Change point detection for load balancing based on content popularity

Sotiris Skaperas [1] Lefteris Mamatas [1]

[1] Department of Applied Informatics,
University of Macedonia,
e-mail: sotskap@uom.edu.gr; emamatas@uom.edu.gr

Mathematics of Networks meeting, University of Sussex
September 12, 2017

Outline

Problem description

- **Traffic load balancing for content distribution**

- **We propose to use lightweight web servers that are ephemeral:**
  - appear and disappear rapidly.
  - consume very low resources.

The idea is to host one video per web server.

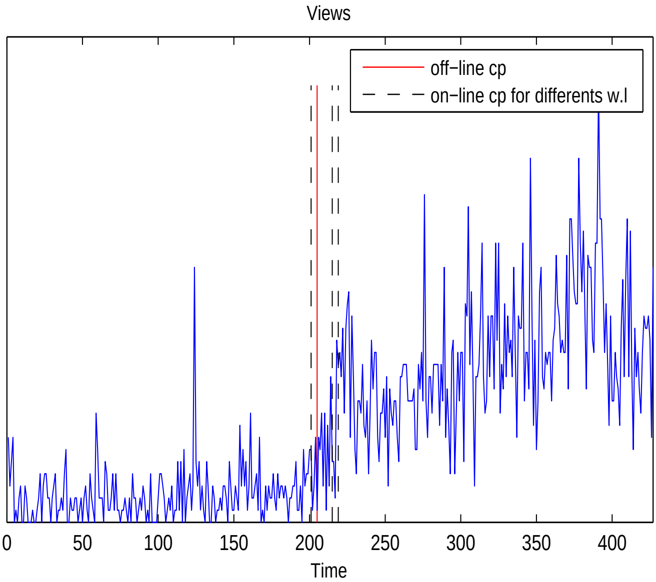
- **Using the unikernel technology is ideal:**
  - Single-purpose lightweight virtual machines.
  - A webserver can boot up in 80 ms and have a size of 10MB.
  - A physical machine can host hundreds (e.g., 500 in a laptop).

- **We have as an input the content popularity**
  - whenever a content starting becoming popular, significant traffic will arrive soon.
  - we require a rapid response, i.e., deploy more unikernels with content replicas

This talk focuses on the detection of content popularity along the above lines.

Requirements for video popularity detection

- **Low complexity while accurate**
  - quick estimation, e.g., low convergence time.
- **Non parametric framework**
  - no restrictive assumptions in the time series structure.
- **Magnitude estimation**
  - Adjusts the volume of traffic peak mitigation strategy.
- **On-line operation**


Views
off−line cp
on−line cp for differents w.l
0
50
100
150
200
250
300
350
400
Time

Our contribution

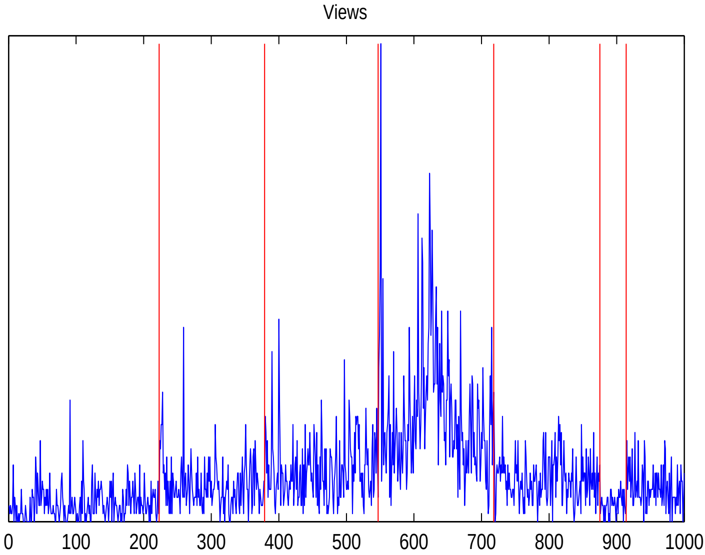
- **Real youtube video views**
  - Selected around 1000 videos.
- **Off-line change point detection (cpd)**
  - Augmented with trend indicator and our heuristic segmentation algorithm.
  - Studying detected change point characteristics.
  - Adjusting online cpd method.
  - Acting as a reference point to validate the on-line cpd.
- **Selecting and adjusting an on-line c.p method along these lines.**
  - produces promising results.
  - will be deployed in a real test-bed.

Mathematical approach

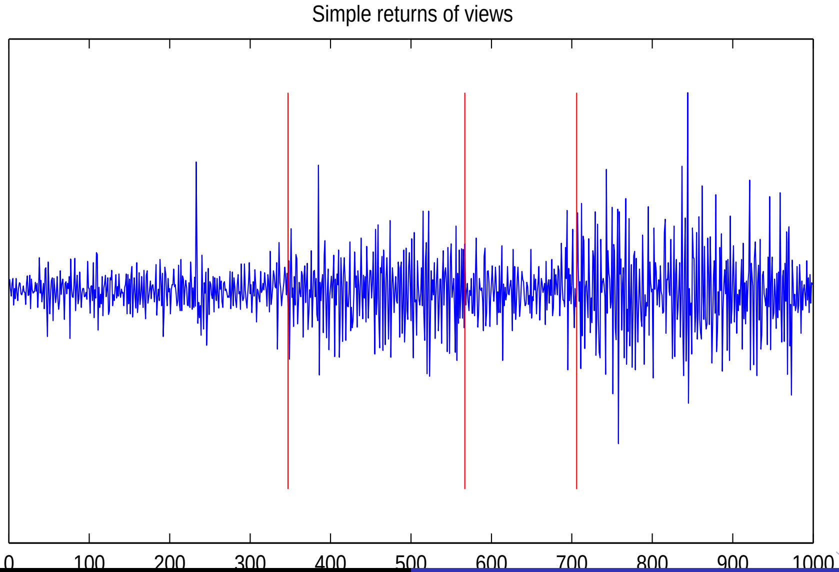
Overview

- Change point analysis
- Moving average convergence divergence

Change point in the mean structure



Change point in the variance structure



Classification of change-point problems

1. How data is received.
    - Off- line data collection i.e, retrospective methods.
    - On- line data collection i.e, sequential methods.
2. Statistical information of the r.v sequence.
    - Non- parametric methods.
    - Parametric methods.
3. Type of change point
    - Mean
    - 2nd order characteristics

Assumption for the process $\{y_t : t \epsilon \mathbb{Z}\}$ under H_0

Assumption

1. $f : \mathbb{R}^{d' x \infty} \to \mathbb{R}^d$
   $y_t = f(\epsilon_t, \epsilon_{t-1}, \cdots), (\epsilon_t : t \epsilon \mathbb{Z})$ iid process.
2. $f^{(m)} : \mathbb{R}^{d' x (m+1)} \to \mathbb{R}^d$
   $y_t^{(m)} = f^{(m)}(\epsilon_t, ..., \epsilon_{t-m}), (y_t^{(m)} : t \epsilon \mathbb{Z})$ a sequence of m-dependent r.v.
3. $\sum_{m \geq 1} \|y_0 - y_0^{(m)}\| < \infty$.

Assumption admits a causal representation (not only linear) in terms of the iid sequence. Also a weak dependence structure is enabled.

Change point in the unconditional mean

- Let $\{Y_t : t \epsilon \mathbb{Z}\}$ be a sequence of d- dimensional random vectors, where,

$$Y_j = \mu_t + \varepsilon_t$$

- Null hypothesis is

$$H_0 : \mu_1 = \cdots = \mu_n = \mu$$

against the hypothesis of at least one change.

Horvath et al. (1999)

CUSUM process Z_n is defined as,

$$Z_n(k) = \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{k} Y_t - \frac{k}{n} \sum_{t=1}^{n} Y_t \right)$$

Change point in the unconditional mean

Test statistic is given by,

$$M_n = \max_{1 \leqslant k \leqslant n} Z_k' \widehat{\Omega}_n^{-1} Z_k.$$

Asymptotic behaviour under H_0,

$$M_n \Rightarrow \sup_{0 \leqslant t \leqslant 1} \sum_{l=1}^{d} B_l^2(t) \quad (n \to \infty)$$

Point estimator,

$$\widehat{cp} = \frac{1}{n} argmax_{1 \leqslant k \leqslant n} M_n$$

Change point in the covariance matrix

- Let $\{Y_t : t \epsilon \mathbb{Z}\}$ be a sequence of d- dimensional random vectors
- Null hypothesis is
$$H_0 : Cov(Y_1) = \cdots = Cov(Y_n)$$
against the hypothesis of at least one change.

Aue et al. (2009)

CUSUM process,

$$S_k = \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{k} vech[\tilde{Y}_t \tilde{Y}_t'] - \frac{k}{n} \sum_{t=1}^{n} vech[\tilde{Y}_t \tilde{Y}_t'] \right),$$

where $\tilde{Y}_t = Y_t - \bar{Y}_t$.

Change point in the covariance matrix

Test statistic is given by,

$$C_n = \max_{1 \leqslant k \leqslant n} S_k' \widehat{\Omega}_n^{-1} S_k.$$

Asymptotic behaviour under H_0,

$$C_n \Rightarrow \sup_{0 \leqslant t \leqslant 1} \sum_{l=1}^{d(d+1)/2} B_l^2(t) \quad (n \to \infty).$$

Point estimator,

$$\widehat{cp} = \frac{1}{n} argmax_{1 \leqslant k \leqslant n} C_n$$

long-run covariance estimation

$$\Omega = \sum_{t \epsilon \mathbb{Z}} Cov\left([Y_0 Y_0'], [Y_t Y_t']\right)$$

$$\widehat{\Omega}_n \to \Omega \quad (n \to \infty)$$

Two general approaches for the non parametric estimation of $\widehat{\Omega}_n$.

- bootstrap [Wied et al. (2014)]
- kernel based

Kernel based estimation

Non-parametric Newey-West estimator [Newey et al. (1989)],

$$\widehat{\Omega}_n = \widehat{\Sigma}_0 + \sum_{h=1}^{H} k\left(\frac{h}{H+1}\right)\left(\widehat{\Sigma}_h + \widehat{\Sigma}_h'\right)$$

- $k(.)$, the Bartlett kernel, where,

$$k(x) = \begin{cases} 1 - |x|, & \text{for } |x| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

- Σ, the covariance matrix
- H, the bandwidth $H = \lfloor n^{1/3} \rfloor$ (minimize overestimation)

Segmentation algorithms

Address the issue of multiple change points detection.

- Class of Binary segmentation algorithms:
  - first, search for a single change point in the whole sequence.
  - then the sequence is split in two subsequences until no more points are detected.

- Our approach combines:
  - ICSS algorithm [Inclan et al. (1994)].
  - BS algorithm [Vostrikova et al. (1981)].
  - the CUSUM test statistics previously presented.

Algorithm steps

1. Obtain the test statistic $ST_{1:N}$.
   - If $ST_{1:N} > CV$, cp_1 the c.p estimation, divide the original sequence into two subsequences and proceed to the Step 2.
   - If $ST_{1:N} \leqslant CV$, no change points are detected.
2. For each subsequence, detect a change (Step 1) and continue the process until no more changes are found.
3. $L = (cp_0, \cdots, cp_{s+1})$, where $cp_0 = 1, cp_s{+}1 = N$ and $cp_1, \cdots, cp_s$ are the c.p previous detected in increasing order. Obtain the test statistic in the intervals (cp_{i-1}, cp_{i+1}). Eliminate the corresponding point if its maximum is not significant.
4. Vector $L = (cp_1 + 1, \cdots, cp_s + 1)$ are the points of change.

On-line test in the unconditional mean

Assumption: $\mu_1 = \cdots = \mu_m$.

$$H_0 : \mu_{m+1} = \mu_{m+2} = \cdots$$

$$H_1 : \mu_{m+1} = \cdots = \mu_{m+k^*-1} \neq \mu_{m+k^*} = \mu_{m+k*+1} = \cdots$$

where $1 \leqslant k^* < \infty$ denotes the unknown time of change.
m: the length of the training period (model is assumed to be stable).

$$g(m,k) = \sqrt{m}\,(1 + k/m)\,(k/k+m)^{\gamma}\,, \gamma \in [0, 1/2)$$

Stefan Fremdt. (2014)

CUSUM detector is given by,

$$Q(m,k) = \sum_{i=m+1}^{m+k} Y_i - \frac{k}{m}\sum_{i=1}^{m} Y_i$$

Stopping time,

$$\tau_m = min\{k \geq 1 : Q(m,k) \geq \widehat{\omega}_m c_{1,a}\, g(m,k)\}$$

Under H_0,

$$= P(\sup_{0<t<1} \frac{W(t)}{t^{\gamma}} > c_{1,a}) = a$$

On-line test in the variance

Pape et al. (2016)

Test statistic is given by,

$$V(m,k) = k\widehat{D}^{-1/2}\left(Var|_{m+1}^{m+k} - Var|_1^m\right)$$

$\widehat{D}$ is an estimator necessary for deriving the asymptotic distribution.

$$\tau_m = \min\{k \leqslant \lfloor mB \rfloor : \|V_k\|_2 > c_a g(m,k)\}$$

B constant that denotes the detection time.
Under the null,

$$\frac{\|V_k\|_2}{g(m,k)} \to \sup_{t\in[0,1]} \left(\frac{B}{1+B}\right)^{1/2-\gamma} \frac{\sqrt{\sum_{l=1}^{d}[W_l(t)]^2}}{t^\gamma}$$

Critical values estimation

Crucial issue for the application. Two general methods:

- Approximation of the quantiles of the limit distributions by simulating Brownian motions or bridges on a fine grid.

alternatively,

- Generation of standard normal distributed random variables and application of the test statistics to the sample.

Direction of changes detection

- Traditional detection methods do not consider the direction of changes.
- We calculate an indicator to solve the problem.

The Moving Average Convergence Divergence (MACD) indicator :

- describes the direction of the trend of the data.
- does not need preliminary learning phase.

although,

- is a trend follower
- does not describe the strength of the trend.

MACD

MACD

The MACD at the observation x associated to the moving lengths N_{short} and N_{long} is defined as,

$$MACD_{N_{\text{short}},N_{\text{long}}}(x) = EMA_{N_{\text{short}}}(x) - EMA_{N_{\text{long}}}(x).$$

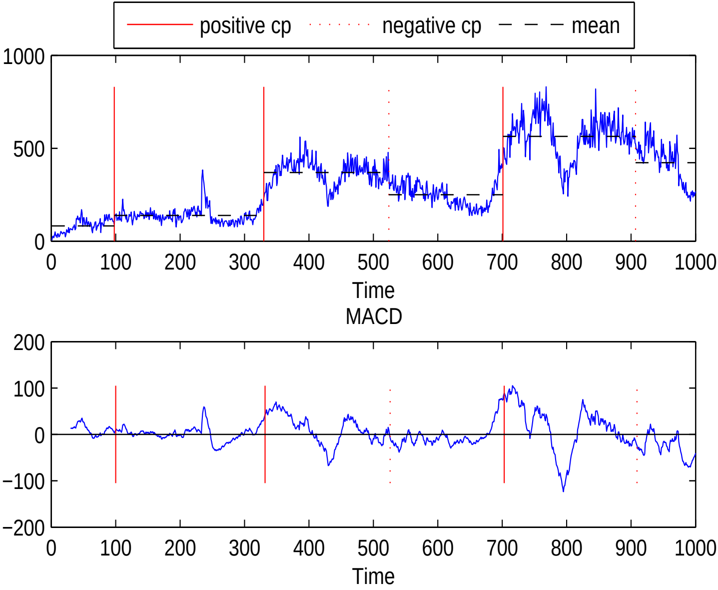
and the corresponding histogram as,

$$h(x) = MACD(x) - EMA_{N_{\text{short-1}}}(MACD(x))$$

- large N produces an output sensitive to slow variations.
- small N results in an output sensitive to fast variations.

EMA

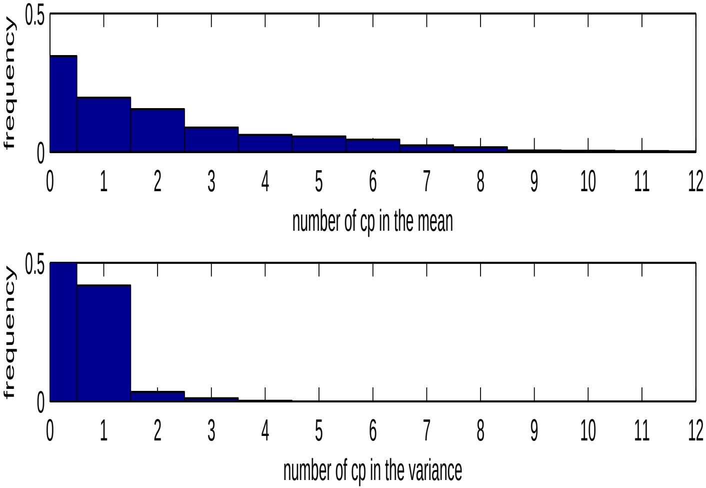
$$y_n = EMA_N(x)_n = \frac{2}{N+1}x_n - \frac{N-1}{N+1}y_{n-1}.$$


positive cp
negative cp
mean
1000
500
0
0
100
200
300
400
500
600
700
800
900
1000
Time
MACD
200
100
0
−100
−200
0
100
200
300
400
500
600
700
800
900
1000
Time

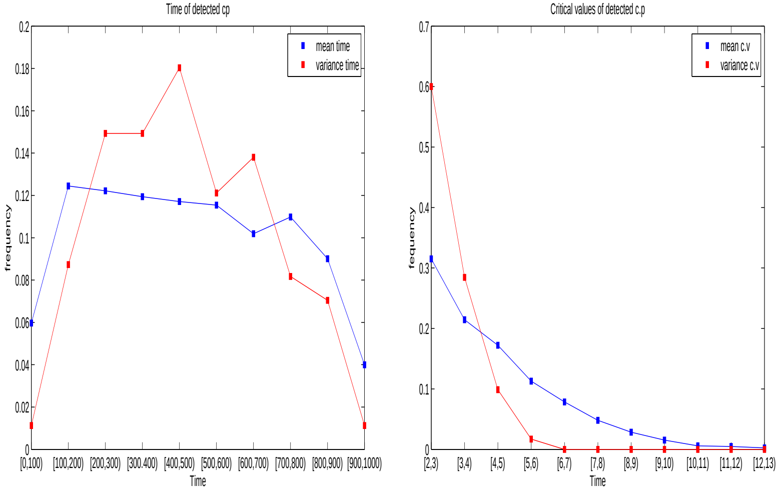
Results

Comparative histograms

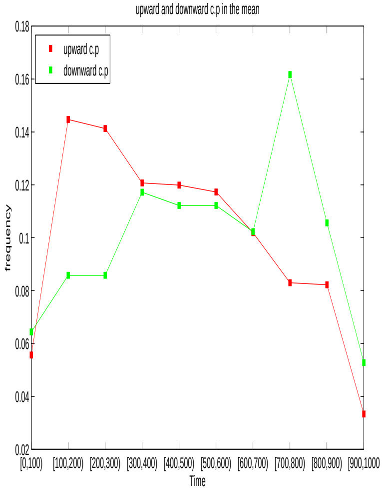
Number of cp

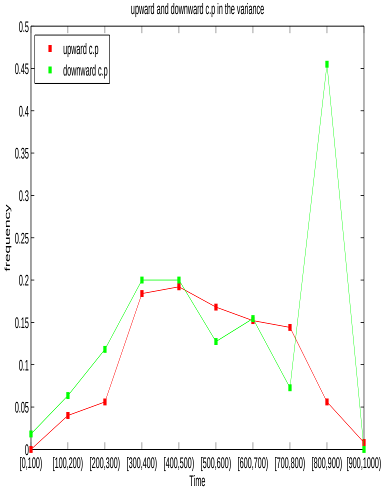


Comparative histograms



Comparative histograms





Successful identifications of upward changes in the mean and the variance structure.

step	**mean**		**variance**	
	critical value		critical value	
	2.4	3	2.4	3
0	0.5938	0.6378	0.8764	0.9310
3	0.8310	0.864	0.9441	1
7	0.8808	0.9026	0.9141	1

On-line detection in the mean structure.

		step=0	
	m=80	m=100	m=150
γ=0	0.9 (22)	0.89 (13)	0.9 (10)
γ=0.25	0.88 (27)	0.87 (14)	0.88 (11)
γ=0.45	0.88 (26)	0.89 (18)	0.87 (12)
		step=5	
	m=80	m=100	m=150
γ=0	0.89 (18)	0.9 (12)	0.9 (8)
γ=0.25	0.9 (22)	0.89 (13)	0.9 (9)
γ=0.45	0.88 (19)	0.88 (16)	0.89 (9)

Conclusions

Conclusions

- CUSUM detector is indeed useful for traffic load balancing.
- Variance approach appears less complicate and more accurate.
- The trend indicator is highly accurate, especially for variance changes.
- The online approach was validated.
  - especially for $\gamma = 0.25$ and $\gamma = 0.45$
- Our next steps:
  - Multi-variate approach with the existing methodology (e.g., multiple videos per unikernel).
  - Real experiments and comparisons.

Acknowledgements

This work is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 644399 (MONROE) through the open call project MEC. The views expressed are solely those of the author(s).

Thank You!