0. 論文

The Annals of Statistics
2013, Vol. 41, No. 2, 670–692
DOI: 10.1214/13-AOS1094

SEQUENTIAL MULTI-SENSOR CHANGE-POINT DETECTION[1]

BY YAO XIE AND DAVID SIEGMUND

Duke University and Stanford University

タイトル：Sequential multi-sensor change-point detection
著者：

arXiv投稿日：2012/7/10 学会/ジャーナル：Annals of Statistics 2013

1. どんなもの？

- データのサブセットのみに影響を与える変化点を監視する手法
    - 系列の選択を行っている
- 変化点のあとでは変化する系列が非ゼロの平均を持つことを仮定
    - 変化点前はゼロの平均を持つということか
- 尤度比統計量を使用して割合p0を用いた混合モデルで全体的な検出統計量を形成する

2. 先行研究

- 非構造化問題
    - 構造化されていないデータを対象とした手法(構造化されたデータとは？)
        - 本手法もこれにあたる
        - Detection and localization of change-points in high-dimensional network traffic data
        - Body area networks: A survey
        - Efficient scalable schemes for monitoring a large number of data streams
        - Detecting the emergence of a signal in a noisy image
- CUSUMを使ったもの
    - Meiの手法
        - 個々のセンサからのCUSUM統計量の和を利用
        - 適切な閾値と比較して停止規則を作る
        - 変化点前後のデータ分布は完全に既知である設定
        - $$T_{\mathrm{Mei}} = \inf\left\{ t : \sum_{n=1}^{N} \max_{0 \le k < t} \ell_n(t,k,\delta) \ge b \right\},$$
    - Tartakovskyらの手法
        - 局所尤度比統計量を合計する基準を提案
        - 変化していないセンサからのノイズが含まれるので検出遅延が遅くなる可能性がある

- $$T_{\mathrm{TV}} \triangleq \inf\left\{t : \max_{0 \le k < t} \sum_{n=1}^{N} \ell_n(t,k,\delta) \ge b\right\}.$$

3. コアアイディア

- データについて
  - N個のセンサーが時間$t = 1, 2, \ldots$で取られるデータ$y_{n,i}$が観測される
  - 変化を検出した停止基準を定めることを目標にする
- 対数尤度を定義する
  - $$\ell_n(t,k,\mu_n) = \sum_{i=k+1}^{t} (\mu_n y_{n,i} - \mu_n^2/2).$$
  - μ_nはセンサーnの未知パラメータ
- 提案手法は4つの停止基準を提案
  - $$T_2 = \inf\left\{t : \max_{0 \le k < t} \sum_{n=1}^{N} \log(1 - p_0 + p_0 \exp[(U_{n,k,t}^{+})^2/2]) \ge b\right\}.$$
  - $$T_{\max} = \inf\left\{t : \max_{0 \le k < t} \max_{1 \le n \le N} (U_{n,k,t}^{+})^2/2 \ge b\right\}.$$
  - $$T_3 = \inf\left\{t : \max_{0 \le k < t} \sum_{n=1}^{N} [\ell_n(t,k,\delta) + \log(p_0)]^{+} \ge b\right\}$$
  - $$T_4 = \inf\left\{t : \max_{0 \le k < t} \sum_{n=1}^{N} [(U_{n,k,t}^{+})^2/2 + \log(p_0)]^{+} \ge b\right\},$$
  - $S_{n,t} = \sum_{i=1}^ty_{n,i}$
  - $U_{n,k,t} = (t-k)^{-1/2}(S_{n,t}-S_{n,k})$
  - $l_n(t,k,\hat{\mu}*n) = (U*{n,k,t}^+)^2/2$
  - x^+はxの正の部分を集めたものになる
- p_0のパラメータに左右されるがその選択も入る検出統計量T_{parallel}も考えられる
- $$T_{\mathrm{parallel}} = \min T_2(p_1), T_2(p_2)$$

4. どうやって有効だと検証した？

- 人工データを用いて期待遅延を比較する
  - データとしては系列長と系列数は固定
    - 変化系列数割合pを変化させたときの遅延の比較になっている
  - maxを取るものはpの値が小さい時には有効に働くが，大きくなると遅延が大きくなる
  - Meiの手法ではpが大きい時に上手くいくが，小さい時には他の手法より劣る結果に
  - T_2とT_3の結果は同等である

TABLE 5
EDD with $N = 100$ obtained from 500 Monte Carlo trials. Thresholds for ARL 5000 are listed in Table 4. Theoretical approximations for EDD are in parentheses

p	Method	EDD, $\mu = 1$	EDD, $\mu = 0.7$	EDD, $\mu = 1.3$
0.01	max	25.5	49.6	16.3
	$T_2(1)$	52.3 (56.9)	105.5 (114.6)	32.9 (34.1)
	$T_2(0.1)$	31.6 (32.5)	59.4 (64.9)	20.3 (19.7)
	Mei	53.2	103.8	38.1
	$T_3(0.1, 1)$	29.1 (29.3)	63.3 (59.0)	19.1 (19.1)
	$T_3(1, 1)$	82.0 (83.6)	213.7 (193.5)	53.3 (53.5)
0.03	max	18.1	33.3	11.6
	$T_2(1)$	18.7 (19.3)	35.8 (38.4)	12.6 (11.7)
	$T_2(0.1)$	14.2 (13.9)	26.7 (27.5)	9.3 (8.5)
	Mei	23.0	41.6	16.4
	$T_3(0.1)$	13.4	26.9	9.2
	$T_3(1)$	27.2	66.0	16.3
0.05	max	15.5	28.4	9.7
	$T_2(1)$	12.2 (11.6)	21.8 (23.0)	7.9 (7.1)
	$T_2(0.1)$	10.4 (10.1)	18.9 (19.9)	6.9 (6.2)
	Mei	15.7	26.9	11.4
	$T_3(0.1, 1)$	9.8 (9.8)	18.6 (21.4)	7.0 (6.8)
	$T_3(1, 1)$	15.5 (16.2)	38.8 (39.8)	9.0 (9.7)
0.1	max	12.6	23.0	8.4
	$T_2(1)$	6.7 (5.9)	11.8 (11.3)	4.7 (3.7)
	$T_2(0.1)$	6.7 (7.2)	11.6 (14.1)	4.6 (4.5)
	Mei	9.6	15.4	7.4
	$T_3(0.1, 1)$	7.1 (7.6)	11.9 (16.7)	5.3 (5.3)
	$T_3(1, 1)$	6.8 (7.3)	15.7 (19.6)	4.6 (4.5)
0.3	max	9.6	16.7	6.6
	$T_2(1)$	3.0 (2.0)	4.4 (3.5)	2.4 (1.4)
	$T_2(0.1)$	3.5 (5.2)	5.6 (10.1)	2.7 (3.3)
	Mei	4.9	7.0	4.0
	$T_3(0.1, 1)$	4.6	6.7	3.9
	$T_3(1, 1)$	3.0	4.3	2.5
0.5	max	8.6	14.4	5.8
	$T_2(1)$	2.3	3.0	2.0
	$T_2(0.1)$	2.8	4.0	2.1
	Mei	3.8	5.0	3.0
	$T_3(0.1, 1)$	4.0	5.4	3.3
	$T_3(1, 1)$	2.3	3.0	2.0
1	max	7.2	12.1	5.1
	$T_2(1)$	2.0	2.0	2.0
	$T_2(0.1)$	2.0	2.6	2.0
	Mei	3.0	3.4	2.3
	$T_3(0.1, 1)$	3.4	4.3	3.0
	$T_3(1, 1)$	2.0	2.1	2.0

5. データセット

- 実データ実験は無し

6. 疑問点

7. 次に読むべき論文は？

- Meiの手法
    - Efficient scalable schemes for monitoring a large number of data streams.
- Tartakovskyらの手法

    - Asymptotically optimal quickest change detection in distributed sensor system

キーワード