 1. Comparison of the finite-sample distribution of the test statistic against its large-sample distribution, based on 200 homogeneous windows of length 128 of the data considered in Section 5.1.

4. どうやって有効だと検証した？

- テレビ番組のデータ
    - 長いセグメントに対応するものがある(映画，音楽)
    - 長さ33の窓にわたって線形回帰を行う
        - そこで得られた傾きと切片を計算し，それらとRBFカーネルを組み合わせてカーネルを作る
    - PrecisionとRecallでは平均て故意に競合するアプローチよりも優れていることを確認
- 同様のデータで統計的要約？に取り組む
    - 長さ15の窓で線形回帰
    - 隠れマルコフモデルは難しい問題設定で訓練手順は非現実的(あまり参考にならない)

- 結果

	Semantic seg.		Speaker seg.	
	Precision	Recall	Precision	Recall
KFDR	0.72	0.63	0.89	0.90
MMD	0.71	0.58	0.76	0.73
KCD	0.65	0.63	0.78	0.74
HMM	0.73	0.65	0.93	0.96

Table 2. Best Precision and Recall for all benchmarked method, for both semantic segmentation and speaker segmentation tasks.

5. データセット

- 1980年代のテレビ番組の約三時間の2つのサウンドトラック
  - 13の特徴がある(12個の係数，切片項)

6. 疑問点

- カーネル法でどの様に変化点を検出しているのかが分からない
  - どの様な基準で変化点を選んでいるのか？
  - 何かを最大or最小にするように選んでいる気はするが
- カーネル法を知りたいならもっと基本的な論文を読むべきなきがする

7. 次に読むべき論文は？

- A kernel method for the two-sample problem

キーワード

- 変化点検出
- カーネル法
- オーディオセグメンテーション