# Information Theory SALZA

#### Marion Revolle, François Cayre and Nicolas Le Bihan

GIPSA-Lab | DIS | CICS

April 28<sup>th</sup>, 2017

# Information Theory (without probabilities)

### WORK IN PROGRESS!

Marion Revolle, François Cayre and Nicolas Le Bihan

GIPSA-Lab | DIS | CICS

April 28<sup>th</sup>, 2017

Probabilistic vs. algorithmic Practical issues

## Probabilistic framework (discrete distributions)

#### Entropy

Let  $X = (x, p_X, A)$  a discrete r.v., its entropy reads :

$$H(\mathcal{X}) = -\sum_{i=1}^{|\mathcal{A}|} p_{\mathcal{X}}[x=i] \log_2 p_{\mathcal{X}}[x=i].$$

Probabilistic vs. algorithmic Practical issues

## Probabilistic framework (discrete distributions)

#### Entropy

Let  $X = (x, p_X, A)$  a discrete r.v., its entropy reads :

$$H(\mathcal{X}) = -\sum_{i=1}^{|\mathcal{A}|} p_{\mathcal{X}}[x=i] \log_2 p_{\mathcal{X}}[x=i].$$

#### **Relative entropy**

Also let  $\mathcal{Y} = (y, p_{\mathcal{Y}}, \mathcal{A})$  another discrete r.v.. Provided  $\forall i, p_{\mathcal{Y}}[y = i] \neq 0$ , the relative entropy (KL-divergence) reads :

$$D_{\mathcal{KL}}(\mathcal{X}||\mathcal{Y}) = \sum_{i=1}^{|\mathcal{A}|} p_{\mathcal{X}}[x=i] \log_2 \frac{p_{\mathcal{X}}[x=i]}{p_{\mathcal{Y}}[y=i]}.$$

SALZA Results Overture Probabilistic vs. algorithmic Practical issues

### Issues with probabilities

#### Robust estimation of $p_X, p_Y$

- Need enough data;
- When using parametric distributions, truth may suffer.

SALZA Results Overture Probabilistic vs. algorithmic Practical issues

### Issues with probabilities

#### Robust estimation of $p_X, p_Y$

- Need enough data;
- When using parametric distributions, truth may suffer.

#### The model shadows the data

• Is information only in the model?

Probabilistic vs. algorithmic Practical issues

### Algorithmic framework

Entropy  $H(X) \rightsquigarrow$  Kolmogorov complexity K(x)

Let  $x \in \mathcal{A}^N$ , K(x) is defined as :

"The length of a shortest program to output x on a universal Turing machine".

Probabilistic vs. algorithmic Practical issues

## Algorithmic framework

Entropy  $H(X) \rightsquigarrow$  Kolmogorov complexity K(x)

Let  $x \in \mathcal{A}^N$ , K(x) is defined as :

"The length of a shortest program to output x on a universal Turing machine".

### Relative entropy $D_{KL}(\mathcal{X}||\mathcal{Y}) \rightsquigarrow$ Relative complexity K(x|y)

Also let  $y \in \mathcal{A}^M$ ,  $\mathcal{K}(x|y)$  is defined as :

"The length of a shortest program to output x on a universal Turing machine, when y is known."

Probabilistic vs. algorithmic Practical issues

## Algorithmic framework

Entropy  $H(X) \rightsquigarrow$  Kolmogorov complexity K(x)

Let  $x \in \mathcal{A}^N$ , K(x) is defined as :

"The length of a shortest program to output x on a universal Turing machine".

### Relative entropy $D_{KL}(\mathcal{X}||\mathcal{Y}) \rightsquigarrow$ Relative complexity K(x|y)

Also let  $y \in \mathcal{A}^M$ ,  $\mathcal{K}(x|y)$  is defined as :

"The length of a shortest program to output x on a universal Turing machine, when y is known."

#### But I don't have a universal Turing machine...

You have something quite close. It is called your PC. Or anything with (plenty of) memory and a if branching instruction.

Probabilistic vs. algorithmic Practical issues

## How did it all start? Information distances!

Definition (Maximum Information Distance [Bennett et al., 1998])

$$E_1(x,y) = \max\{K(x|y), K(y|x)\}.$$

Will go back to this later.

Probabilistic vs. algorithmic Practical issues

How did it all start? Information distances!

Definition (Maximum Information Distance [Bennett et al., 1998])

$$E_1(x,y) = \max\{K(x|y), K(y|x)\}.$$

Will go back to this later.

Definition (Normalized Information Distance [Li et al., 2004])  $NID(x,y) = \frac{\max\{K(x|y), K(y|x)\}}{\max\{K(x), K(y)\}}.$ 

Probabilistic vs. algorithmic Practical issues

How did it all start? Information distances!

Definition (Maximum Information Distance [Bennett et al., 1998])

$$E_1(x,y) = \max\{K(x|y), K(y|x)\}.$$

Will go back to this later.

Definition (Normalized Information Distance [Li et al., 2004])

$$\mathsf{NID}(x,y) = \frac{\max\{K(x|y), K(y|x)\}}{\max\{K(x), K(y)\}}.$$

#### Applicability

Binary objects of arbitrary sizes.

SALZA Results Overture Probabilistic vs. algorithmic Practical issues

### Issues with this framework

#### Length of "a shortest program"

Not computable on a universal Turing machine.

SALZA Results Overture Probabilistic vs. algorithmic Practical issues

### Issues with this framework

### Length of "a shortest program"

Not computable on a universal Turing machine.

### Approximating K(x)

Length of compressed data (length of decompressor code is constant).

 $K(x) \simeq C(x).$ 

Highly questionable. May work well in practice.

SALZA Results Overture

Probabilistic vs. algorithmic Practical issues

## Issues with this framework

### Length of "a shortest program"

Not computable on a universal Turing machine.

### Approximating K(x)

Length of compressed data (length of decompressor code is constant).

 $K(x) \simeq C(x).$ 

Highly questionable. May work well in practice.

Lacking a pure conditional estimate...

Let *xy* denote the concatenation of two strings *x* and *y*.

 $K(x|y) \simeq C(xy) - C(y).$ 

SALZA Results Overture Probabilistic vs. algorithmic Practical issues

### First practical embodiment

### The intuition behind

Let *z* be another string  $(x, y, z \text{ defined over } \mathcal{A})$ :

• Intuitively : if C(xy) < C(xz) then y is *closer* to x than z;

• Recall approx. :  $K(x|y) \simeq C(xy) - C(y)$ .

SALZA Results Overture Probabilistic vs. algorithmic Practical issues

## First practical embodiment

### The intuition behind

Let z be another string  $(x, y, z \text{ defined over } \mathcal{A})$ :

- Intuitively : if C(xy) < C(xz) then y is *closer* to x than z;
- Recall approx. :  $K(x|y) \simeq C(xy) C(y)$ .

Definition (Normalized Compression Distance [Li et al., 2004])

$$\mathsf{NCD}(x,y) = \frac{C(xy) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}}$$

What people do when they don't want to start from scratch.

Probabilistic vs. algorithmic Practical issues

### Issues when using a real-word compressor

Built-in compressor limitations [Cebrián et al., 2005]

- Length of block in Burrows-Wheeler transform (bzip2);
- Length of sliding window in LZ77 (gzip: 32KiB, lzma: 4GiB).

Probabilistic vs. algorithmic Practical issues

## Issues when using a real-word compressor

Built-in compressor limitations [Cebrián et al., 2005]

- Length of block in Burrows-Wheeler transform (bzip2);
- Length of sliding window in LZ77 (gzip: 32KiB, lzma: 4GiB).

### Computing C(xy) is another approximation

Does not guarantee that *only* data from *y* will be used to encode *x*.

- Even with lzma;
- [Ziv and Merhav, 1993] factorization would be best (below).

Probabilistic vs. algorithmic Practical issues

## Issues when using a real-word compressor

Built-in compressor limitations [Cebrián et al., 2005]

- Length of block in Burrows-Wheeler transform (bzip2);
- Length of sliding window in LZ77 (gzip: 32KiB, lzma: 4GiB).

### Computing C(xy) is another approximation

Does not guarantee that *only* data from *y* will be used to encode *x*.

- Even with lzma;
- [Ziv and Merhav, 1993] factorization would be best (below).

Breaking another dogma : departing from pure data compression

- Limited only by machine specs (CPU, RAM, 32/<u>64</u> bits);
- Much cleaner computations.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

## What did we implement?

Estimates for classical operators

- K(x), K(x|y);
- K(x,y): "length of a shortest program to encode x and y, plus a means to separate the two".

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

## What did we implement?

Estimates for classical operators

- K(x), K(x|y);
- K(x,y): "length of a shortest program to encode x and y, plus a means to separate the two".

#### Universal normalized semi-distance

Compute a semi-distance between arbitrary binary objects.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

## What did we implement?

Estimates for classical operators

- K(x), K(x|y);
- K(x,y): "length of a shortest program to encode x and y, plus a means to separate the two".

#### Universal normalized semi-distance

Compute a semi-distance between arbitrary binary objects.

#### Algorithmic directed information estimates

Enables model-free causality inference.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# What did we implement?

Estimates for classical operators

- K(x), K(x|y);
- K(x,y): "length of a shortest program to encode x and y, plus a means to separate the two".

#### Universal normalized semi-distance

Compute a semi-distance between arbitrary binary objects.

### Algorithmic directed information estimates

Enables model-free causality inference.

### Building on the Lempel-Ziv family with an unbounded buffer Unbounded (up to sizeof(size\_t)) : semi-infinite sliding window.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# What conditional information? [Revolle et al., 2016]

#### Let x and y be two strings



Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# What conditional information? [Revolle et al., 2016]

#### Let's encode x knowing y, LZ style



Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# What conditional information? [Revolle et al., 2016]

#### Definition (Set of allowed references : $\mathcal{R}$ )

Where to draw references (below) from.

### $x|y: \mathcal{R} = \text{past of } y$



• x|y: Usual LZ77 factorization when x = y;

 $\rightarrow$  Estimation of self-complexity.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# What conditional information? [Revolle et al., 2016]

#### Definition (Set of allowed references : $\mathcal{R}$ )

Where to draw references (below) from.

 $x|^{+}y$  :  $\mathcal{R}$  = all of y



• x|y: Usual LZ77 factorization when x = y;

- $x|^{+}y$ : Usual Ziv-Merhav factorization;
  - $\rightarrow$  Estimation of conditional complexity.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# What conditional information? [Revolle et al., 2016]

### Definition (Set of allowed references : $\mathcal{R}$ )

Where to draw references (below) from.

### x-|y: $\mathcal{R}$ = past of both x and y



- x|y: Usual LZ77 factorization when x = y;
- x|\*y : Usual Ziv-Merhav factorization ;
- *x*<sub>-</sub>|*y* : Previously undefined ;
  - $\rightarrow$  Estimation of directed algorithmic information.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# What conditional information? [Revolle et al., 2016]

### Definition (Set of allowed references : $\mathcal{R}$ )

Where to draw references (below) from.

 $x_{-}|^{+}y : \mathcal{R} = \text{past of } x \text{ and all of } y$ 



- x|y: Usual LZ77 factorization when x = y;
- x|\*y : Usual Ziv-Merhav factorization ;
- $x_{-}|y$ : Previously undefined;
- $x_{-}|^{+}y$ : Previously undefined;
  - $\rightarrow$  Estimation of x and y joint complexity.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# What conditional information? [Revolle et al., 2016]

### Definition (Set of allowed references : $\mathcal{R}$ )

Where to draw references (below) from.

#### Collectively referred to as : $x \wr y$



- x|y: Usual LZ77 factorization when x = y;
- x|\*y : Usual Ziv-Merhav factorization ;
- x<sub>-</sub>|y : Previously undefined ;
- $x_{-}|^{+}y$ : Previously undefined;
- $x \wr y$ : Derive generic properties.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

## Generic factorization

Definition (Factorization symbols)

- <u>References</u> : (*I*, *d*)
  - $\rightarrow$  *Copy I*  $\geq$  2 literals from  $\mathcal{R}$ .

Note : d = "distance" in  $\mathcal{R}$  (we do not use it).

Literals : (1, d)

 $\rightarrow$  Output l = 1 literal of value d.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

## Generic factorization

Definition (Factorization symbols)

- References : (I, d)
  - $\rightarrow$  *Copy I*  $\geq$  2 literals from  $\mathcal{R}$ .

Note : d = "distance" in  $\mathcal{R}$  (we do not use it).

Literals : (1, d)

 $\rightarrow$  Output l = 1 literal of value d.

Definition (Generic factorization and  $\mathcal{L}_{x \wr y}$ )

$$x \wr y \rightsquigarrow (I_1, d_1) \dots (I_n, d_n).$$

 $\mathcal{L}_{x \wr y} = \{l_1, \dots, l_n\}$  : set of lengths produced during the factorization.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

### Some more definitions

#### Definition (Set value)

Let  $f : \mathbb{N}^* \to \mathbb{R}$  be a mapping and let S be a finite set of non-zero natural numbers. The image of S by f is defined as :

$$|\mathcal{S}|_f = \sum_{s\in\mathcal{S}} f(s).$$

Note :  $|\mathcal{S}| = |\mathcal{S}|_{\mathbb{1}_{\mathcal{S}}}$  denotes Card( $\mathcal{S}$ ).

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

### Some more definitions

#### Definition (Set value)

Let  $f : \mathbb{N}^* \to \mathbb{R}$  be a mapping and let S be a finite set of non-zero natural numbers. The image of S by f is defined as :

$$|\mathcal{S}|_f = \sum_{s\in\mathcal{S}} f(s).$$

Note :  $|\mathcal{S}| = |\mathcal{S}|_{\mathbb{1}_{\mathcal{S}}}$  denotes Card( $\mathcal{S}$ ).

#### Definition (Admissible function)

A function  $f : \mathbb{N}^* \to [0, 1]$  is said to be admissible iff it is monotonically increasing.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

SALZA conditional complexity estimate [Revolle et al., 2016]

Definition (SALZA conditional complexity estimate)

Let |x| be the length of x. Given an admissible function f, and two non-empty strings  $x \in \mathcal{A}_x$  and  $y \in \mathcal{A}_y$ , the SALZA conditional complexity estimate of x given y, denoted  $S_f(x \wr y)$ , is defined as :

$$S_{f}(x \wr y) = \underbrace{\frac{|\mathcal{L}_{x \wr y}| - 1}{|x|}}_{\substack{\mathcal{Z} \\ \text{Usual} \\ \text{compl.}}}$$
Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

SALZA conditional complexity estimate [Revolle et al., 2016]

### Definition (SALZA conditional complexity estimate)

Let |x| be the length of x. Given an admissible function f, and two non-empty strings  $x \in \mathcal{A}_x$  and  $y \in \mathcal{A}_y$ , the SALZA conditional complexity estimate of x given y, denoted  $S_f(x \wr y)$ , is defined as :

$$S_{f}(x \wr y) = \underbrace{\frac{|\mathcal{L}_{x \wr y}| - 1}{|x|}}_{\mathcal{Z}} \underbrace{\left(1 - \frac{\sum_{\mathcal{L}_{x \wr y}} |f(l) - (|\mathcal{L}_{x \wr y}|_{f} - 1)}{|x|}\right)}_{\mathcal{S}}_{\text{Spreading factor}}.$$

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

SALZA conditional complexity estimate [Revolle et al., 2016]

### Definition (SALZA conditional complexity estimate)

Let |x| be the length of x. Given an admissible function f, and two non-empty strings  $x \in \mathcal{A}_x$  and  $y \in \mathcal{A}_y$ , the SALZA conditional complexity estimate of x given y, denoted  $S_f(x \wr y)$ , is defined as :

$$S_{f}(x \wr y) = \underbrace{\frac{|\mathcal{L}_{x \wr y}| - 1}{|x|}}_{\mathcal{Z}} \underbrace{\left(1 - \frac{\sum_{\mathcal{L}_{x \wr y}} lf(l) - (|\mathcal{L}_{x \wr y}|_{f} - 1)}{|x|}\right)}_{\mathcal{S}}_{S}$$
Spreading factor

Lemma :  $0 \le S_f(x \wr y) < 1$ . Proof : see paper.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# Comparing SALZA discriminative power

### SALZA vs. LZ complexity alone



Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# Choosing an admissible function

#### Hard vs. soft



Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# Effect of the unbounded buffer

#### Constant-quality results : independent of string lengths



Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# Complexity [Revolle et al., 2016]

#### Definition (SALZA complexity of self)

Let  $\mathcal{L}_x = \mathcal{L}_{x|x}$  be the set of lengths produced during a regular LZ77 factorization.

Given an admissible function *f* and a non-empty string  $x \in \mathcal{A}_x$ , the SALZA complexity of *x*, denoted  $S_f(x)$ , is defined as :

$$S_f(x) = S_f(x|x)$$
$$= \left(1 - \frac{\sum_{\mathcal{L}_x} lf(l) - (|\mathcal{L}_x|_f - 1)}{|x|}\right) \frac{|\mathcal{L}_x| - 1}{|x|}$$

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# Joint complexity [Revolle et al., 2016]

### Definition (SALZA joint complexity)

Given an admissible function f, and two non-empty strings  $x \in \mathcal{A}_x$  and  $y \in \mathcal{A}_y$ , the SALZA joint complexity of x and y, denoted  $S_f(x, y)$ , is defined as :

$$S_f(x,y) = S_f(y_-|^*x) + S_f(x) + \log_{|\mathcal{A}_x|}\left(\frac{|x|}{|y|}\right).$$

Note :  $S_f(x, x) = S_f(x)$  because  $S_f(x_-|^+x) = 0$ .

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# Joint complexity (cont'd)

### How does it perform in practice ?

Let  $\varepsilon = |S_f(x, y) - S_f(y, x)|$ .

х	у	$\mathbb{E}[\varepsilon]$	Var [ɛ]	$\min(\epsilon)$	$max(\varepsilon)$
UDHR	UDHR	1.43e-3	1.38e-6	5e-6	7.96e-3
DNA	DNA	1.23e-3	8.11e-7	6e-6	4.98e-3
UDHR	DNA	6.84e-2	2.49e-6	6.28e-2	7.17e-2

Data :

- UDHR : Universal Declaration of Human Rights (various languages);
- DNA : Mitochondrial DNA samples (various mammal species).

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# Normalized semi-distance [Revolle et al., 2016]

Recalling the mother of information distances

$$E_1(x,y) = \max\{K(x|y), K(y|x)\}.$$

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

Normalized semi-distance [Revolle et al., 2016]

Recalling the mother of information distances

$$E_1(x,y) = \max\{K(x|y), K(y|x)\}.$$

### Definition (Normalized SALZA semi-distance)

Given an admissible function *f*, and two non-empty strings  $x \in \mathcal{A}_x$  and  $y \in \mathcal{A}_y$ , the normalized SALZA semi-distance, denoted NSD<sub>*f*</sub>, is defined as :

$$\mathsf{NSD}_f(x,y) = \max\left\{S_f(x|^{\scriptscriptstyle +}y), S_f(y|^{\scriptscriptstyle +}x)\right\}.$$

Note : The triangle inequality may be violated. Not observed during simulations.

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

# Algorithmic directed information

#### Local Markov condition on DAGs [Janzing and Schölkopf, 2010]



Let T denote the action of a Turing machine :

$$x_j = T(x_a, \dots, x_b, s_j)$$

Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

Causal algorithmic directed information [Revolle et al., 2016]

#### Definition (Causal directed information : online applications)



Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

Causal algorithmic directed information [Revolle et al., 2016]

#### Definition (Causal directed information : online applications)



Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

Causal algorithmic directed information [Revolle et al., 2016]

#### Definition (Causal directed information : online applications)



Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

Full algorithmic directed information [Revolle et al., 2016]

#### Definition (Full directed information : offline applications)



Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

Full algorithmic directed information [Revolle et al., 2016]

#### Definition (Full directed information : offline applications)



Teaser Redefining conditional information Complexity, joint complexity and semi-distance Directed information

Full algorithmic directed information [Revolle et al., 2016]

#### Definition (Full directed information : offline applications)



Introduction SALZA Results

Clustering Causality inference An experiment in literature

# NCD/gzip: Mitochondrial DNA



Clustering Causality inference An experiment in literature

### SALZA NSD : Mitochondrial DNA



Clustering Causality inference An experiment in literature

# NCD/gzip: Writing systems (UDHR)



Clustering Causality inference An experiment in literature

## SALZA NSD : Writing systems (UDHR)



Clustering Causality inference An experiment in literature

# NCD/gzip : Markov chains ( $|\mathcal{A}| = 64$ )



Clustering Causality inference An experiment in literature

## SALZA NSD : Markov chains ( $|\mathcal{A}| = 64$ )



Clustering Causality inference An experiment in literature

# NCD/gzip: Full books



Clustering Causality inference An experiment in literature

### SALZA NSD : Full books



Introduction SALZA Results

Clustering Causality inference An experiment in literature

# Sample DAG #1







Introduction SALZA Results

Clustering Causality inference An experiment in literature

# Sample DAG #2









Clustering Causality inference An experiment in literature

### Drafts from La Réticence by Jean-Philippe Toussaint

que caner, wille an ongivery pro prolonge to chant de mar le intites al, cut a an ip signals hypons play -----in it a marti Je na rentrai nos à l'hôtel tout de muite ne moir-là, je n'éloionai de marca vara la grande plage do sable qui s'étandait derrière le village sur regeria plusieurs bilosètres. J'avais déjà lafasé le village ferrière noi, et je nin tens ap geomin qui menait See dans les ornières cheerin taly and at - detined risilie clôture tout située, et de claus "baseder de an de de de plage mit. J'entendais la ser bet en male devinitions dans he muit. J'entendais in ser tost in the second to a service and Ex one de ser la plage, j'erieval ner vers under the average of the contract to an pieds rus dans le sable. Je sentais le co the a male will get packing the a mail local in plante de nas plants, en establist de la constitución de la c + + dana-ta nutt, anno autago, puis, tentesent, fa ein un pied dans l'eau marrie . . . . durb land generale d'une vacualitte et le sentia un frissen dans tout le coros, son sant aprelas, s'activant et me montant à la tôte, puis je déposai l'autre pide ed dans l'any of set of the set of second at the contract of the termined of the aling, d, white - The la muit, et j'apercevais les contours de l'île au loin, sur lesquels or ulini ales. chaque passage de la lunière tetait un églairage furitaif J'at s'emp p- 20794 samis 14 on manteou sombre sur la plage, et je ne bougais par les plede Salare شسوار start of P-1dans l'emo et les yeux ouverts dans l'obscurité, et je vis un bateur the section qui glissait immobile ada operante Tother, à l'heatern, qui glissait tri -The disparat bissis townt mi, due, put po diquile soften has be channed whether an Corre illestrate and Rodgen Generation . Infin South and 1. 5-1. 12 activio when along that and meliaci. Pa Quinte das Est do la la la ini ultures of lem que ja cadaglina vilence .

Je ne rentral pas à l'hôtel tout de suite ce soir-là, je n'éloignai vers la grande place de sable cui s'étendait derrière le village sur plusieurs kilomètres. J'avais déjà laissé le village derrière moi, et je longeais le petit chemin de terre qui menait à la plage, évitant çà et là les grandes flaques d'eau amobiles dans les ornières que la lumière de la lune delairaient faiblement. Il y avait un ghamp dans l'obscurité en bordure du chemin, un themp aundoms at direct us protégeait une vieille clôture tout a ablaée, et, continuant de suive le chemin dans levait, je començai bientôt à entendre le bruit de la mer au loin, le nurmure résulier de la mer qui n'apporta peu à peu come un soulagement des sens et de l'esprit. Arrivé sur la place, i'ôtai mes chaussures et mes chaussettes et je continual pleds nus der le suite vers le rivage, me demansures à la main. Je sentais le contact froid et junide du sable sous la plante de mes piefs, le sable nouillé qui pénétraill'entre mes prteils, et je marchais devent mes dans le rent aus la personal vertex pairs and an and a second se et je regardais la ver en face de moi, izzanitiz beine et repesente, tertile que les contours de l'fie de Sesuelo se devinaient au toin. Quelques vagaelettes hésitantes venaient mourir devait noi dans le sable, et. soutevant la lante, #-> je nis-un-pied-dans l'ensiet je sentis un frisson dans tout le corps, mon sang s'activant et me montant à la tête, puis je déposai l'autre pied dans 💪 🛶 Press giaciete et mes pieds peu à peu s'accoutunérent à la température de l'eau. et, eane je se relevais pour regagner l'hôtel, je vis un bateau apparaître ----nicon, un ferry qui glissait instille à leven auface de l'eau, qui glissait have so mus ! wole les lablots et la ar detecte ende et cui sente en dispartifice derrière les contours rocheux de l'île de Sascelo, dent le prore pentiment de tourner even régularité dans la quit. minilation linge. a brand and derived death alg- pageone. and and it des print donale materie ~ appendentes Serpiedo mollos das norm dar le saille

afrend in pieds

Clustering Causality inference An experiment in literature

### Drafts from La Réticence by Jean-Philippe Toussaint

In bullet alle pate dai is the to buis

Je ne rentral cas à l'hôtel tout de suite ce soir-là, je n'éloisnai vers la grande plage de sable qui s'étendait derrière le village sur plusieurs kilonètres. J'avais détà laissé le village derrière moi, et je longeais le petit chemin de terre qui menait à la plage, évitant çà et là les grandes flaques d'eau faiblement éclairées par la lune qui s'étaient formées dans les proières. Il y sonit un chemp dans l'obscurité en bordure du chemin, un champ abandonné et silencieux que protégeait une vieille clôture tout abînée, et, continuant de suivre le chemin désert dans la muit, je commençai bientôt à entendre le bruit de la mer au loin, le nurmure régulier de la mer qui n'apporta peu à peu comme un sculagement des sens et de l'asprit. Arrivé sur la nlare. 1'Stai mes chaussures et nes chaussettes et je n'avancal'aieda mus vers le rivage, Je sentais le contact froid et mente du sable sous la plante de mes pieds, le sable moullé qui pénétrait entre mes orteils , et je marchais dans la nuit vere la rer en enfongant nes pieds à chaque pas davantage dans le sette pour n'imprégner toujours plus de la sensation de bien-être and i ne procural. J'avais finis par a'ansectr au bord de l'eau, et je ne bougeais plus, je regardais la mer en face de moi, J'étais avais là en merteau sorbre as level de l'esu, et je vis in bateau annaraître à l'herizon, un ferry cui elissait lentement devent noi tout illuminé dere la muit, qui glianait involte à la service de l'em et qui finit par discussification de contrato de l'île de Secolo, al legle last and reach to day a mit.



Je ne rentrai pas à l'hôtel tout de suite ce soir-là, je n'éloignai vers la grande plage de sable qui s'étendait derrière le village sur plusieurs kilomètres, J'avais détà laissé le village derrière moi, et je longeais le petit chemin de terre qui menait à la plage, évitant çà et là les grandes flaques d'eau faiblement éclairées par la lune qui s'étaient formées dans les ornières. Il y avait un chem dens l'obscurité en bordure du chemin, un champ abandonné et silencieux que protégeait une vieille clôture tout abfinée, et, continuant de suivre le chemin désert dans la nuit, je commencai bientôt à entendre le bruit de la mer au loin. le numure régulier de la per qui n'anporta peu à peu comme un soulagement des sens et de l'esprit. Arrivé sur la place, i'ôtai mes chaussures et mes chaussettes et le m'avancai lentement dans la muit vers le rivage, les pieds mus et mes chaussures à la main. Je sentais le contact froid du sable sous la plante de mes pieds, le sable humide qui pénétrait entre mes orteils, et j'enfonçais à chaque sons (mes pes) davantage dans le sol pour me pénétrer toujours plus de la sensation de bien-être que ne procurait le contact du sable mouillé sour ses pions. J'avais fini par n'assecir au bord de l'eau, et je ne bougeais plus, je regardais la mer en face de moi. Le phare de l'île de Savaelo tournait avec régularité dans la muit. et tout était ailencieux autour de moi enr-ha-ploge diensie. J'étais assis là tout seul en manteau sombre anter a seble mouillé, et je vis in bateau apparaître attende de sameto, un ferry qui glissait lentement devant moi tout illuminé dans la nuit, qui glissait inmobile à leverteen et qui finit par disparaître en alleren derrière l'île de Sasselo. to streat them an inyest

Clustering Causality inference An experiment in literature

## Drafts clustering : Neighbor-Joining



Introduction SALZA Results

Clustering Causality inference An experiment in literature

### Drafts clustering : UPGMA



Clustering Causality inference An experiment in literature

### Drafts causality inference (full directed information)



Biomedical data Imaging ?

# Sample problem

### Description

Decide between two states (eyes closed/open) based on EEG signals. EEG data exhibits features at known frequencies ( $\alpha$ ,  $\beta$ , etc.) Data : courtesy [Andrzejak et al., 2001].

Biomedical data Imaging ?

# Sample problem

### Description

Decide between two states (eyes closed/open) based on EEG signals. EEG data exhibits features at known frequencies ( $\alpha$ ,  $\beta$ , etc.) Data : courtesy [Andrzejak et al., 2001].

### The usual approach

Let s(t) be a signal,  $s_{\min} \le s(t) \le s_{\max}$ . One computes the Power Spectral Density (PSD) :

$$\mathsf{PSD}_{s}(t) = \int_{-\infty}^{\infty} \mathbb{E}\left[s(t)s(t+ au)
ight] e^{-2i\pi t au} d au$$

Then, feature extraction, etc.

Biomedical data Imaging ?

### Fitting the AIT framework

### Accessing frequency information

Compute successive residuals  $R_f$  in Butterworth filter bank.

Note :  $s_{\min} \leq R_f(t) \leq s_{\max}$ , too.

Biomedical data Imaging ?

# Fitting the AIT framework

### Accessing frequency information

Compute successive residuals  $R_f$  in Butterworth filter bank.

Note :  $s_{\min} \leq R_f(t) \leq s_{\max}$ , too.

### Quantization

Signals (usually) have continuous values. We can only handle *discrete* alphabets ! Compute complexity over bytes :

$$x_f(t) = \operatorname{Rint}\left(255 \times \frac{R_f(t)}{s_{\max} - s_{\min}}\right)$$

.

Note : Many other choices in the literature, sometimes quite involved.
Biomedical data Imaging ?

# Comparing power vs. complexity [dB] of EEG signals



Biomedical data Imaging ?

# Comparing power vs. complexity [dB] of EEG signals



Biomedical data Imaging?

# Some thoughts on AIT for 2D data

### "Copy from the past" in 2D?

Block matching ! Think of various block sizes in H.26x standards.

#### Issue

Handle block residual information.

## Acknowledgements and paper draft

### Transcripts of Jean-Philippe Toussaint's drafts

Profs. Brigitte Combes and Thomas Lebarbé (Université Grenoble Alpes), heads of the project *La Réticence* (supported by the CAHIER consortium, TGIR Huma-Num).

Proof-reading of the paper

Many thanks to Steeve Zozor !

#### Paper draft

http://arxiv.org/abs/1607.05144

### **References** I

Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., and Elger, C. E. (2001).

Indications of Nonlinear Deterministic and Finite-dimensional Structures in Time Series of Brain Electrical Activity : Dependence on Recording Regions and Brain State.

Physical Review E, 64 :061907.



Bennett, C. H., Gacs, P., Li, M., Vitániy, P. M., and Zurek, W. H. (1998). Information Distance.

IEEE Transactions on Information Theory, 44 :1407–1423.



Cebrián, M., Alfonseca, M., and Ortega, A. (2005).

Common Pitfalls Using the Normalized Compression Distance : What to Watch Out for in a Compressor.

Communications in Information and Systems, 5:367-384.

### **References II**

### Janzing, D. and Schölkopf, B. (2010).

Causal Inference Using the Algorithmic Markov Condition.

IEEE Transactions on Information Theory, 56 :5168-5194.

Li, M., Chen, X., Li, X., Ma, B., and Vitányi, P. M. (2004).

#### The Similarity Metric.

IEEE Transactions on Information Theory, 50 :3250-3264.



Revolle, M., Cayre, F., and Le Bihan, N. (2016).

SALZA : Soft Practical Algorithmic Complexity Estimates for Clustering and Causality Inference.

Submitted.



#### Ziv, J. and Merhav, N. (1993).

A Measure of Relative Entropy Between Individual Sequences with Application to Universal Classification.

IEEE Transactions on Information Theory, 39 :1270–1279.