LIGHT

Hoang Vu Nguyen
Jilles Vreeken

UNIVERSITÄT DES SAARLANDES
M2CI CLUSTER OF EXCELLENCE
max planck institut informatik

Linear-time Detection of Non-linear Changes in Massively High Dimensional Time Series

Hoang Vu Nguyen
Jilles Vreeken

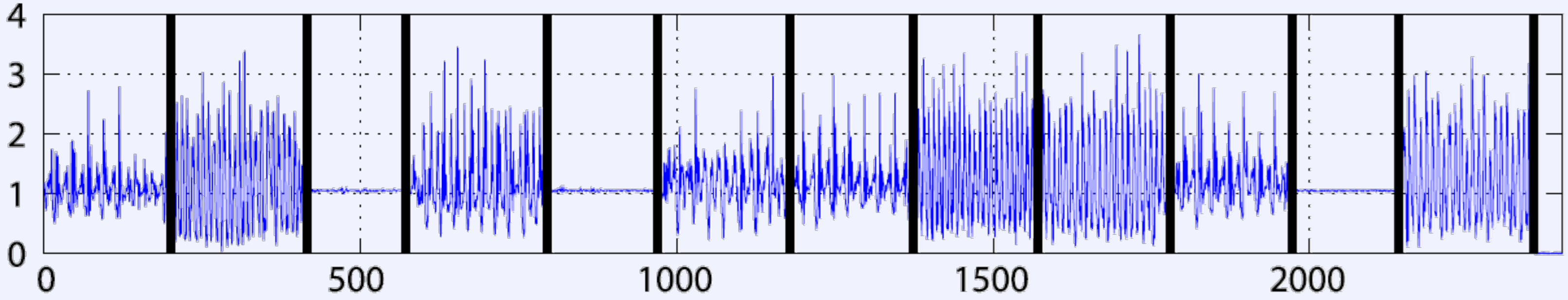
UNIVERSITÄT DES SAARLANDES
CLUSTER OF EXCELLENCE
max planck institut informatik

Question of the day

Suppose we have a time series, of, say **50 000** dimensions.

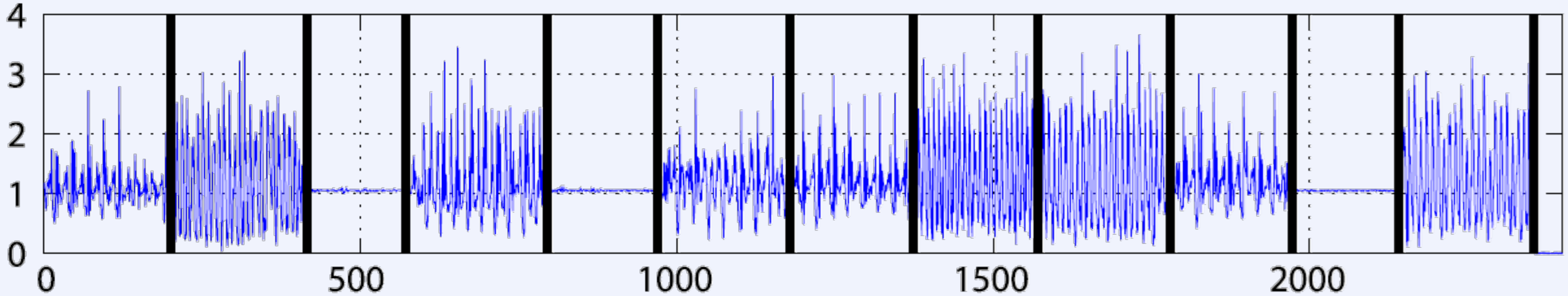
How can we detect **change points** in its distribution?

Change points



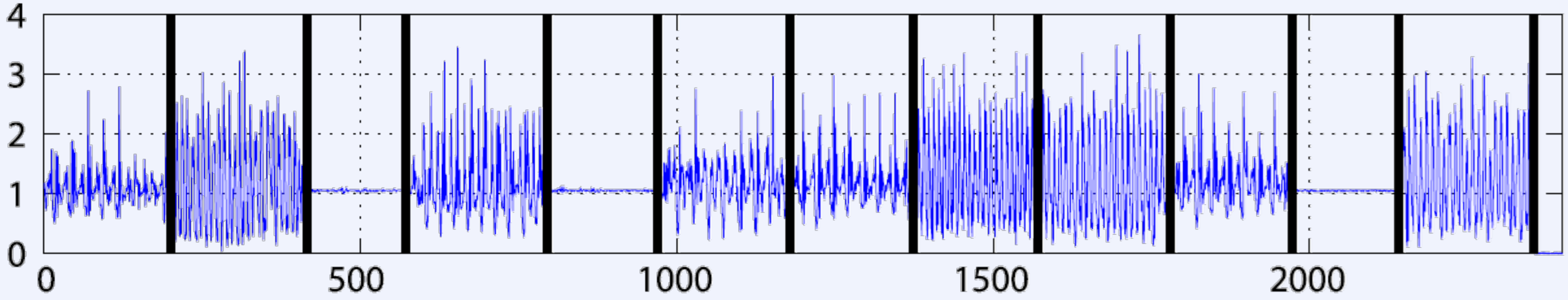
Points in time when **important statistical properties** change

Our goal



We aim to do this for time series with **50k dimensions** while ensuring both **quality** and **efficiency**

Common strategy



Simply sweep a **test window** over the data, measure **divergence** of its distribution against a **reference window**

Curse of Dimensionality

Directly considering **full joint distributions** to compute divergence does not work

- distribution unknown
- estimation requires large samples
  - especially for high dimensional data

Naively, we can simply use **very large windows**

- this has many undesired effects
- high delay, missed alarms, and high runtime

(Desorby et al, '05; Harchaoui et al '08; Kawahara et al '09)

Principally

Instead, we can consider **lower dimensional spaces**
e.g. through Principal Component Analysis (PCA)

- measure divergence over only the projected space

Much better, but does not solve our problem

- covariance estimation also requires **large windows**
- PCA is **cubic** in number of dimensions

Lower dimensional is often still **high** dimensional

- $100 \ll 50\,000$, but 100 dimensions are still challenging

(Kuncheva & Faithfull '14; Qahtan et al, '15)

LIGHT

We propose LIGHT

- LInear-time change detection in hiGH dimensional Time series

In short, LIGHT

- performs scalable PCA to reduce dimensionality
- factorises joint distribution in PCA space for efficient computation
- scales linearly in both data size and dimensionality

Let there be LIGHT

Consider a **reference window** $\mathcal{W}_{ref}$ of m instances

- we use PCA to map to space $\mathbb{S}$ of $k \ll n$ dimensions
- we then work with transformed window $\mathcal{W}'_{ref}$

In particular, we map **test windows** $\mathcal{W}_{test}$ to $\mathbb{S}$

- and consider the difference between $\mathcal{W}'_{ref}$ and $\mathcal{W}'_{test}$

Scalable PCA

We want to do PCA, but…

- robustly, and not at $O(n^2)$

We use matrix sampling for fast and reliable PCA

- consider data of $\mathcal{W}_{ref}$ as matrix A
- we sample with replacement $c \ll n$ columns, according to relative variance, and obtain matrix $C \in \mathbb{R}^{m \times c}$
- we perform PCA on $C^T C$ to compute $k \leq c$ eigenvectors of C, such that ~95% of the variance of $\mathcal{W}_{ref}$ is maintained

Eigenvectors of C approximate those of A

- error bounds that hold with high probability
- $O(mc^2 + mnk)$ instead of $O(mn^2 + n^3)$

(Drineas et al, SIAM J. Computing, 2006)

Factorising the distribution

We do not want to consider the full joint over 100 dimensions...

Instead, we consider a **factorised** distribution

- graphical model $G = (V, E)$ over the k dimensions of $\mathbb{S}$ representing the joint by 1d and 2d distributions

$$p(Y_1, \dots, Y_k) = \frac{\prod_{Y_i, Y_j \in E} p(Y_i, Y_j)}{\prod_{Y \in V} p(Y)^{\deg(Y)-1}}$$

We obtain G by

1. computing all pairwise correlations
2. initialising G with all pairwise edges
3. and simplifying G to one of its maximum spanning trees

Non-linear changes

We want **non-linear** change detection

- so, we need a **non-linear correlation measure**

Quadratic measure of dependency

$$corr(Y_i, Y_j) = \int\int \left(P(y_i, y_j) - P(y_i)P(y_j)\right)^2 dy_i dy_j$$

- uses cumulative distribution functions
- permits computation in **closed form on empirical data**
- for all dimension pairs would take $O(m^2k^2)$
- we can **estimate** unbiased and error-bounded in $O(mk^2)$

(Seth et al, IEEE TSP 2011)

Measuring divergence

Next, we need to determine the of $\mathcal{W}'_{test}$ to $\mathcal{W}'_{ref}$

We want to do so **efficiently** and **non-parametrically**

- quadratic measure of divergence
  - leveraging our factorisation
- initial cost $O(m^2k)$, update cost $O(mk)$

To detect changes we use an adaptive threshold

- Page-Hinkley test

(Nguyen et al. DAMI 2014; Qahtan et al. KDD 2015)

LIGHT

In sum, LIGHT

- performs scalable PCA to reduce dimensionality, and
- factorises joint distribution in PCA space for efficient computation
- scales linearly in both data size and dimensionality

Complexity analysis

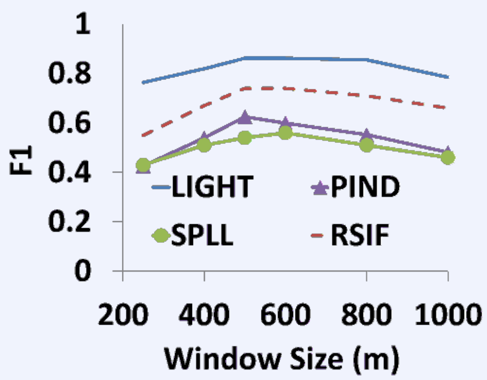
- for time series with r changes, and $m = O(n)$ we have

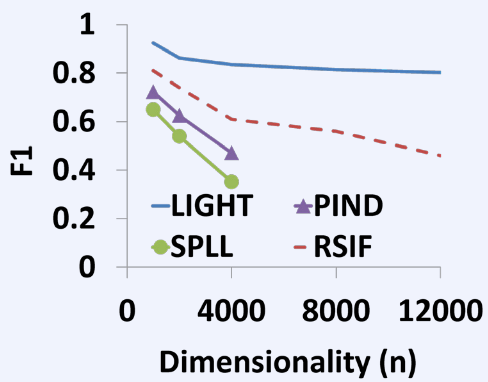
$$O((c^2 + nk)mr + (M - r)mk)$$

code available at eda.mmci.uni-saarland.de/light

Experiments

Experiments show that LIGHT outperforms the state of the art in both quality and efficiency.

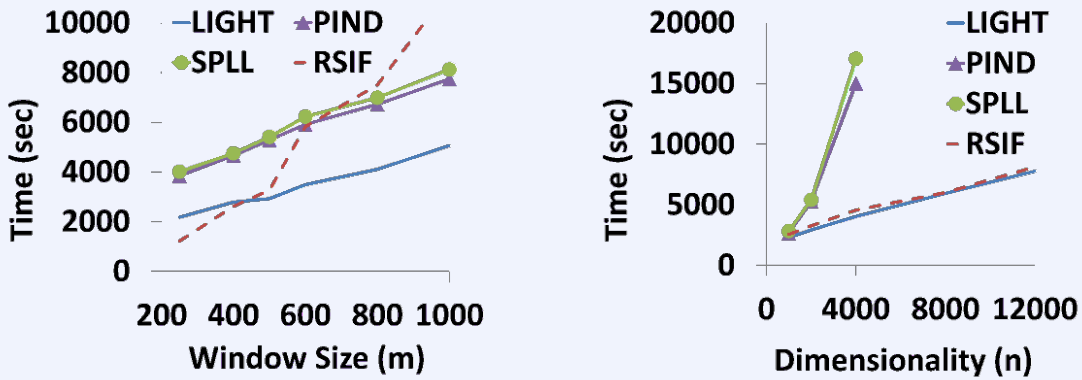




$(n = 2000,\ m = 500,\ c = 200,\ v = 90\%,\ s_1 = 50,\ s_2 = 3)$

Experiments

Experiments show that LIGHT outperforms the state of the art in both quality and efficiency.



$(n = 2000, m = 500, c = 200, v = 90\%, s_1 = 50, s_2 = 3)$

Experiments

Experiments show that LIGHT outperforms the state of the art in both **quality** and **efficiency**.

Data	Dimensionality
Amazon	20 000
EMG1	3 000
EMG2	2 500
Sport	5 625
Youtube	50 000
Average	

$(m = 100,\ c = 50,\ v = 90\%,\ s_1 = 50,\ s_2 = 3)$

Experiments

Experiments show that LIGHT outperforms the state of the art in both quality and efficiency.

F1 score

Data	Dimensionality	**LIGHT**	**PIND**	**SPLL**	**RSIF**
Amazon	20 000	**0.91**	-	-	0.64
EMG1	3 000	**0.77**	0.48	0.45	0.72
EMG2	2 500	**0.84**	0.41	0.44	0.67
Sport	5 625	**0.94**	0.51	0.46	0.84
Youtube	50 000	**0.93**	-	-	0.76
Average		**0.87**	0.54	0.50	0.72

$(m = 100, c = 50, v = 90\%, s_1 = 50, s_2 = 3)$

Experiments

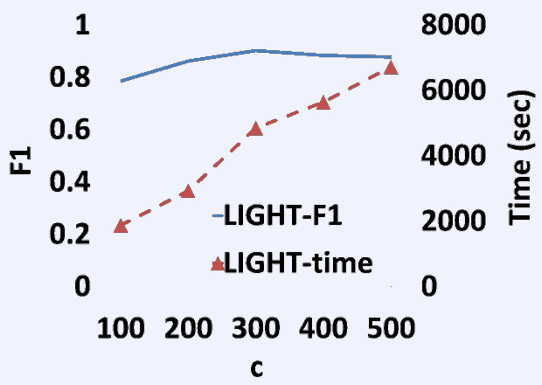
Experiments show that LIGHT outperforms the state of the art in both quality and efficiency.

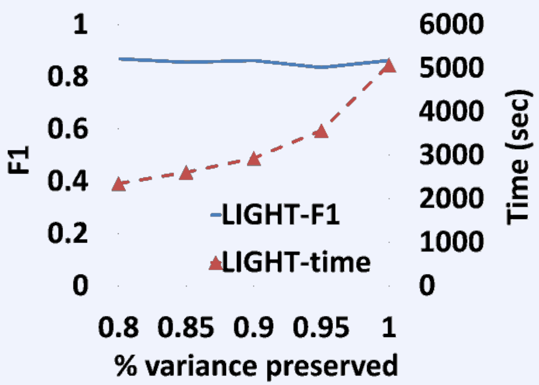
		F1 score				Runtime			
Data	Dimensionality	**LIGHT**	**PIND**	**SPLL**	**RSIF**	**LIGHT**	**PIND**	**SPLL**	**RSIF**
Amazon	20 000	**0.91**	-	-	0.64	**1273.6**	∞	∞	1944.5
EMG1	3 000	**0.77**	0.48	0.45	0.72	**1.2**	92.6	98.1	3.1
EMG2	2 500	**0.84**	0.41	0.44	0.67	**1.1**	345.7	341.5	2.3
Sport	5 625	**0.94**	0.51	0.46	0.84	**5.6**	1295.7	1280.4	11.9
Youtube	50 000	**0.93**	-	-	0.76	**4863.5**	∞	∞	7338.4
Average		**0.87**	0.54	0.50	0.72	**878.6**	∞	∞	1329.5

$(m = 100,\ c = 50,\ v = 90\%,\ s_1 = 50,\ s_2 = 3)$

Experiments

Experiments show that LIGHT is very **robust** with regard to parameter settings.





Conclusions

We studied LInear-time detection of
non-linear changes in hiGH dimensional Time series

In short, LIGHT

- performs scalable PCA, factorises the joint distribution
- efficient, non-parametric, non-linear
- scales linearly in both data size and dimensionality
- permits incremental calculation

Future work

- SLIGHT, for detecting non-linear changes in streaming data

(source code available at: eda.mmci.uni-saarland.de/light)

Thank you!

We studied LInear-time detection of
non-linear changes in hiGH dimensional Time series

In short, LIGHT

- performs scalable PCA, factorises the joint distribution
- efficient, non-parametric, non-linear
- scales linearly in both data size and dimensionality
- permits incremental calculation

Future work

- SLIGHT, for detecting non-linear changes in streaming data

(source code available at: eda.mmci.uni-saarland.de/light)