Some work on event relevance

Adrian Dimulescu
May 10th, 2010
Telecom ParisTech
Infres
Phd supervisor: Jean-Louis Dessalles

Plan

- Intro : algorithmic complexity
- Inversion problem
- Encoding, meaning
- Surprise as quantity of change
- Etat de l'art

Algorithmic complexity

- ...or Kolmogorov complexity
- Minimum length of code you need to feed a Turing Machine in order to produce a string
- Independent of TM (up to an additive “translation” constant)
- Defines the distinction between regular / random strings
  - 01010101010101010101
  - 00101110110111011110

Minimal encoding

Dans le métro parisien, trouver l'itinéraire le plus court pour un trajet

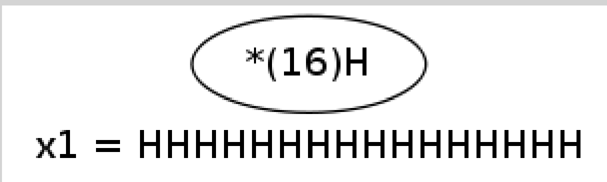
Inversion problem

- Encode/compress a piece of data, ex: x = 'HHHHHHHH'
- Possible encodings, examples:
  - 8*H
  - 4*H4*H
  - 2*H3*H2*HH
  - ...
- Which one to choose? The shortest.
- Incomputable in general case but maybe we can get around it

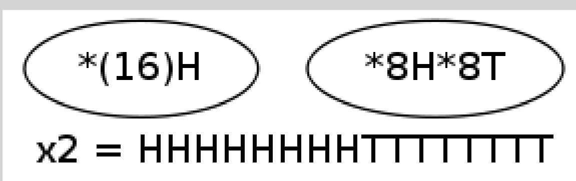
Meaning as encoding

The world is not random, like a 'white snow' (no channel) TV screen, how to grasp regularities?

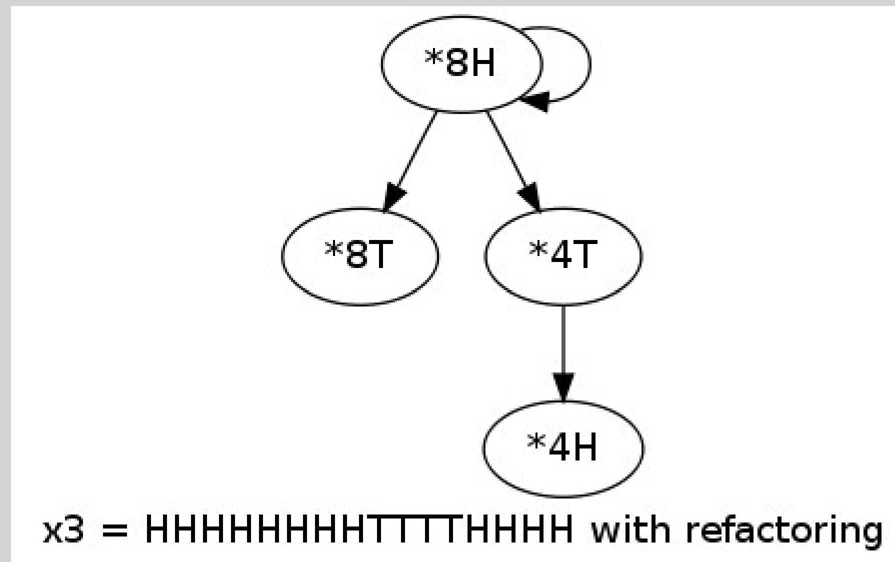
A first sample, *x1*, is presented



Second sample is presented



Third sample, model changes



ressembles a bayesian network,
with programs in nodes

Concept = Program

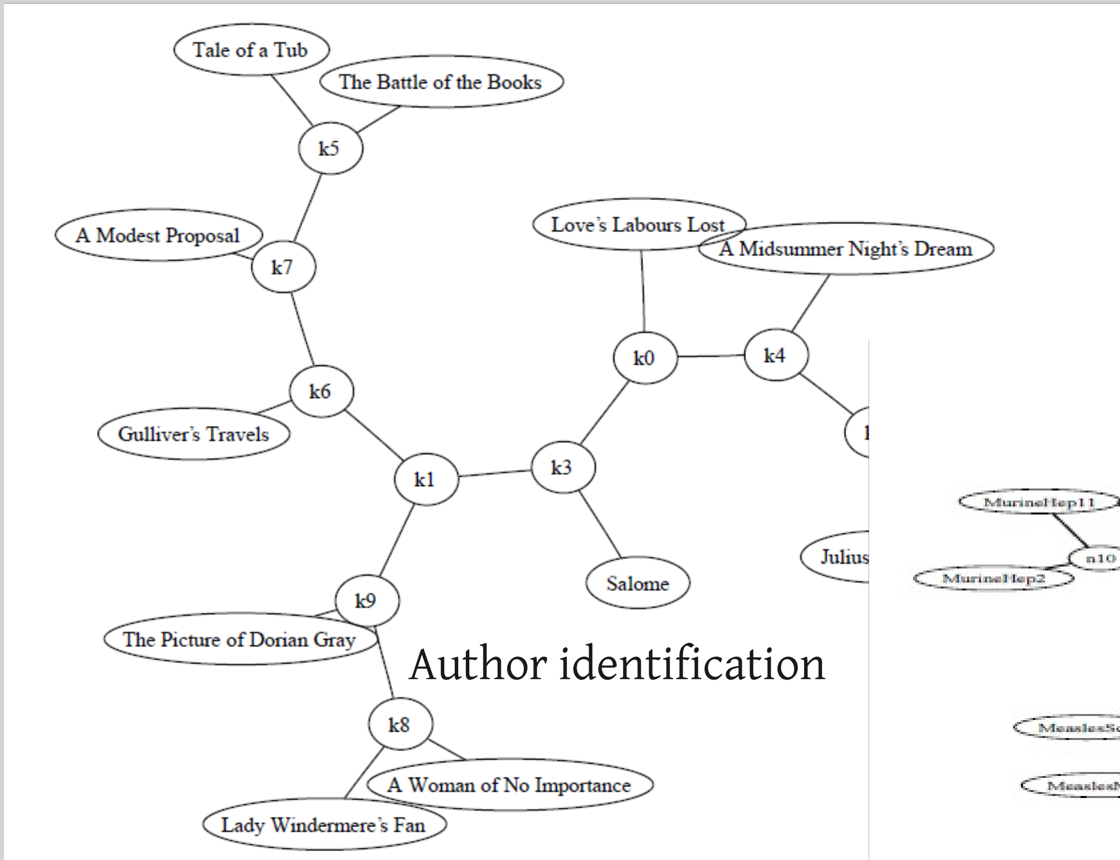
- “Meaning” of data x : a program p so that $M(p) = x$
  - Aka *re-presentation*
  - There can be several meanings (interpretations) for a sign (piece of data)
  - How to choose: favor the shortest or maybe take'em all?
- So concepts like “man”, “horse”, “penguin” : programs (or, equivalently, codes)

Property = Program

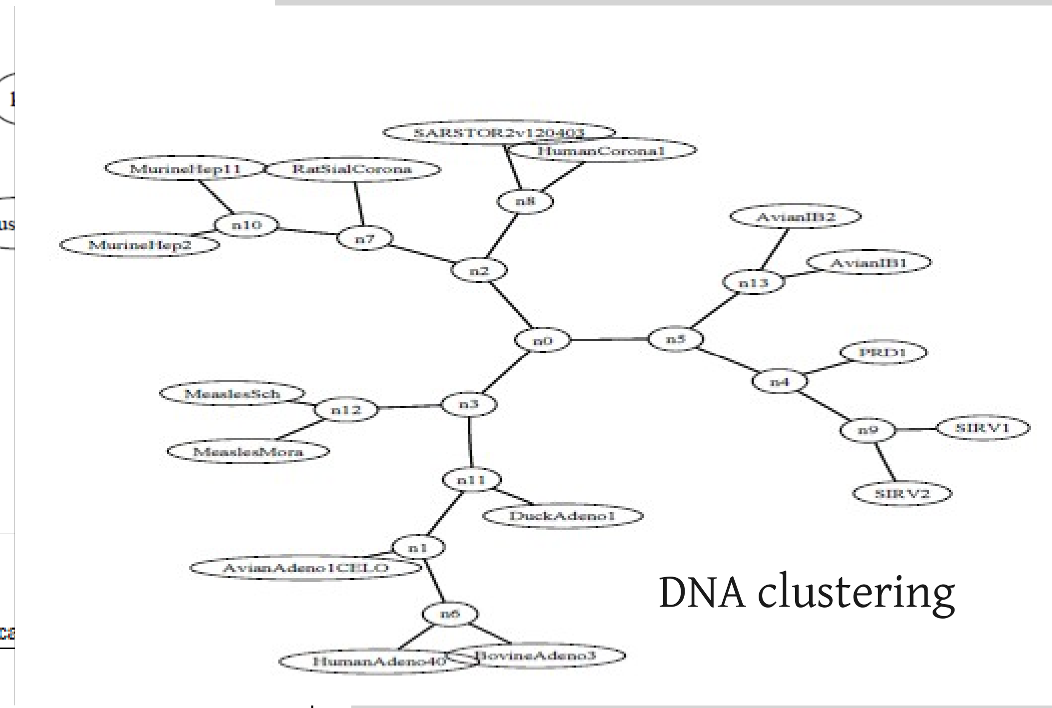
Semantic applications (Vitanyi & al)

- Information distance
- Semantic distance between texts, music, DNA sequences
- Normalized "Google" Distance
  - Semantic distance between words (concepts)

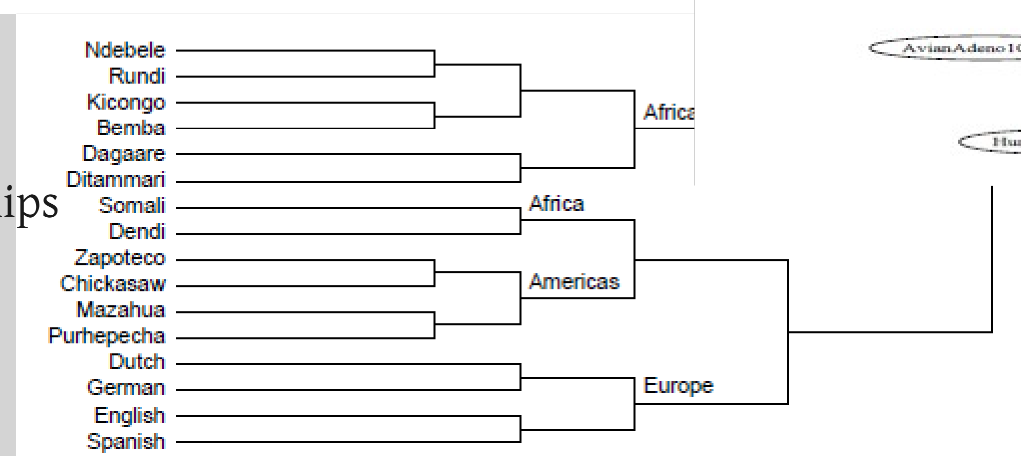
Computable applications of a theoretically incomputable complexity (stop saying it's useless)



Author identification



DNA clustering

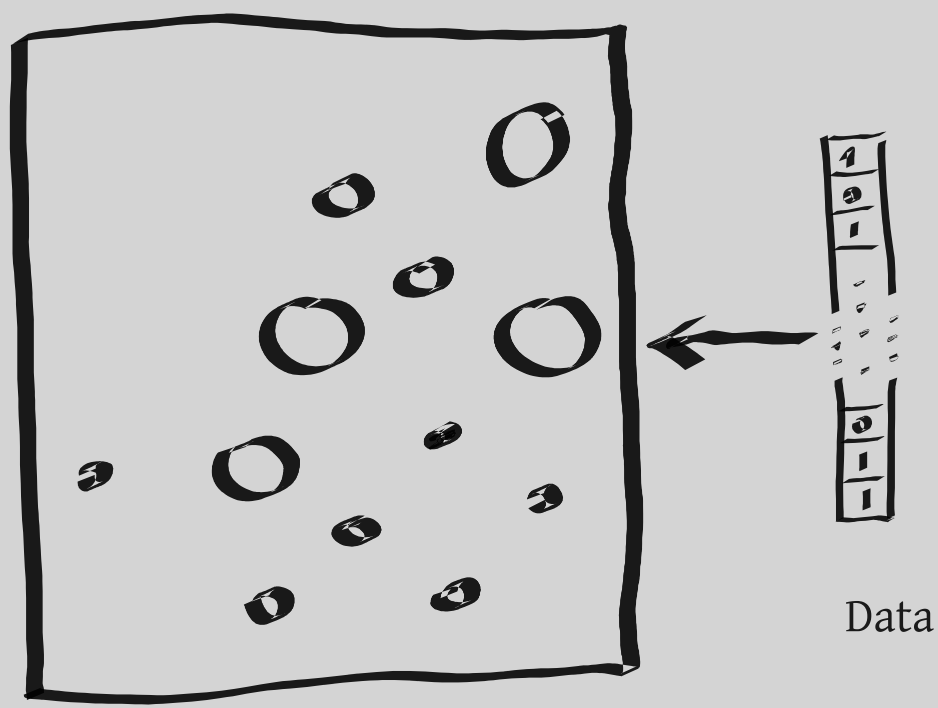


Language relationships

Cilibrasi&Vitany, 2005,2007

An encoding machine

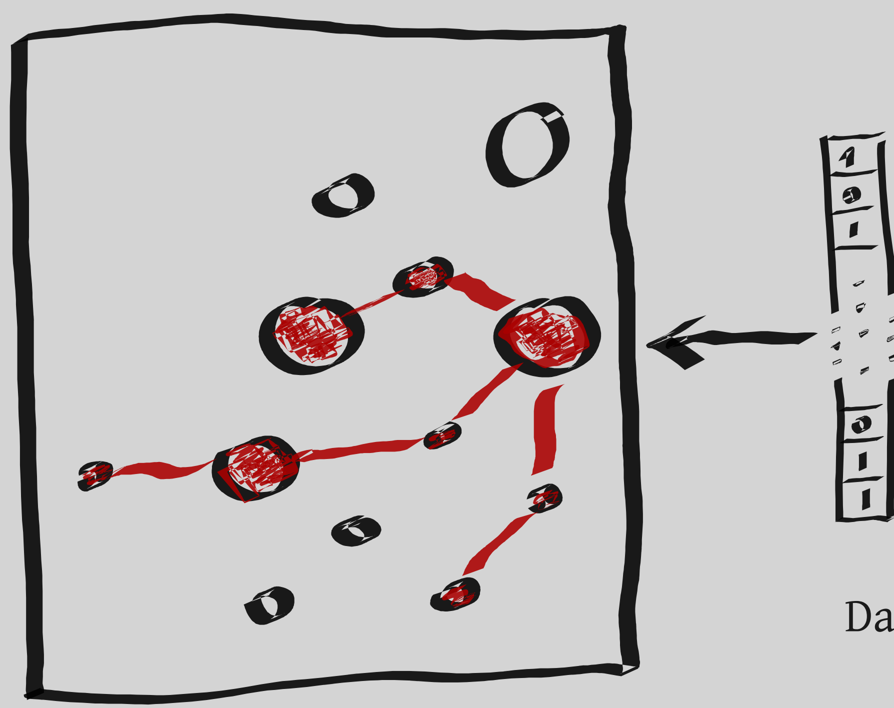
- Inversion problem (Solomonoff, 1965): given datum x, find program p so that $M(p) = x$.



Box of reusable programs (subroutines)

An encoding machine (2)

“all possible causes of emergence”



Data

Box of reusable programs (subroutines)

Two-part codes

- Encoding with two parts
  - One containing regularities, already in the model
  - One containing the relatively-random part
- Example
  - Suppose CB = { p_1 : '8*H' }, data x = 'HHHHHHHHT'
  - $p(x) = p_1\ T$

Regular part

Random part

Algorithmic / subjective probability

- Algorithmic pr. : $Q_M(x) = \sum_{M(p)=x} 2^{-l(p)}$
- Subjective pr. : representativeness (Kahneman, Tversky,... 1982)
  - *representativeness* : “what is the probability that event A originates from process B”
  - “the representativeness hypothesis states that predictions do not differ from evaluations or assessements of similarity”
  - “the assessment of similarity is described as a feature matching process.” - Tversky, *Features of similarity*

An encoding machine (3)

Let CB be a finite set of programs for a given universal Turing machine, "cognitive box of programs". Let *x* be a piece of data presented to CB.

Let $H \subset CB = \{h_i \in CB | C(x|h_i) < b\}$ contain reasonable candidates for encoding *x* in two parts: C(h) + C(x|h), with less than *b* extra bits, an effort barrier).

Let $h_f \in H$ be the "favored" = shortest hypothesis.

Let $h_b \in H$ be the "best" program, i.e. with the two-part code $C(h_b) + C(x|h_b)$ is minimal. Then *x* is surprising for CB if $h_f \neq h_b$

and the quantity of unexpectedness is given by the difference between the two attempted encodings: $U(x, CB) = C(h_f) + C(x|h_f) - C(h_b) - C(x|h_b)$

(which is also the difference of a posteriori complexity of the two hypothesis, i.e. the quantity of change if the model "wants" to change)

Some psychological support : Keren&Teigen, 2003

Surprise as quantity of change

- Previous formula does not take into account model change
- Model change = programs/concepts changing size, sharing common subroutines/properties, new programs
- On this model, surprise = quantity of change
  - How to measure it? Number of bits changed, on all programs
  - How to decide what/how to change?
  - Too surprising – not accepted
- “adaptive compressor”- Schmidhuber

Example

CB = {alt1 = ε, alt2 = *8H, *h* = *7HT}

x = HHHHHHHT

C(x|h) = 0; C(x|alt1) = 9; If alt2 is used for encoding, the resulting program would be *8HXT so C(x|alt2) = 2; All existing hypothesis can encode x, alt2 is the favored smallest hypothesis:

C(h) + C(x|h) = 5 + 0 <- best encoding

C(alt1) + C(x|alt1) = 0 + 9C(alt2) + C(x|alt2) = 4 + 2 = 6 <- favorite hypothesis loses by 1 bit, so small surprise.

U(x) = 6 - 5 = 1 bit

Example2

What if lots of HHHHHHHH come first and only then one HHHHHHHT ?

Suppose there are lots of HHHHHHHH that are presented to the cognitive box. Each presentation of HHHHHHH is followed by pressure to the overall compression device in order for a minimization of the representation of the stimulus (corresponding to the bayesian posterior probability update), ending up with having a one-bit representation.

CB = {alt1 =ε, alt2 = # = *8H, h = *7HT}

-
- C(h) + C(x|h) = 5 + 0
- C(alt1) + C(x|alt1) = 0 + 9
- C(alt2) + C(x|alt2) = 2 + 2 = 4 <- favorite hypothesis wins, no surprise
-

Model organization: logical depth?

- "We propose depth as a formal measure of value." (Bennett 88)
- "Some mathematical and natural objects (a random sequence, a sequence of zeros, a perfect crystal, a gas) are intuitively trivial, while others (e.g. the human body, the digits of Pi) contain internal evidence of a nontrivial causal history."
- "Logical depth is the necessary number of steps in the deductive or causal path connecting an object with its plausible origin. Formally, it is the time required by a universal computer to compute the object from its compressed original description." (Li&Vitanyi, 97)
- Slow growth : deep objects cannot be *quickly* produced from shallow ones by any deterministic process
- Examples: math book, DNA, "wisdom"

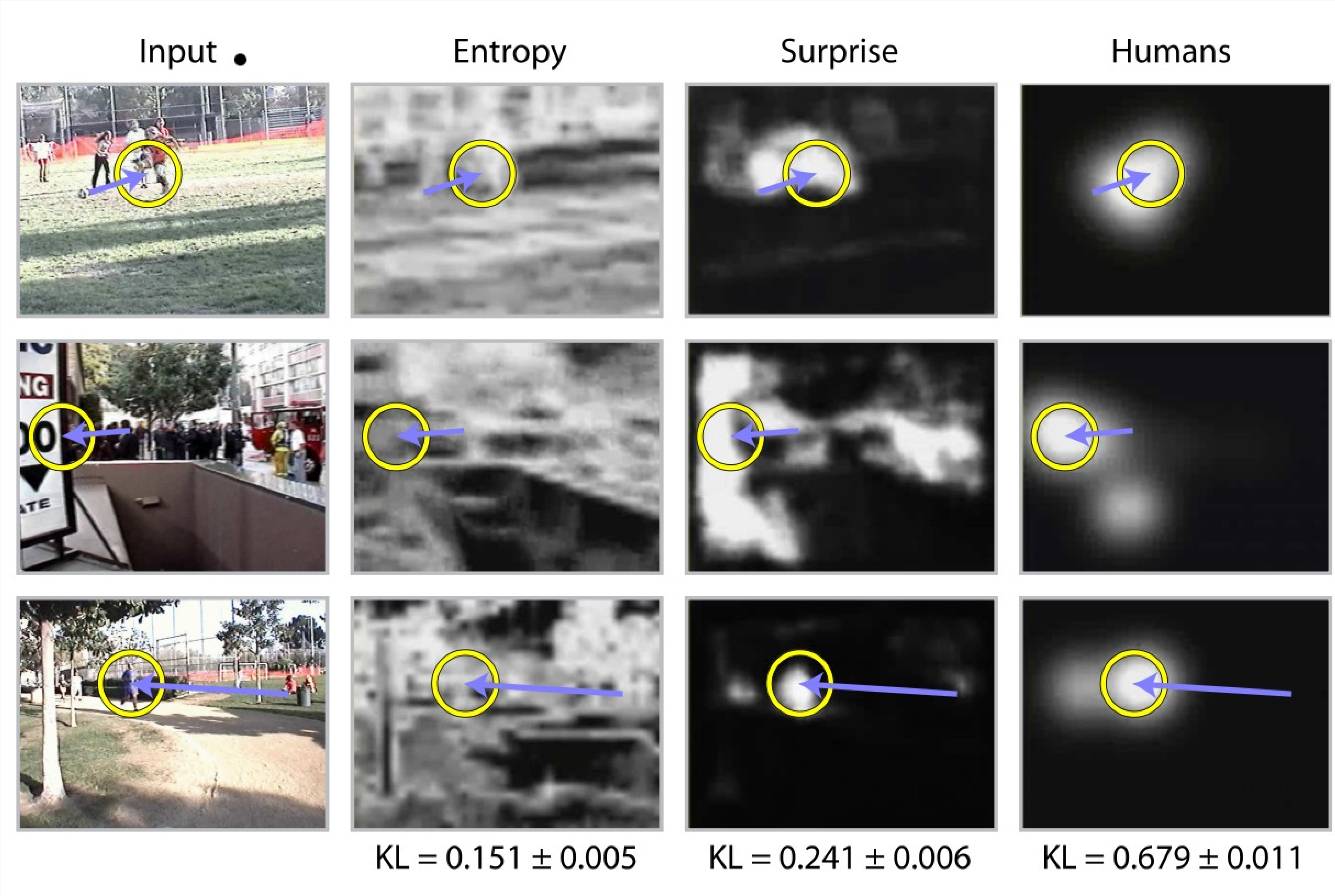
$$depth_\epsilon(x) = \min\left\{ t : \frac{Q_U^t(x)}{Q_U(x)} \geq \epsilon \right\}$$

Parallel views

Bayesian theory of surprise (P. Baldi & friends)

- Model = probability distribution modeled in a bayesian network
- Surprise = difference between posterior/prior probability distributions (beliefs)
- Diff = Kullback-Leibler distance between the two distributions

$$S(D,M) = KL(P(M|D),P(M)) = \int_M P(M|D)\log\frac{P(M|D)}{P(M)}dM$$



(Itti, Baldi, *Bayesian surprise attracts human attention,* 2009)

Surprise & beauty (J. Schmidhuber)

- Beauty : data is a two-part code with overwhelming regular part ("already known")
- Surprise : first derivative of beauty (compression improvement)
  - "as the learning agent improves its compression algorithm, formerly apparently random data parts become subjectively more regular and beautiful" (Schmidhuber,2009)

$$I(D,O(t)) \sim \frac{\partial B(D,O(t))}{\partial t}$$

- Possibility of beautiful and surprising at the same time (perpetual discovery of known)? - relate to logical depth

Novelty

- “red spot”
- “three-legged chicken”
- “blue man”
- “happy phd student”

Novelty as co-occurrence

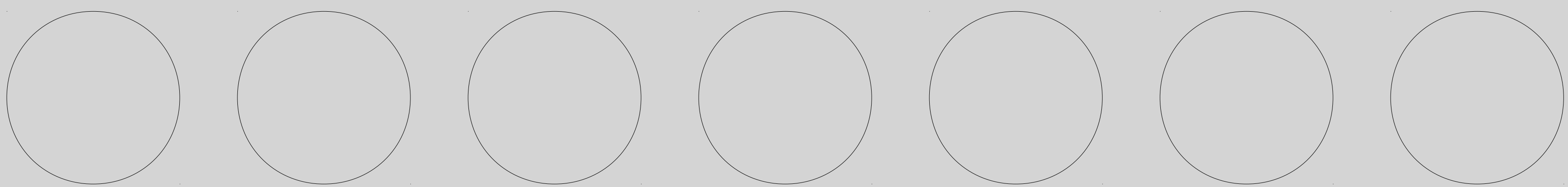
Why is novelty interesting?

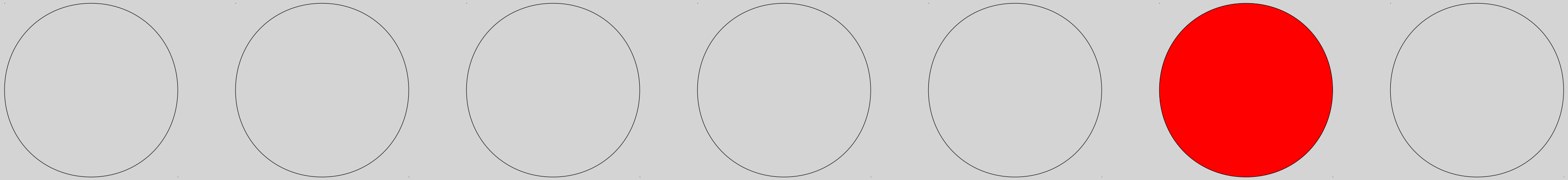
- Overall minimization constraint
- Incentive to code sharing
- Green grass / verdure

Thanks!

Advertisement

Please stop ugly Times New Roman proliferation !





Focus / attention

- “the *red* spot got the ball”
  - ok, what about:
- “the *white* spot got the ball”
  - Which white spot, exactly?
  - Correct: “a white spot got the ball”
- “The” / “a” → focus, attention
  - Focus: complete detailed encoding
  - Background: partial (generic) encoding?