Coding of Non-Stationary Sources as a Foundation for Detecting Change Points and Outliers in Binary Time-Series

Peter Sunehag, Wen Shao, Marcus Hutter

Research School of Information Science and Engineering
Australian National University
Canberra, ACT, 0200, Australia

{peter.sunehag, wen.shao, marcus.hutter}@anu.edu.au

Highlights

The Cranberries

"Life is **changing** everyday, in every possible way."

"You have to be **fast** on your feet and **adaptive** or else a strategy is useless."

Charles de Gaulle

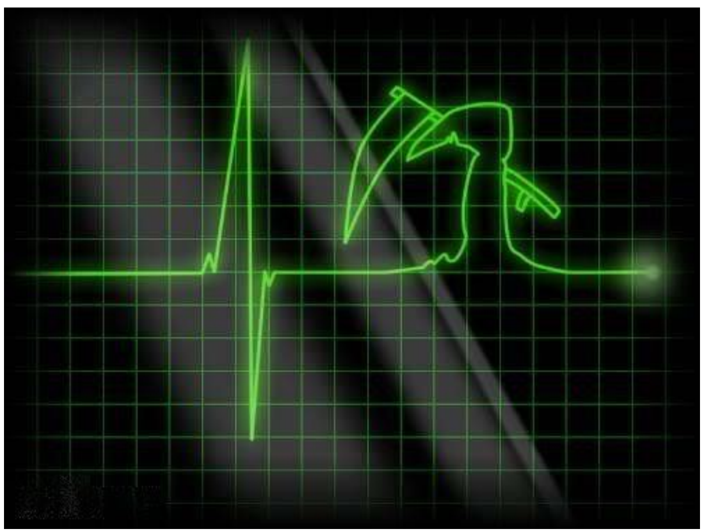
Provided **theoretical justifications** for **adaptive** schemes in **non-stationary** cases as a foundation for detecting change points and outliers in binary time-series and many other applications.

Motiving Application

Detecting **change points** and **outliers** in time series

- Outliers in transactions (fraud)
- Outliers in network access log (criminal and suspicious activities)
- Changes in sales of Wal-Mart Stores (global economics changes)
- Changes in health data (pandemic)

Online and **adaptive to non-stationary data sources**



Preliminaries

KT estimator: a simple method of estimating the probability distribution of the next bit in a (binary) sequence, by assuming a particular Dirichlet prior on the parameter(s) in the Bernoulli distribution.

$$P(x_{n+1} = 1 \mid x_{1:n}) = \frac{\#1 + 1/2}{\#1 + \#0 + 1}$$

Pros

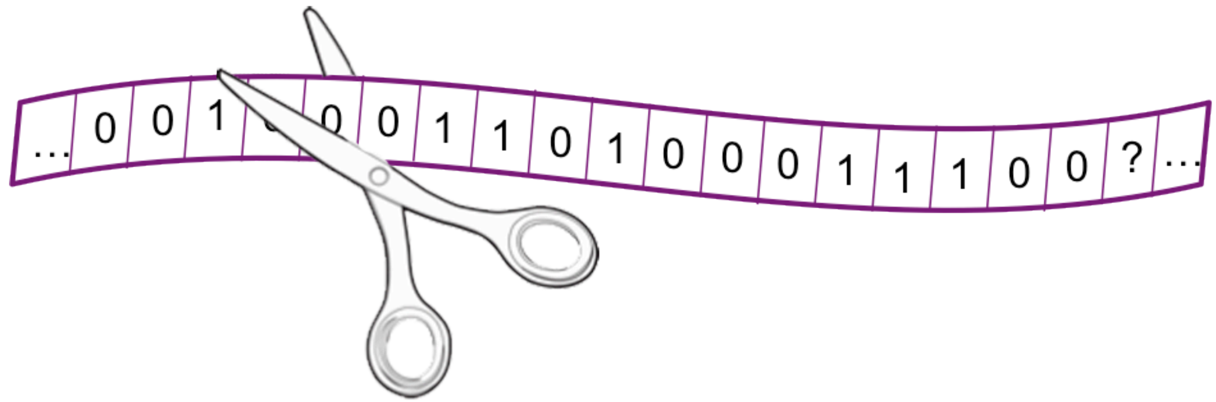
- Simple calculation, only counts of #0 and #1 need to be stored.
- For stationary cases, asymptotically optimal (worst-case redundancy).

Cons

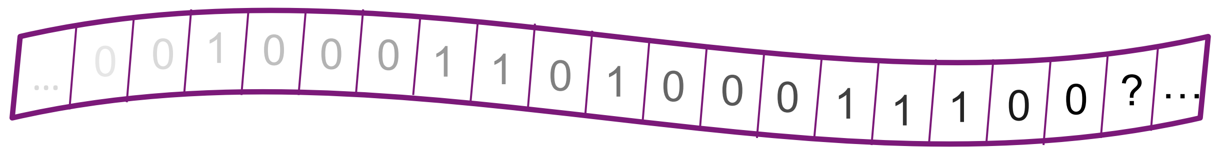
- Non-adaptive, i.e. slow to adapt to a change in the source.

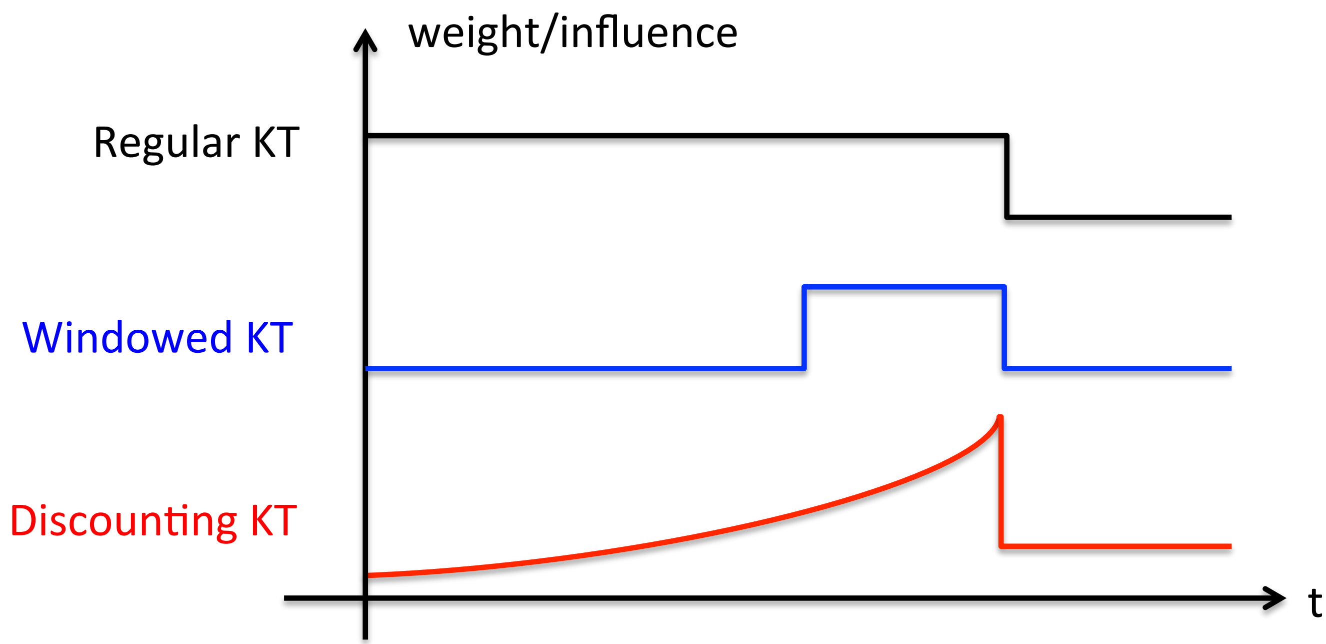
Adaptive Schemes

Windowed KT



Discounting KT





Results Redundancy Bound

General one step prediction redundancy, loosely speaking

$$R_m = E[\textcolor{red}{your\_code\_len}(x_{m+1}|x_{1:m}) - \textcolor{blue}{optimal\_code\_len}(x_{m+1}|x_{1:m})]$$

Theorem

- $x_{1:m}$ *from a non-stationary Bernoulli process with* x_i *being sampled according to* θ_i.
- *with a KT estimator with a moving window of size* $n < m$.
- i *is such that* $m - n + 1 \leq i \leq m + 1$, $\theta_i \in [\theta_L, \theta_R]$, $\theta_L, \theta_R \in (0, 1)$ *and* $\theta_L \leq \theta_R$.

$$R(n) \leq \frac{1 + o(1)}{n} + \max\{KL(\theta_L||\theta_R), KL(\theta_R||\theta_L)\} + \textcolor{red}{O(\frac{1}{n})}$$

Results Redundancy Bound

Problem with the previous bound

Last term **grows unboundedly** when the parameters tend to 0 or 1.

Corollary

- $x_{1:m}$ *from a non-stationary Bernoulli process with* x_i *being sampled according to* $\theta_i \in [L, R]$ $(0 < L \leq R < 1)$.
- *with a KT estimator with a moving window of size* $n < m$.
- i *is such that* $m - n + 1 \leq i \leq m + 1$, $\theta_i \in [\theta_L, \theta_R]$, $\theta_L, \theta_R \in [L, R]$ *and* $\theta_L \leq \theta_R$.

$$R_m(n) \leq \max\{KL(\theta_L || \theta_R), KL(\theta_R || \theta_L)\} + \frac{C}{n}$$

where C *does depend on* L *and* R *but not on* θ_L *or* θ_R.

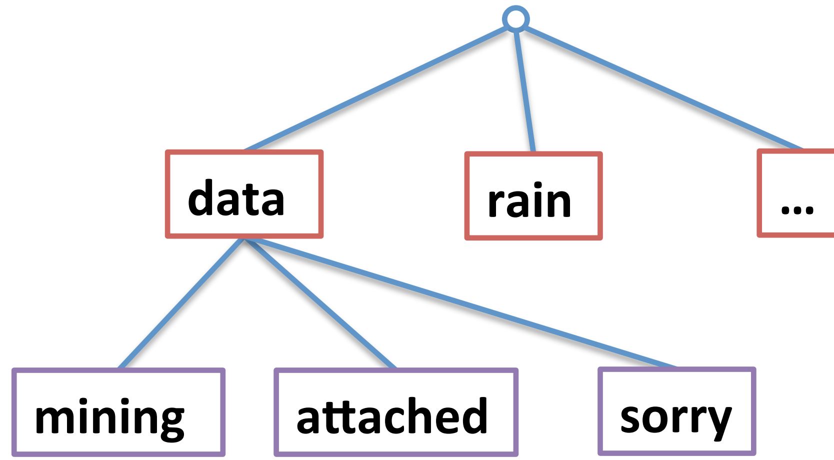
Geometrically drifting sources

Suppose a sequence $x_{1:\infty}$ is generated by a non-stationary Bernoulli process, identified by $\theta_{1:\infty}$ ($\theta_1 \in (0,1)$) with each x_i sampled according to θ_i. We say that the source is geometrically drifting if and only if $1 \le \max\{\frac{\theta_i}{\theta_{i+1}}, \frac{1-\theta_i}{1-\theta_{i+1}}\} \le c$ for all i and some constant $c \ge 1$.

Slowly drifting

Ratio matters; difference not

Application - Compression



Total redundancy for Context Tree Weight algorithm (for sources with finite dependency).

1. Model redundancy (finding the right structure)
2. **Parameter redundancy (KT estimator)**
3. Coding redundancy (Arithmetic coding)

Thank you