

0. 論文

Fast detection of multiple change-points shared by many signals using group LARS

Jean-Philippe Vert and Kevin Bleakley
Mines ParisTech CBIO, Institut Curie, INSERM U900
{firstname.lastname}@mines-paristech.fr

タイトル : [Fast detection of multiple change-points shared by many signals using group LARS](#)

著者 : Jean-Philippe Veart and Kevin Bleakley

arXiv投稿日 :

学会/ジャーナル : NeurIPS2010

1. どんなもの？

- 複数の一次元信号が頻繁に変化点を共有している場合の高速な変化点検出アルゴリズムを提案
- 信号数が増えた場合でも一貫した変化点検出ができる
- 今回はゲノム解析に焦点を当てているので特異な状況になっている
 - 系列数は増加する可能性があるが、系列長は固定

2. 先行研究

- 動的計画法による探索がある
 - これは計算量が $O(n^2pk)$ となり、系列長に関して2乗のオーダーになっている
 - Estimating the number of change-points via schwarz Criterion
 - Gaussian model selection
 - Detection of multiple change-points in multivariate time series
- バイナリセグメンテーションを用いた手法
 - これは1つの変化点を探索するアルゴリズムを再帰的に適用するもの
 - $O(np \cdot \log(k))$ となり計算量の問題は解決
 - 動的計画法は大域的な最適解を返すのに対し、こちらは解の品質に問題あり(おそらく近似的な解になる)
- 1次元系列の場合
 - セグメンテーション結果を凸最適化問題として定式できる(Lassoに落とし込むことができる)

3. コアアイデア

- 変化点検出の問題をLassoの問題に落とし込んでいる

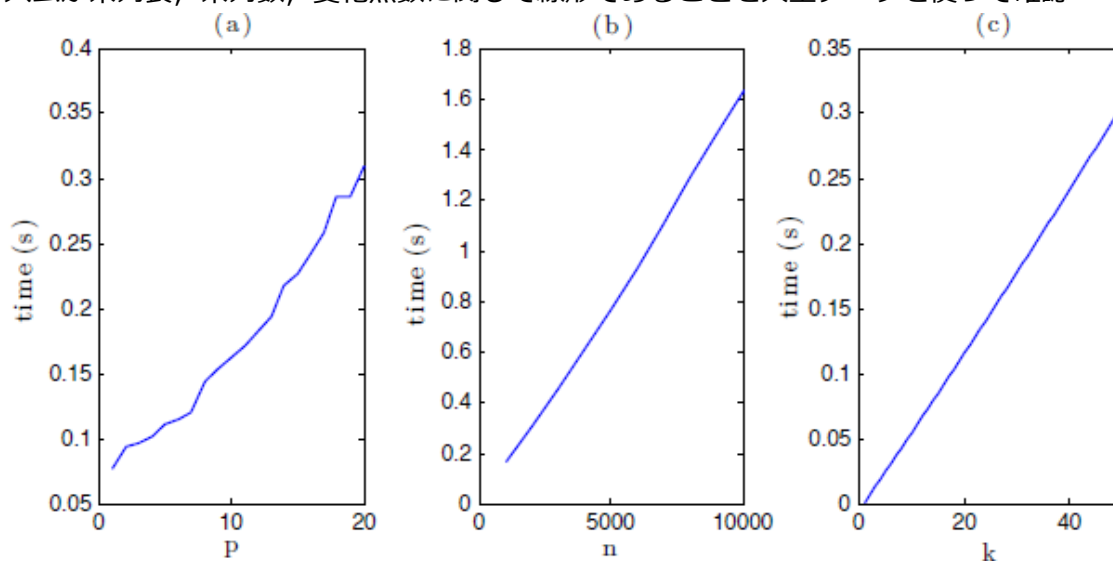
- 具体的には以下の式の最適化を考えている

$$\min_{U \in \mathbb{R}^{n \times p}} |Y - U| + \lambda \sum_{i=1}^{n-1} |U_{i+1, \cdot} - U_{i, \cdot}|$$

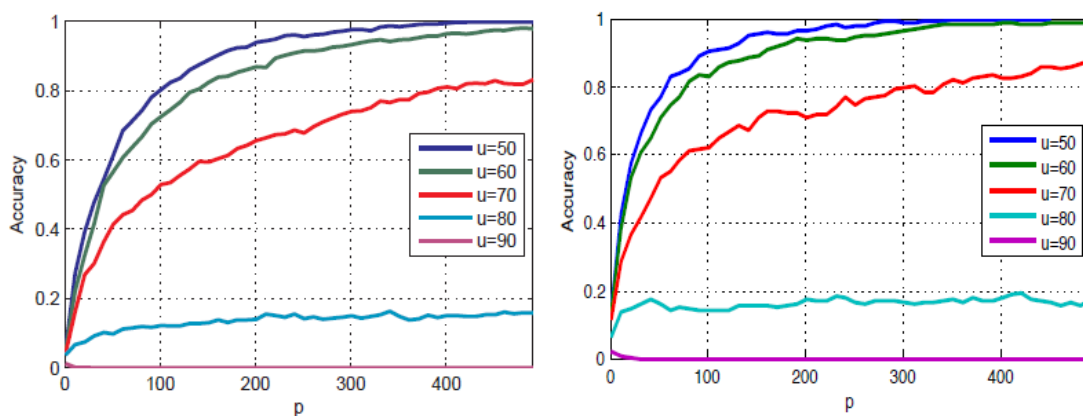
- ここでのUは正規分布の平均行列Mだと思えばよい
- 提案されているアルゴリズムはLARSアルゴリズムがベースになったもの
 - これは解の個数を1つずつ増やしていく手法

4. どうやって有効だと検証した？

- アルゴリズムが系列長，系列数，変化点数に関して線形であることを人工データを使って確認

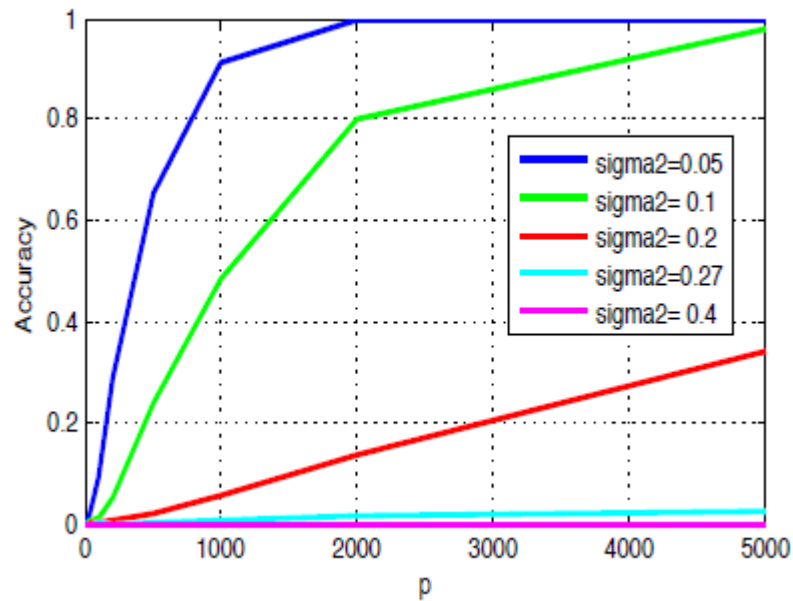


- - u(変化時点)の違いによる精度の確認
 - uを80より大きくすると0に近づくことがよくわからない
 - プログラムの性質上仕方がないものなのか，どこかで議論している可能性はある

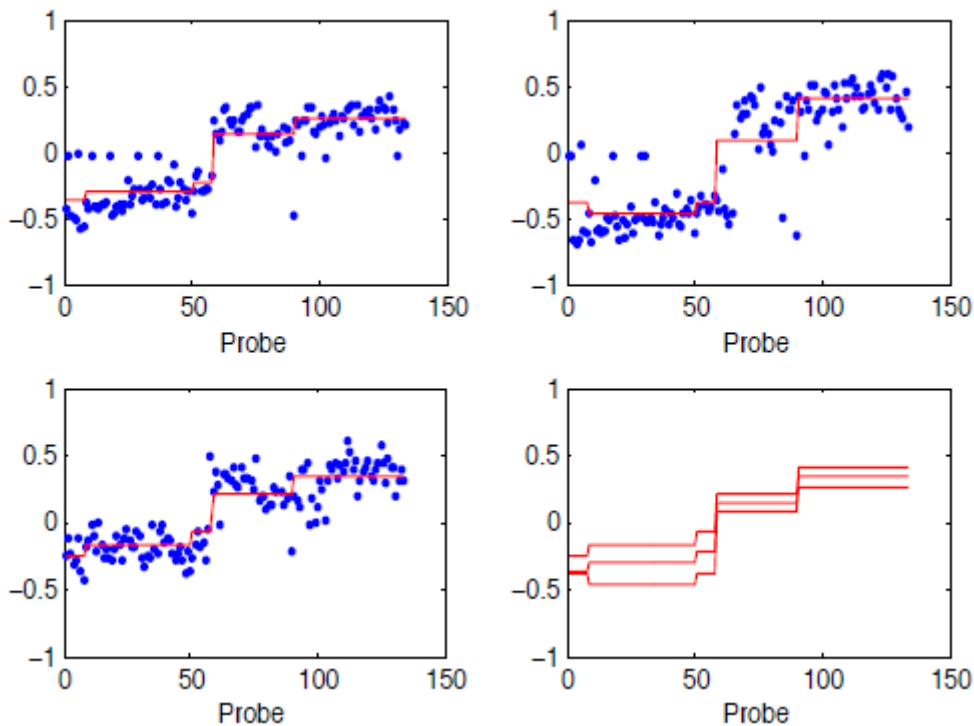


- 複数変化時点における精度の確認

- σ の値を正しく設定しないと精度は大幅に変わる



- ある共通した病気を持つ個人の遺伝子に関してゲノムのコピー数変化を分析(array-CGHみたいなもの)
 - ある遺伝子では3人の患者が同様の変化を持つことが確認できた
 - 最先端のH-HMMと呼ばれる手法がある



5. データセット

- 膀胱がんにおけるコピー数変化のデータ

6. 疑問点

- group lassoの問題に落としていることは分かった
 - 何をスパースにしている？
 - 何をグループ化している？
 - 見た感じでは系列数に関してスパースな変化点を検出するかどうかは分かっていない

7. 次に読むべき論文は？

- アップデートした論文
 - The group fused Lasso for multiple change-point detection
- 動的計画法を用いた手法
 - Estimating the number of change-points via schwarz Criterion
 - Gaussian model selection
 - Detection of multiple change-points in multivariate time series
- バイナリーセグメンテーションを用いた手法
 - Detection of disorder in multidimensional stochastic processes
 - Adaptive detection of multiple change-points in asset price volatility

キーワード

- 変化点検出
- 多次元系列データ
- group lasso