

0. 論文

The group fused Lasso for multiple change-point detection

Kevin Bleakley
INRIA Saclay, Orsay, France
kevbleakley@gmail.com

Jean-Philippe Vert
Mines ParisTech CBIO, Fontainebleau, France
Institut Curie, Paris, France
INSERM U900, Paris, France
Jean-Philippe.Vert@mines.org

タイトル : [The group fused Lasso for multiple change-point detection](#)

著者 : Kevin Bleakley, Jean-Philippe Vert

arXiv投稿日 : 2011/6/21 学会/ジャーナル :

1. どんなもの ?

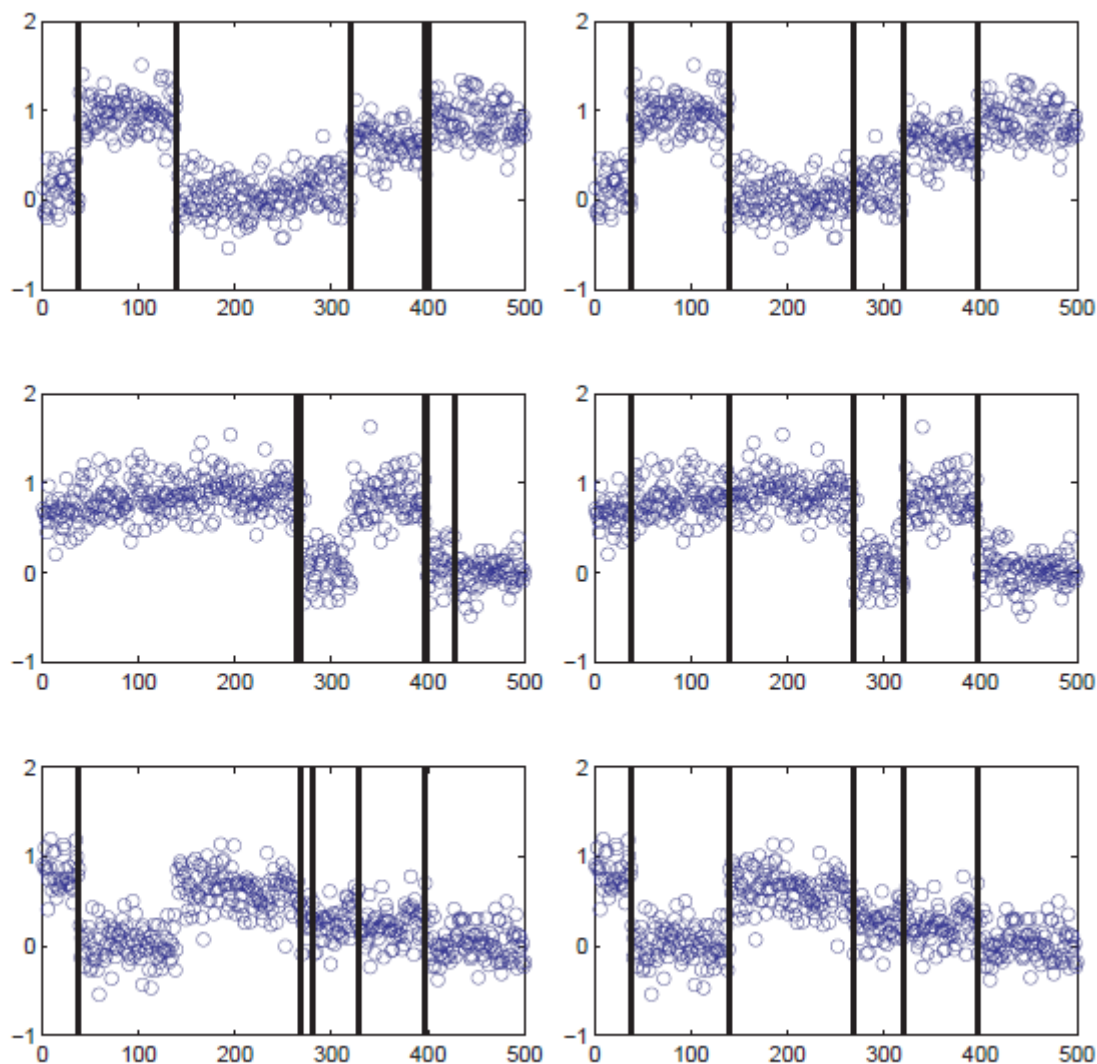
- 一次元信号の集合が共有する変化点を検出するためにgroup fused lassoを用いたものの提示
- 元信号の全変動に制約を与えて近似することを目標にする
- betaが変化するところが変化点とする ?
- 01の論文とそんなに変わらない(論文中では01が予備のバージョンと書いてある)

2. 先行研究

- この論文での議論はピースワイズコンスタントなモデルを二次誤差基準を用いて近似する手法に限定する
- 動的計画法を用いた変化点検出
 - 計算量に問題あり
- バイナリーセグメンテーションを用いた変化点検出
 - 解の信頼性が動的計画法よりもない

3. コアアイデア

- 変化点検出手法をグループfusedラッソに落とし込む



-
- 個別で変化点検出したものとの比較
- 一番下のものに関してはすべて手の変化点を得られている
- 01論文との差別化点は d_j と書かれているもので重みづけされていること
 - 位置依存の重みで異なる位置の異なる大きさのジャンプに対応している
- 提案されているアルゴリズムは二つ
 - 座標降下法: 厳密に解く方法, 変化点数に関して3乗のオーダーが掛かる
 - Larsアルゴリズム: group LARSをもとにしたアルゴリズム
 - 変化点を1つずつ追加していき収束するまでアルゴリズムが続く
- 変化点数未知の場合の拡張もある
 - ある λ で選ばれる変化点数 k が存在する(OverEstimateするように設定)
 - $[1, \dots, k]$ の中で一番よさげなものを選択
 - 具体的にはセグメント間で平均を取ってその平均の差の和が一番小さくなるような k を選択する
 - 動的計画法を用いれば効率的に探索することができる

4. どうやって有効だと検証した？

- 計算効率性の比較
 - Larsアルゴリズム(反復最適化)vs座標降下法の比較
 - 系列長や系列数が増えると違いがなくなってくる
 - 基本的にLARSアルゴリズムが早いことが分かっているが, データが大きくなるにつれて最適化問題が解きやすくなった可能性あり

- 変化点数が大きくなった時はLarsが圧倒的に勝つ
- アルゴリズムの中身で k^3 と k に関して線形の違いがある

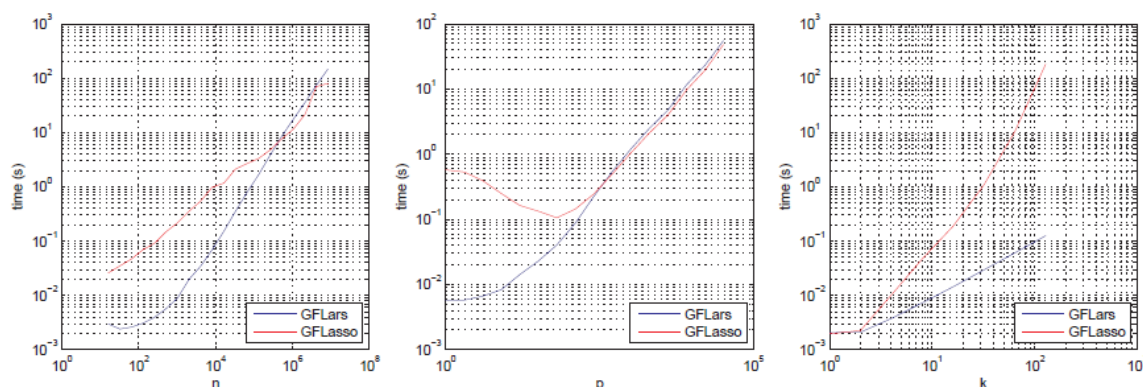


Figure 2: **Speed trials for group fused LARS (top row) and Lasso (bottom row).** Left column: varying n , with fixed $p = 10$ and $k = 10$; center column: varying p , with fixed $n = 1000$ and $k = 10$; right column: varying k , with fixed $n = 1000$ and $p = 10$. Figure axes are log-log. Results are averaged over 100 trials.

- 精度の確認
 - 変化点位置を変更した場合の精度比較
 - オプションとして重み付きと重み無しがある
 - 重み無しは $u > 80$ になると0に収束することが分かる(01もそうだったけどなぜか分からん)
 - 重み付きはある程度ロバストな性能になっている

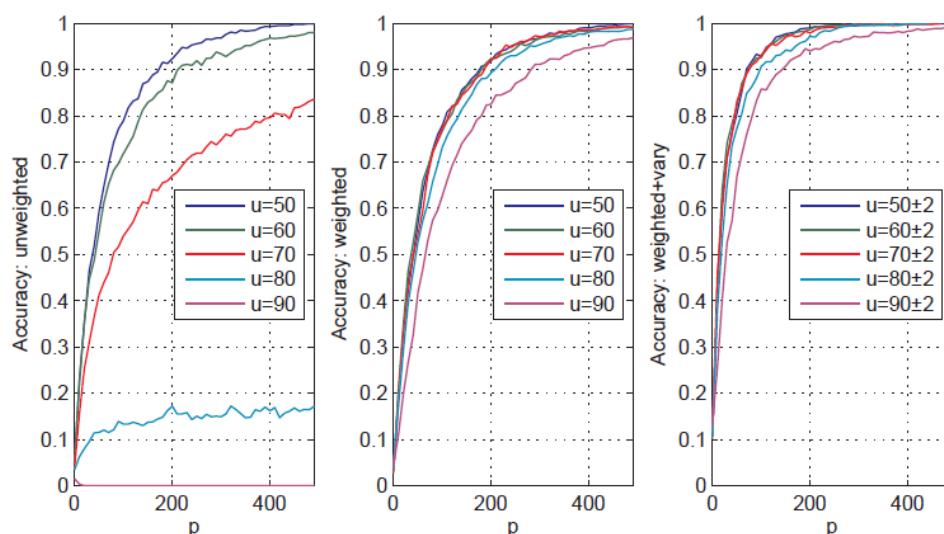


Figure 3: **Single change-point accuracy for the group fused Lasso.** Accuracy as a function of the number of profiles p when the change-point is placed in a variety of positions $u = 50$ to $u = 90$ (left and centre plots, resp. unweighted and weighted group fused Lasso), or: $u = 50 \pm 2$ to $u = 90 \pm 2$ (right plot, weighted with varying change-point location), for a signal of length 100.

- arrayCGHデータを使った実験
 - arrayCGHデータはDNAのコピー数変化を解析したデータ
 - コピー数が増大, 減少している領域を検出したいモチベーションがある
 - 全ての系列で同じ変化点を共有している仮定ではないことに注意
 - 次元を減らし, ノイズ除去の適応的な方法としている
 - 変化点に対して統計的検定をすることができると記述あり
 - やっているわけではない
 - H-HMMと呼ばれる手法との比較を行っていて同様の変化が検出がされている

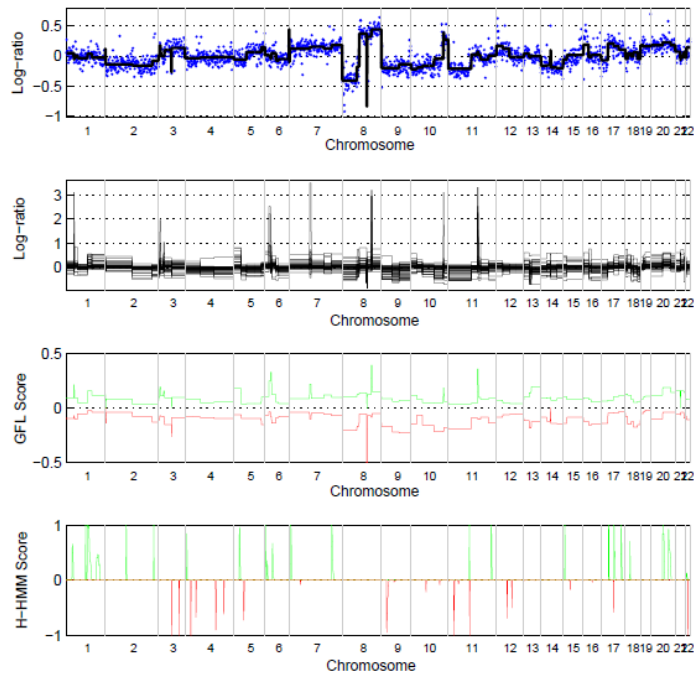


Figure 6: **Bladder cancer profiles**. (a) shows one of the original 57 profiles and its associated smoothed version. (b) shows the result of superimposing the smoothed versions of the 57 bladder tumor aCGH profiles obtained using weighted group fused LARS followed by dimension-selection. (c) shows the result of transforming the set of smoothed outputs into “scores” for amplification/deletion (see Section 6.4) and (d) the corresponding output for the H-HMM method [27]. Vertical black lines indicate chromosome boundaries.

○

5. データセット

- 膀胱がん患者のデータセット

6. 疑問点

- 図を見てみた感じ共有されている変化点の検出となっている
 - やはり変化している系列としていない系列が混在していて選択できるモデルにはなっていない
- グループ化しているのはあくまでも周りの要素をグループ化している

7. 次に読むべき論文は？

- 金融部分のもの: Structural learning with time-varying components: tracking the crosssection of financial time series

キーワード