

## 0. 論文

### RANK-BASED MULTIPLE CHANGE-POINT DETECTION IN MULTIVARIATE TIME SERIES

*F. Harlé<sup>\*†</sup>, F. Chatelain<sup>†</sup>, C. Gouy-Pailler<sup>\*</sup>, S. Achard<sup>†</sup>*

<sup>\*</sup> CEA, LIST, LADIS, 91191 Gif-sur-Yvette CEDEX, France

<sup>†</sup> University of Grenoble, GIPSA-Lab, 11 rue des Mathématiques, 38402 St Martin d'Hères, France

タイトル : RANK-BASED MULTIPLE CHANGE-POINT DETECTION IN MULTIVARIATE TIME SERIES

著者 :

arXiv投稿日 :

学会/ジャーナル : European Signal Processing Conference 2014

## 1. どんなもの？

- 多次元系列データのセグメンテーションのためのベイズ的アプローチ
- ランク検定のp値をモデル化して、有意な変化点に有利な疑似尤度の提案
- 事後分布に対してMCMC法が適用される

## 2. 先行研究

- 1次元系列データの変化点検出(多次元に拡張可能)
  - Detection of abrupt changes: theory and application(1993)
  - Kernel-based methods for hypothesis testing(2013)
  - Empirical likelihood ration test for the change-point problem(2007)
- ゲノミクスで提案されているもの
  - fused lasso latent feature model
- 階層的ベイズモデルの導入したもの
- 今回の問題設定は断片的に定常な共同セグメンテーションなので少し合わないか

## 3. コアアイデア

- データ :  $X \in \mathbb{R}^{K \times N}$
- 例 :  $x_{(j,i)}$  はセンサー  $j$  の  $i$  時点目での観測とする
- 変化点かどうかを示す行列  $R$  を導入
  - $r_{(j,i)} = 1$  なら  $(j,i)$  が変化点とする
- この  $R$  に加えて共依存性の推定も行う必要がある
  - $j$  と依存性のある信号は  $j$  に変化があれば、その信号も同様の変化がある
- 尤度関数として以下を考える

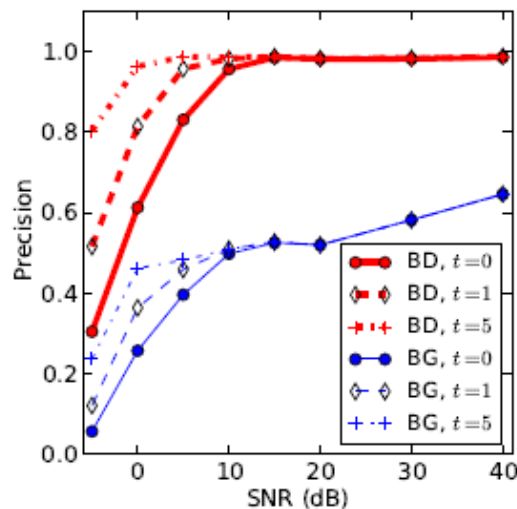
$$f(R, P|X) \propto L(X|R)f(R|P)f(P)$$

- 変化点検出として以下の流れ(あまり詳しく追っていない)
  - 時系列間の依存性構造をモデル化するために  $R, P$  の事前分布を選択
  - 順位検定をベースにした疑似尤度  $L$  を設定

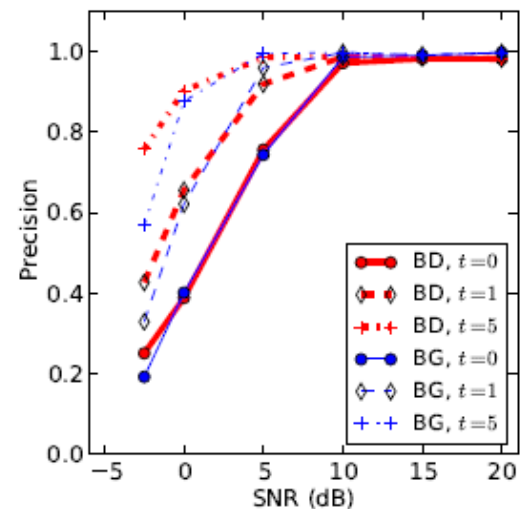
- 上の尤度関数から変化点に関するパラメータ $R$ の事後分布が推定される
- MCMC法とギブズサンプリングを用いてMAP推定を行い変化点検出とする

#### 4. どうやって有効だと検証した？

- 系列長100の二つの系列を人工的に作成した設定
  - パラメータ $k$ を持つ非標準化 $t$ 分布に従うデータ
  - 正規分布 $N(k_i)$ に従うデータ
  - 古典的なベルヌーイガウスモデルとの比較



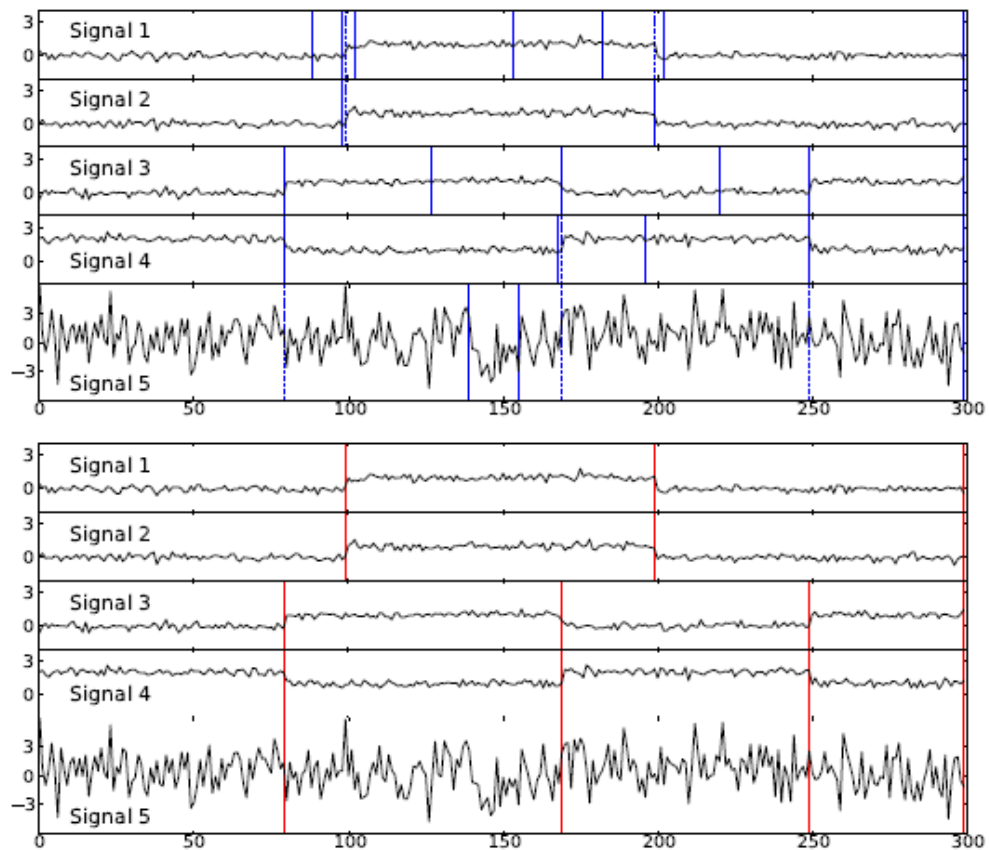
(a) Student's  $t$ -distribution



(b) Normal distribution

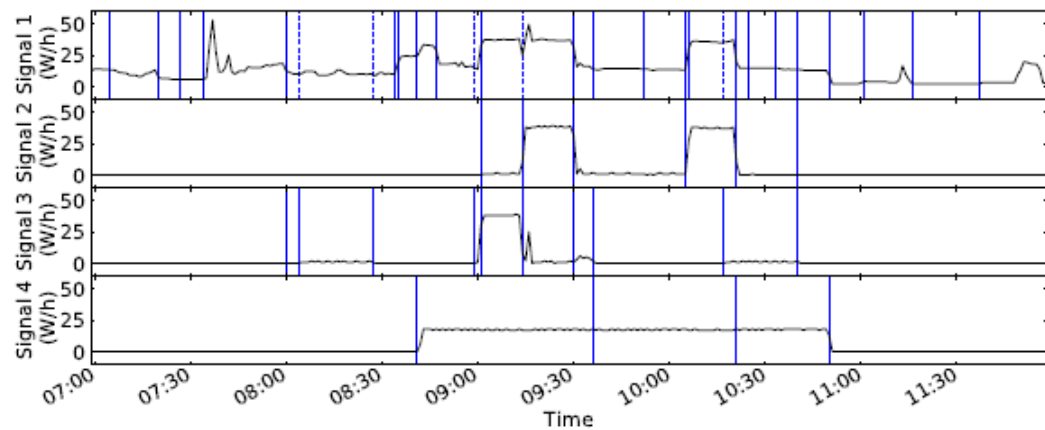
- - 裾の思い $t$ 分布の方で提案法(赤線)がより良い結果
    - 既存手法よりロバストなモデル
  - 正規分布の方でも既存手法と同程度
    - 一般性を保つことを確認できる
- 二つのグループに従うデータセット
  - グループ1 : 信号1, 2(2つの変化点を持つ)
  - グループ2 : 信号3, 4, 5(3つの変化点を持つ)
  - 計算コストが高いものの信号5ではすべての変化点を見つけられる

○ 結果

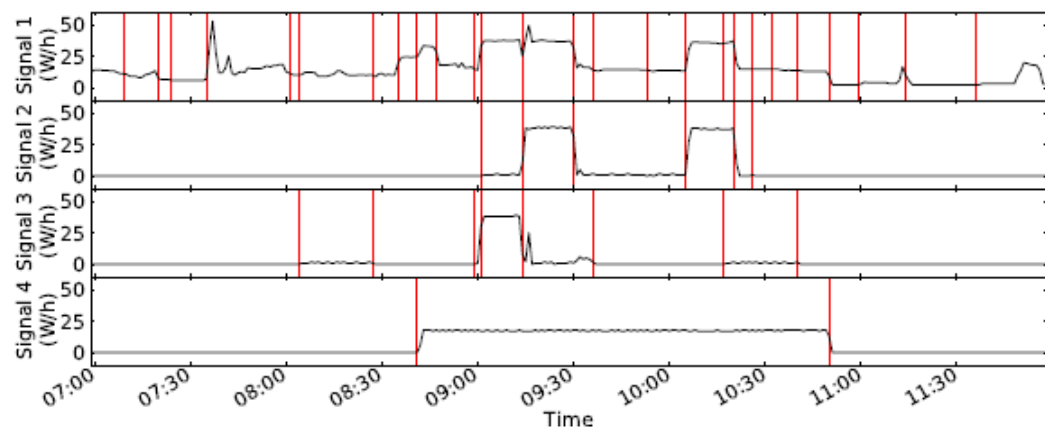


**Fig. 2.** At top, independent MAP estimation of each signal. At bottom, joint estimation with a non-informative prior on the change-points structure. Real change-points are represented in dashed lines.

- Household electric power consumption data
  - 4つの信号からなる(1つは家の中でグローバルな消費電力, 他はある特定の家電の消費電力)
  - 信号1は他の2, 3, 4の信号に変化があったら変化が起こりやすい
  - 他の信号同士は独立



(a) Non-informative prior



(b) Informative prior

**Fig. 4.** Change-point detection in real data (MAP). At top (a), change-points found for signals 2, 3 and 4 but not for global signal 1 are represented in dashed lines.

- 事前分布が適切に設定されると上手く変化点の検出ができるようになる

## 5. データセット

- Household electric power consumption data

## 6. 疑問点

- ベイズ的アプローチを使うことによる利点は
  - 平均の変化を検出するだけならこんな回りくどい手法を使わなくても良い気がする
  - 事後分布の推定にサンプリング(ギブスサンプリング)を使う必要があるのであまり妥当な検定と言わない気がする

## 7. 次に読むべき論文は？

- 応用的な論文の方がいいかもしれない
- 応用分野(脳波, 財政, センサーなど)でよく使われている変化点検出手法にフォーカスしたい

## キーワード

- Rank統計量
- jointセグメンテーション
- ベイズ推定
- MCMC法
- ギブズサンプリング