Information Theory

SALZA

Marion Revolle, François Cayre and Nicolas Le Bihan

GIPSA-Lab | DIS | CICS

April 28th, 2017

Information Theory (without probabilities)

WORK IN PROGRESS !

Marion Revolle, François Cayre and Nicolas Le Bihan

GIPSA-Lab | DIS | CICS

April 28th, 2017

Probabilistic framework (discrete distributions)

Entropy

Let $\mathcal{X} = (x, p_{\mathcal{X}}, \mathcal{A})$ 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 framework (discrete distributions)

Entropy

Let $\mathcal{X} = (x, p_{\mathcal{X}}, \mathcal{A})$ 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 :

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

Issues with probabilities

Robust estimation of $p_{\mathcal{X}}, p_{\mathcal{Y}}$

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

Issues with probabilities

Robust estimation of $p_{\mathcal{X}}, p_{\mathcal{Y}}$

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

The model shadows the data

- Is information *only* in the model ?

Algorithmic framework

Entropy $H(\mathcal{X}) \rightsquigarrow$ 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".

Algorithmic framework

Entropy $H(\mathcal{X}) \rightsquigarrow$ 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 $y \in \mathcal{A}^M$, $K(x|y)$ is defined as :
"The length of a shortest program to output x on a universal Turing machine, when y is known."

Algorithmic framework

Entropy $H(\mathcal{X}) \rightsquigarrow$ 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 $y \in \mathcal{A}^M$, $K(x|y)$ 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.

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.

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])

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

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])

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

Applicability

Binary objects of arbitrary sizes.

Issues with this framework

Length of "a *shortest* program"

Not computable on a universal Turing machine.

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.

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.

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

First practical embodiment

The intuition behind

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

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

First practical embodiment

The intuition behind

Let z be another string (x, y, z 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])

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

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

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).

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).

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.

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*".

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.

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.

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

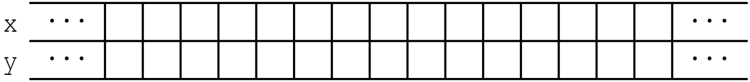
Enables model-free causality inference.

Building on the Lempel-Ziv family *with an unbounded buffer*

Unbounded (up to `sizeof(size_t)`) : semi-infinite sliding window.

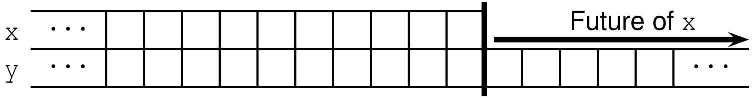
What conditional information ? [Revolle et al., 2016]

Let x and y be two strings



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

Let's encode x knowing y, LZ style

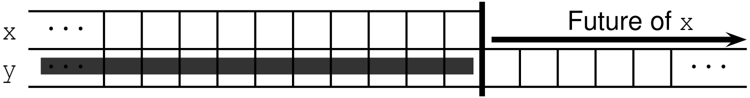


What conditional information ? [Revolle et al., 2016]

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

Where to draw references (below) from.

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



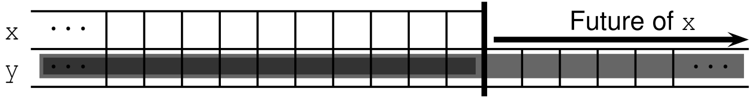
- $x|y$: Usual LZ77 factorization when $x = y$;
  $\rightarrow$ Estimation of self-complexity.

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



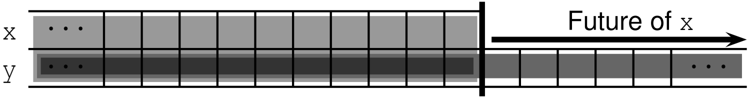
- $x|y$: Usual LZ77 factorization when $x = y$;
- $x|^{+}y$: Usual Ziv-Merhav factorization ;
  $\rightarrow$ Estimation of conditional complexity.

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

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

Where to draw references (below) from.

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



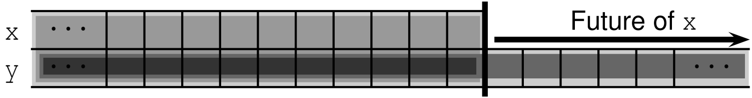
- $x|y$: Usual LZ77 factorization when $x = y$;
- $x|^{+}y$: Usual Ziv-Merhav factorization ;
- $x_{\text{-}}|y$: Previously undefined ;
  $\rightarrow$ Estimation of directed algorithmic information.

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

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

Where to draw references (below) from.

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



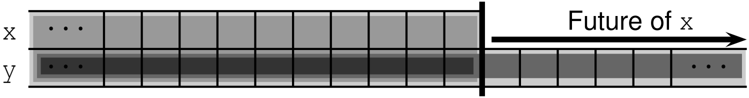
- $x|y$: Usual LZ77 factorization when $x = y$;
- $x|^{+}y$: Usual Ziv-Merhav factorization ;
- $x_{\text{-}}|y$: Previously undefined ;
- $x_{\text{-}}|^{+}y$: Previously undefined ;
  $\rightarrow$ Estimation of x and y joint complexity.

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 ;
- $x_{-}|y$: Previously undefined ;
- $x_{-}|^{+}y$: Previously undefined ;
- $x \wr y$: Derive generic properties.

Generic factorization

Definition (Factorization symbols)

- References : (l, d)
  $\rightarrow$ *Copy* $l \geq 2$ 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.

Generic factorization

Definition (Factorization symbols)

- References : (l, d)
  $\rightarrow$ *Copy* $l \geq 2$ 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 $\mathcal{L}_{x \wr y}$)

$$x \wr y \rightsquigarrow (l_1, d_1) \ldots (l_n, d_n) .$$

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

Some more definitions

Definition (Set value)

Let $f : \mathbb{N}^\star \to \mathbb{R}$ be a mapping and let $\mathcal{S}$ be a finite set of non-zero natural numbers. The image of $\mathcal{S}$ 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 $\mathrm{Card}(\mathcal{S})$.

Some more definitions

Definition (Set value)

Let $f : \mathbb{N}^\star \to \mathbb{R}$ be a mapping and let $\mathcal{S}$ be a finite set of non-zero natural numbers. The image of $\mathcal{S}$ 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 $\mathrm{Card}(\mathcal{S})$.

Definition (Admissible function)

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

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 :

$$S_f(x \wr y) = \underbrace{\frac{|\mathcal{L}_{x \wr y}| - 1}{|x|}}_{\substack{\mathcal{Z} \\ \text{Usual} \\ \text{compl.}}}$$

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 :

$$S_f(x \wr y) = \underbrace{\frac{|\mathcal{L}_{x \wr y}| - 1}{|x|}}_{\mathcal{Z}} \underbrace{\left(1 - \frac{\sum_{\mathcal{L}_{x \wr y}} lf(l) - (|\mathcal{L}_{x \wr y}|_f - 1)}{|x|}\right)}_{\mathcal{S}} .$$

Spreading factor

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

Definition (SALZA conditional complexity estimate)

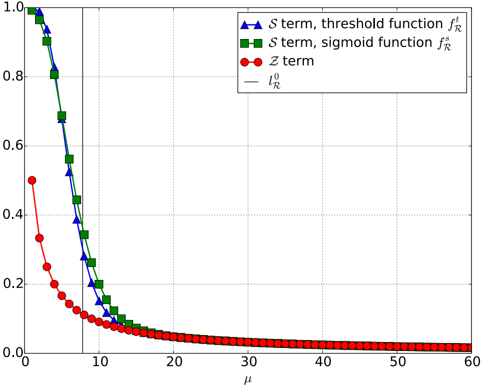
Let $|x|$ be the length of x. Given an admissible function f, and two non-empty strings $x \in \mathcal{A}_x$ and $y \in \mathcal{A}_y$, the SALZA conditional complexity estimate of x given y, denoted $S_f(x \wr y)$, is defined as :

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

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

Comparing SALZA discriminative power

SALZA vs. LZ complexity alone



Min. length for meaningful refs :

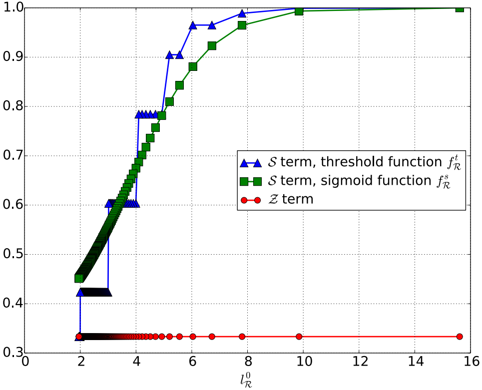
Let $l^0_{\mathcal{R}} = \log_{|\mathcal{A}_{\mathcal{R}}|} |\mathcal{R}|$.

Generate synthetic Poisson-distributed lengths.

$\mu = \mathbb{E}\left[\mathcal{L}_{x \wr y}\right]$.

Choosing an admissible function

Hard vs. soft



Hard thresholding :

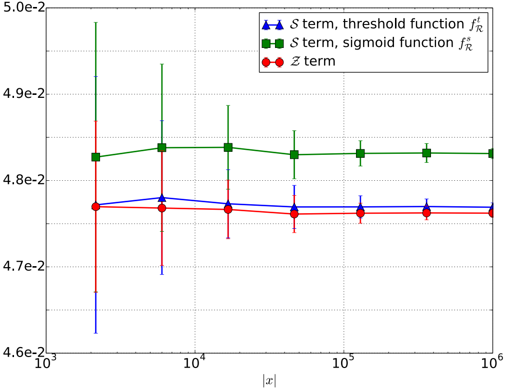
$$f^t_{\mathcal{R}}(l) = \begin{cases} 1 \text{ if } l > l^0_{\mathcal{R}} \\ 0 \text{ otherwise} \end{cases}$$

Soft sigmoid :

$$f^s_{\mathcal{R}}(l) = \frac{1}{1+e^{-l+l^0_{\mathcal{R}}}}$$

Effect of the unbounded buffer

Constant-quality results : independent of string lengths



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

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

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 :

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

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

Joint complexity (cont'd)

How does it perform in practice ?

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

x	y	$\mathbb{E}[\varepsilon]$	$\mathrm{Var}[\varepsilon]$	$\min(\varepsilon)$	$\max(\varepsilon)$
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).

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

Recalling the mother of information distances

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

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)

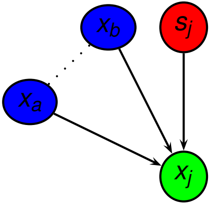
Given an admissible function f, and two non-empty strings $x \in \mathcal{A}_x$ and $y \in \mathcal{A}_y$, the normalized SALZA semi-distance, denoted NSD_f, is defined as :

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

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

Algorithmic directed information

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



Parent strings

Causal mechanism

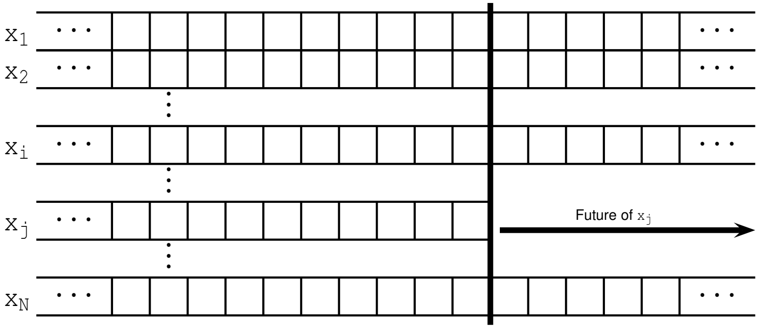
Observed string

Let T denote the action of a Turing machine :

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

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

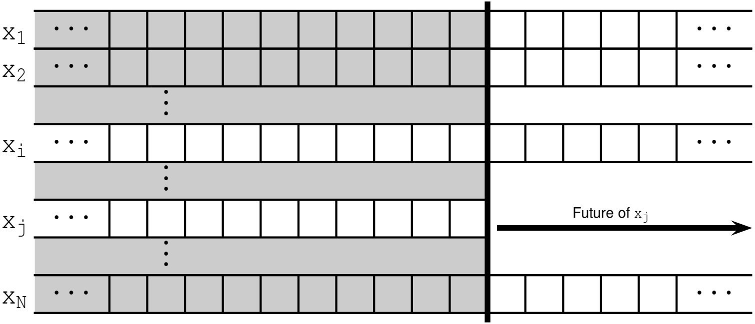
Definition (Causal directed information : online applications)



$$C_S(x_i \to x_j)$$

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

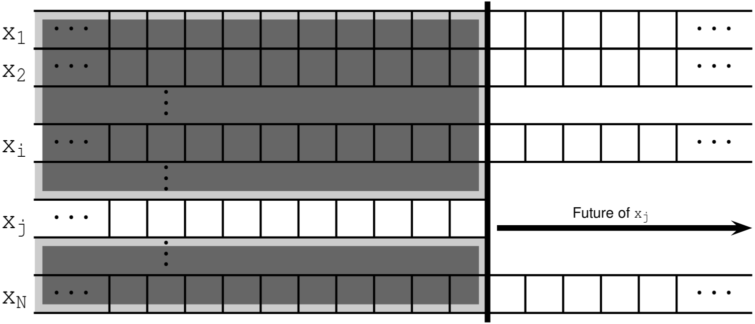
Definition (Causal directed information : online applications)



$$C_S(x_i \to x_j) = S(x_{j\text{-}} | X \setminus \{x_i, x_j\})$$

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

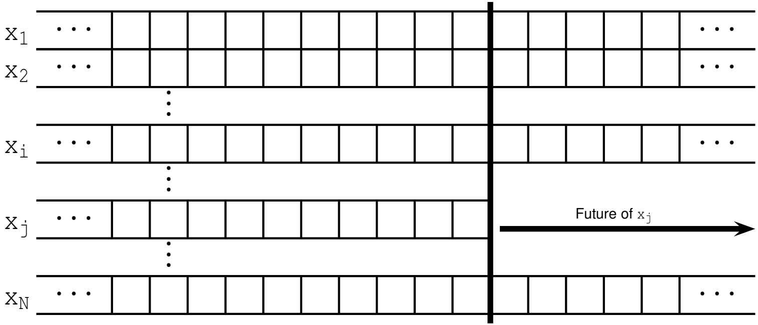
Definition (Causal directed information : online applications)



$$C_S(x_i \to x_j) = S(x_{j\text{-}} | X \backslash \{x_i, x_j\}) - S(x_{j\text{-}} | X \backslash x_j).$$

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

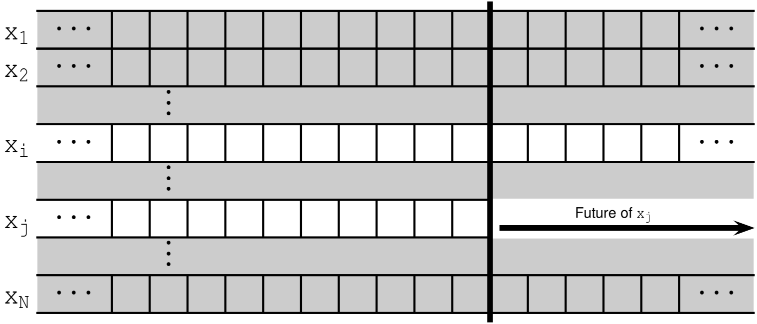
Definition (Full directed information : offline applications)



$$F_S(x_i \to x_j)$$

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

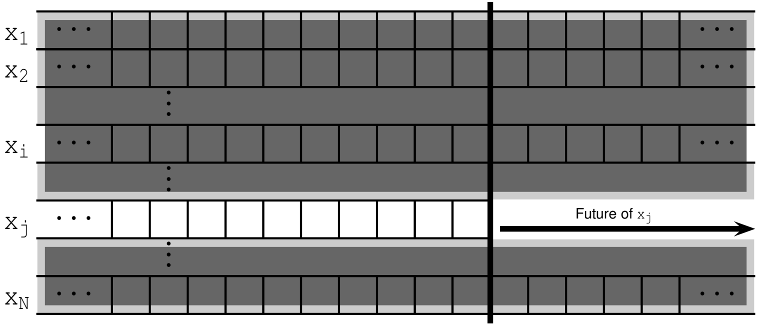
Definition (Full directed information : offline applications)



$$F_S(x_i \to x_j) = S(x_{j_-} |^{+} X \setminus \{x_i, x_j\})$$

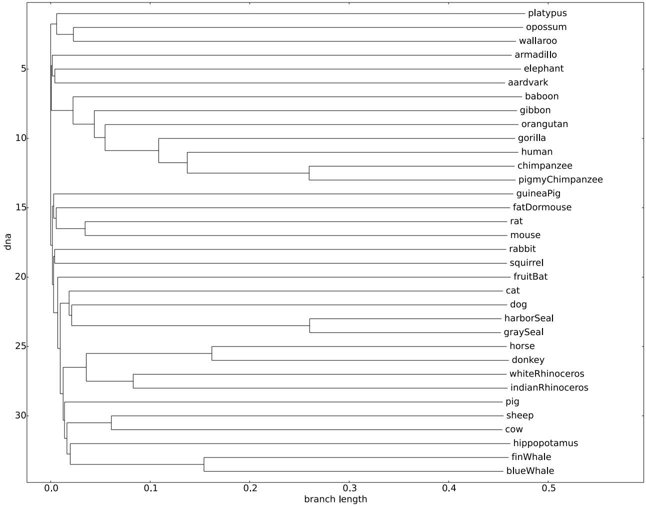
Full algorithmic directed information [Revolle et al., 2016]

Definition (Full directed information : offline applications)

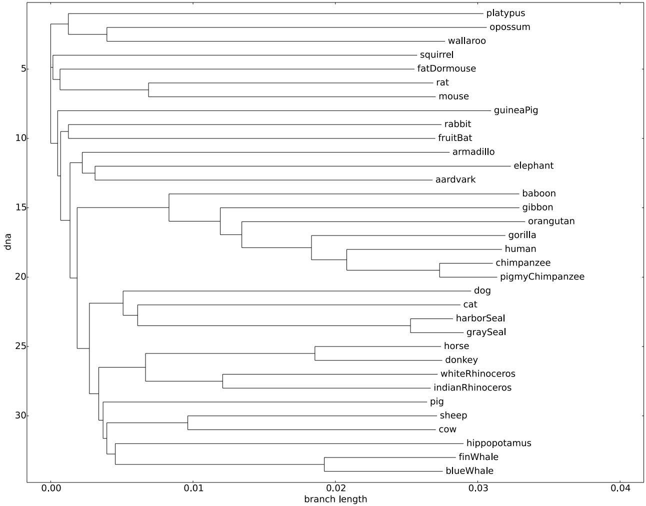


$$F_S(x_i \to x_j) = S(x_{j\text{-}}|^{+} X \backslash \{x_i, x_j\}) - S(x_{j\text{-}}|^{+} X \backslash x_j).$$

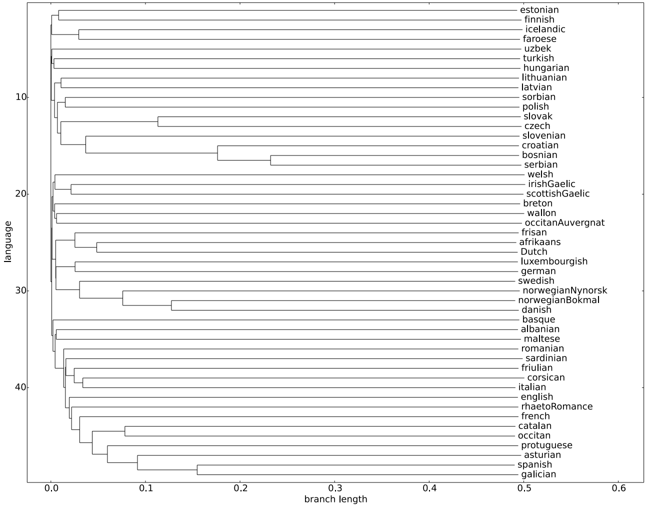
NCD/`gzip` : Mitochondrial DNA



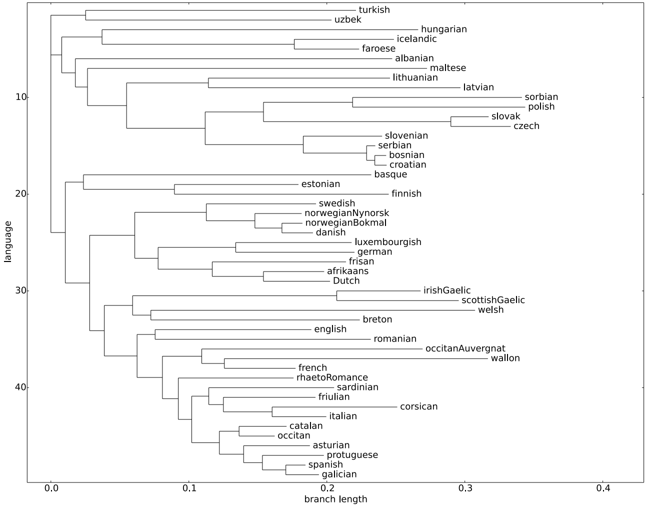
SALZA NSD : Mitochondrial DNA



NCD/`gzip` : Writing systems (UDHR)

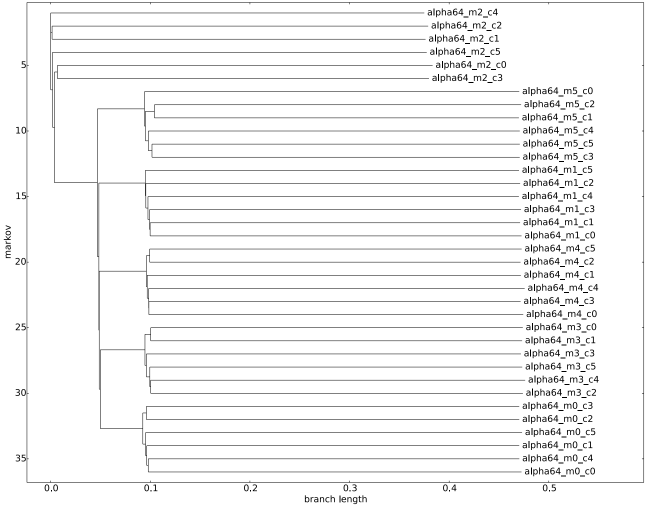


SALZA NSD : Writing systems (UDHR)

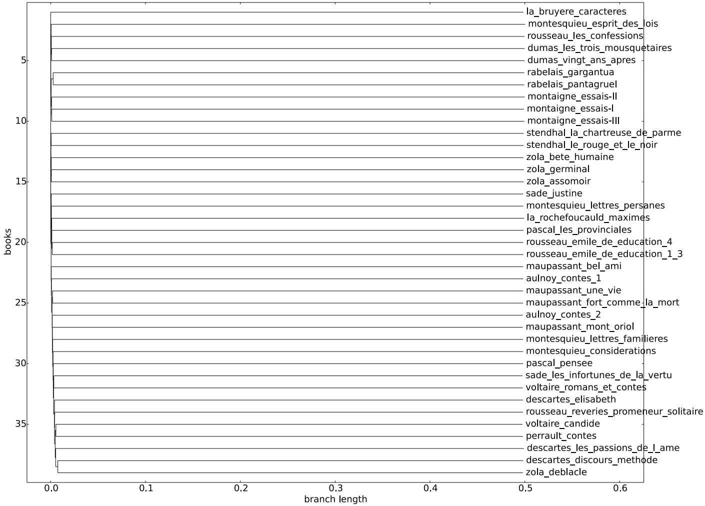


NCD/`gzip` : Markov chains ($|\mathcal{A}| = 64$)

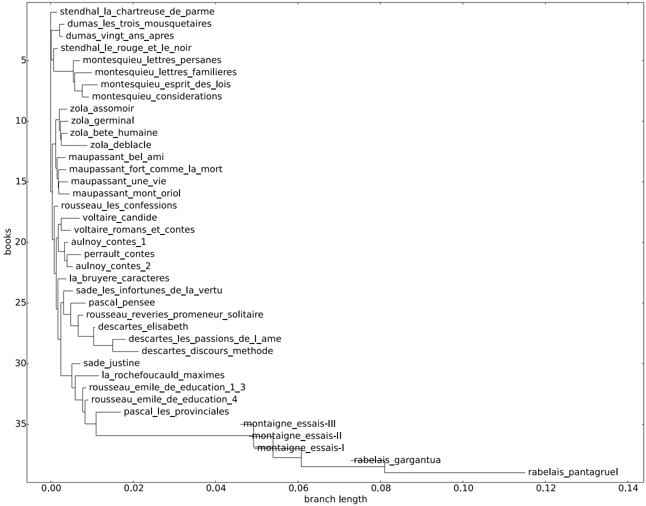
SALZA NSD : Markov chains ($|\mathcal{A}| = 64$)



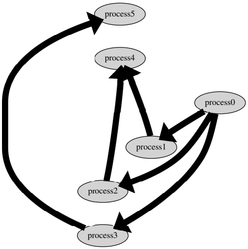
NCD/`gzip` : Full books

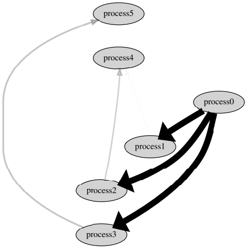


SALZA NSD : Full books



Sample DAG #1

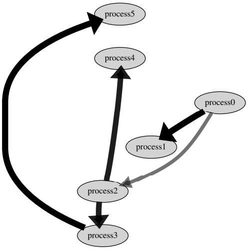


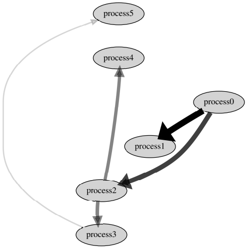


0	0	0	0	0	0	1.0
.50	0	0	0	0	0	.50
.50	0	0	0	0	0	.50
.50	0	0	0	0	0	.50
0	.50	.50	0	0	0	0
0	0	0	.50	0	0	.50

0	0	0	0	0	0
.60	0	0	0	0	0
.60	0	0	0	0	0
.60	0	0	0	0	0
0	.02	.10	0	0	0
0	0	0	.10	0	0

Sample DAG #2

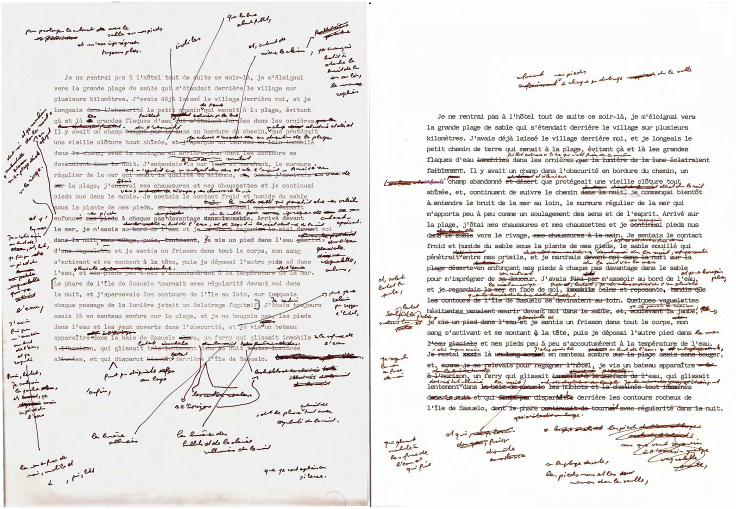




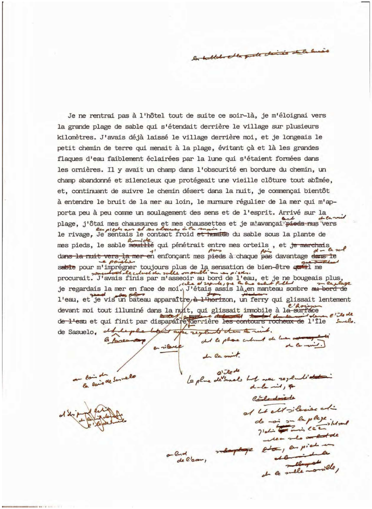
0	0	0	0	0	0	1.0
.90	0	0	0	0	0	.10
.50	0	0	0	0	0	.50
0	0	.80	0	0	0	.20
0	0	.80	0	0	0	.20
0	0	0	.90	0	0	.10

0	0	0	0	0	0
.50	0	0	0	0	0
.40	0	0	0	0	0
0	0	.30	0	0	0
0	0	.30	0	0	0
0	0	0	.10	0	0

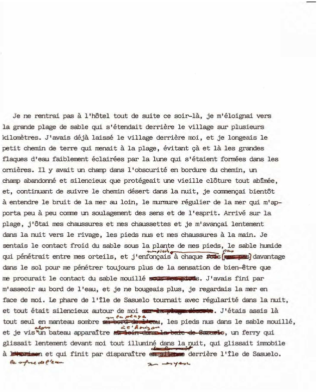
Drafts from *La Réticence* by Jean-Philippe Toussaint



Drafts from *La Réticence* by Jean-Philippe Toussaint

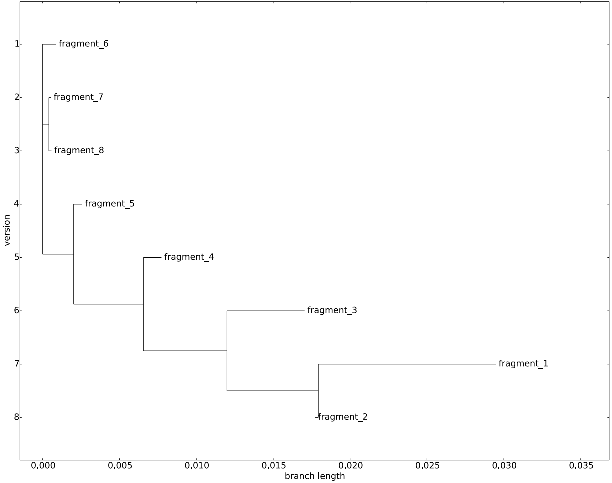


Je ne rentrai pas à l'hôtel tout de suite ce soir-là, je m'éloignai vers la grande plage de sable qui s'étendait derrière le village sur plusieurs kilomètres. J'avais déjà laissé le village derrière moi, et je longeais le petit chemin de terre qui menait à la plage, évitant çà et là les grandes flaques d'eau faiblement éclairées par la lune qui s'étaient formées dans les ornières. Il y avait un champ dans l'obscurité en bordure du chemin, un champ abandonné et silencieux que protégeait une vieille clôture tout abîmée, et, continuant de suivre le chemin désert dans la nuit, je commençai bientôt à entendre le bruit de la mer au loin, le murmure régulier de la mer qui m'apporta peu à peu comme un soulagement des sens et de l'esprit. Arrivé sur la plage, j'ôtai mes chaussures et mes chaussettes et je m'avançai pieds nus vers le rivage. Je sentais le contact froid et humide du sable sous la plante de mes pieds, le sable mouillé qui pénétrait entre mes orteils, et je marchais dans la nuit vers la mer en enfonçant mes pieds à chaque pas davantage dans le sable pour m'imprégner toujours plus de la sensation de bien-être qu'il me procurait. J'avais finis par m'asseoir au bord de l'eau, et je ne bougeais plus, je regardais la mer en face de moi. J'étais assis là en manteau sombre au bord de l'eau, et je vis un bateau apparaître à l'horizon, un ferry qui glissait lentement devant moi tout illuminé dans la nuit, qui glissait immobile à la surface de l'eau et qui finit par disparaître derrière les contours rocheux de l'île de Sasuelo,

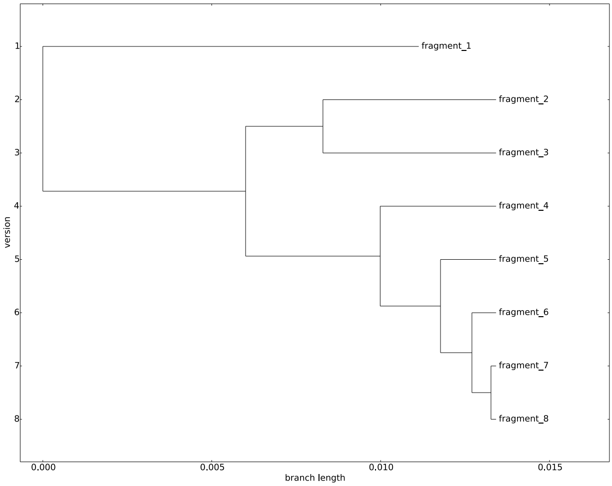


Je ne rentrai pas à l'hôtel tout de suite ce soir-là, je m'éloignai vers la grande plage de sable qui s'étendait derrière le village sur plusieurs kilomètres. J'avais déjà laissé le village derrière moi, et je longeais le petit chemin de terre qui menait à la plage, évitant çà et là les grandes flaques d'eau faiblement éclairées par la lune qui s'étaient formées dans les ornières. Il y avait un champ dans l'obscurité en bordure du chemin, un champ abandonné et silencieux que protégeait une vieille clôture tout abîmée, et, continuant de suivre le chemin désert dans la nuit, je commençai bientôt à entendre le bruit de la mer au loin, le murmure régulier de la mer qui m'apporta peu à peu comme un soulagement des sens et de l'esprit. Arrivé sur la plage, j'ôtai mes chaussures et mes chaussettes et je m'avançai lentement dans la nuit vers le rivage, les pieds nus 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 pas davantage dans le sol pour me pénétrer toujours plus de la sensation de bien-être que me procurait le contact du sable mouillé sous mes pieds. J'avais fini par m'asseoir au bord de l'eau, et je ne bougeais plus, je regardais la mer en face de moi. Le phare de l'île de Sasuelo tournait avec régularité dans la nuit, et tout était silencieux autour de moi sur la plage déserte. J'étais assis là tout seul en manteau sombre au bord de l'eau, les pieds nus dans le sable mouillé, et je vis un bateau apparaître au loin dans la baie de Sasuelo, un ferry qui glissait lentement devant moi tout illuminé dans la nuit, qui glissait immobile à l'horizon et qui finit par disparaître en silence derrière l'île de Sasuelo.

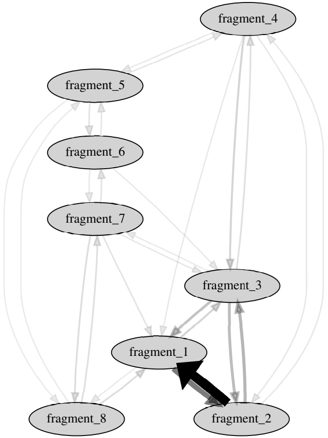
Drafts clustering : Neighbor-Joining



Drafts clustering : UPGMA



Drafts causality inference (full directed information)



Sample problem

Description

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

Sample problem

Description

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

The usual approach

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

$$PSD_s(f) = \int_{-\infty}^{\infty} \mathbb{E}\left[s(t)s(t+\tau)\right] e^{-2i\pi f\tau} d\tau$$

Then, feature extraction, etc.

Fitting the AIT framework

Accessing frequency information

Compute successive residuals R_f in Butterworth filter bank.
Note : $s_{\text{min}} \leq R_f(t) \leq s_{\text{max}}$, too.

Fitting the AIT framework

Accessing frequency information

Compute successive residuals R_f in Butterworth filter bank.
Note : $s_{\text{min}} \leq R_f(t) \leq s_{\text{max}}$, too.

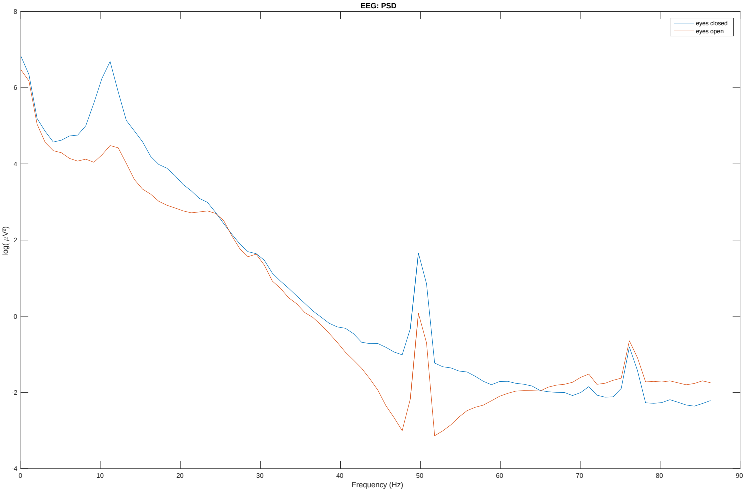
Quantization

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

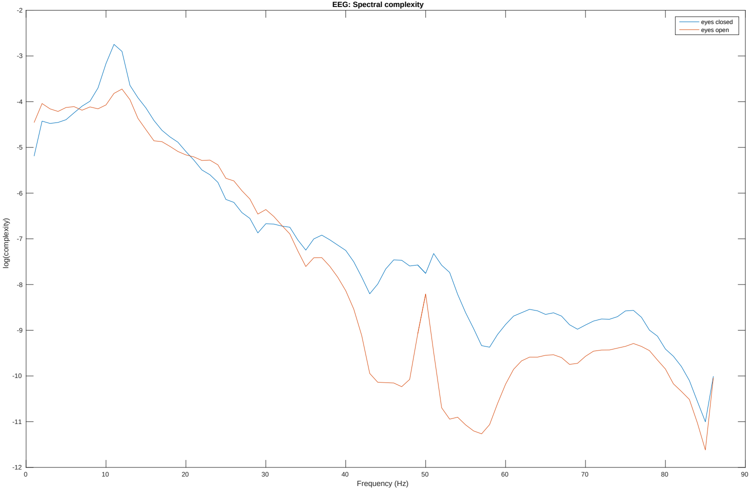
$$x_f(t) = \text{Rint}\left(255 \times \frac{R_f(t)}{s_{\text{max}} - s_{\text{min}}}\right).$$

Note : Many other choices in the literature, sometimes quite involved.

Comparing power vs. complexity [dB] of EEG signals



Comparing power vs. complexity [dB] of EEG signals



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.