Complexity Theory — Part 6: Algorithmic information and logical depth

Dusko Pavlovic

RHUL

Spring 2012

Outline

Outline

What did we achieve?

What did we achieve?

The question was:

Why are there complex phenomena?

Where does the complexity come from?

Why are there complex phenomena?

Why does the evolution increase complexity?

Why are there complex phenomena?

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Why are we so complex?

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Did complexity theory bring us closer to an answer?

What is complexity?

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Which strings are complex?

- $\underbrace{01010101010101010101010101010101010\cdots01}_{\text{100 digits}}$
- $\underbrace{0100110001110000111100000111110\cdots1101}_{\text{100 digits}}$
- $\underbrace{1101000100110101001010111010100\cdots1101}_{\text{100 digits}}$

What is complexity?

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

How do you know which strings are complex?

- $\underbrace{01010101010101010101010101010101010\cdots01}_{\text{100 digits}}$
- $\underbrace{0100110001110000111100000111110\cdots1101}_{\text{100 digits}}$
- $\underbrace{1101000100110101001010111010100\cdots1101}_{\text{100 digits}}$

What is complexity?

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

- $\underbrace{010101010101010101010\cdots01}_{100}$ can be written as
  - $(01)^{50}$

What is complexity?

- $\underbrace{01010101010101010101\cdots01}_{100}$ can be written as
  - $(01)^{50}$
  - do i=1..50 write 01 od

What is complexity?

- $\underbrace{01010101010101010101\cdots01}_{100}$ can be written as
  - $(01)^{50}$
  - do i=1..50 write 01 od
- $\underbrace{01001100011100001111 0\cdots11}_{100}$ can be written as
  - $\underbrace{0^1 1^1 0^2 1^2 \cdots 0^i 1^i \cdots}_{100}$
  - i=1; do until length=100 write 0^i1^i; i = i+1 od

What is complexity?

- $\underbrace{01010101010101010101\cdots01}_{100}$ can be written as
  - $(01)^{50}$
  - do i=1..50 write 01 od
- $\underbrace{01001100011100001111 0\cdots11}_{100}$ can be written as
  - $\underbrace{0^1 1^1 0^2 1^2 \cdots 0^i 1^i \cdots}_{100}$
  - i=1; do until length=100 write 0^i1^i; i = i+1 od
- $\underbrace{11010001001101010 0101\cdots00}_{100}$ can be written as
  - print 110100010011010100101 $\cdots$ 00

Complexity = length of description

Idea

Fix a language and define for any x

$$C(x) \quad = \quad \text{length of the simplest description } x$$

Complexity = length of description

$$C(0101010\ldots) < C(010011000\ldots) < C(110100010\ldots)$$

What is a description?

What is a description?

- $\beta = \bigwedge_{C(x)>20} x$

What is a description?

- $\beta = \bigwedge_{C(x)>20} x$
- The length of the expression "$\bigwedge_{C(x)>20} x$" is < 20

What is a description?

- $\beta = \bigwedge_{C(x)>20} x$
- The length of the expression "$\bigwedge_{C(x)>20} x$" is < 20
- Therefore β does not exist.

Berry Paradox

- $\beta = \bigwedge_{C(x)>20} x$
- The length of the expression "$\bigwedge_{C(x)>20} x$" is < 20
- Therefore β does not exist.
- Therefore all x satisfy $C(x) \leq 20$

Berry Paradox

- $\beta = \bigwedge_{C(x)>20} x$
- The length of the expression "$\bigwedge_{C(x)>20} x$" is < 20
- ~~Therefore β does not exist.~~
- ~~Therefore all x satisfy $C(x) \leq 20$~~
- "$\bigwedge_{C(x)>20} x$" is not a valid description

Moral

We need a formal definition of a description.

Outline

What did we achieve?

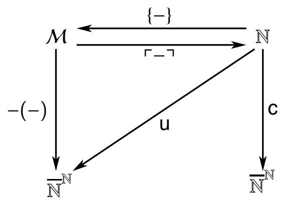
Kolmogorov: Algorithmic information

Solomonoff: Algorithmic probability

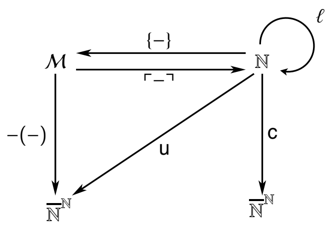
Chaitin: The number of wisdom

Bennett: Logical depth

Idea: descriptions = programs

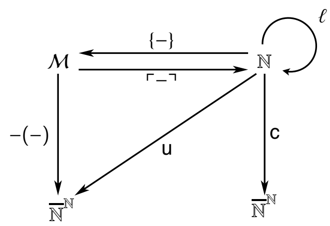


Assume programs with a "length" measure



$\ell(p)$ = a *code length* of p

Assume programs with a "length" measure



$\ell(p)$ = a *code length* of p

(e.g. the number of symbols in p)

Complexity = length of program

Idea

Kolmogorov complexity of x is the length of the simplest program that outputs x

$$K(x) = \bigwedge_{\{p\}()=x} \ell(p)$$

Escape from Berry Paradox

Fact

The predicate

$$K(x) \leq y$$

is not decidable, or else

$$\beta = \bigwedge_{K(x)>20} x$$

would be a program.

Consequence

Komogorov complexity does not satisfy the axioms of complexity measures.

Description distance

Definition

For a given 2-tape Turing Machine M, the *M-description distance* from a string x to a string y is

$$K_M(y \mid x) \quad = \quad \bigwedge_{M(p,x)=y} \ell(p)$$

The machine M is called the *interpreter*.

Description distance

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Definition

For a given 2-tape Turing Machine M, the *M-description distance* from a string x to a string y is[1]

$$K_M(y \mid x) \quad = \quad \bigwedge_{M(p,x)=y} \ell(p)$$

The machine M is called the *interpreter*.

[1]The definition of infimum implies $\bigwedge \emptyset = \infty$.

Description distance

Notation

For any pair of functions $f, g : \mathbb{N} \to \mathbb{N}$ we write

$$f \overset{+}{\leq} g \iff \exists c \forall n.\ f(n) \leq g(n) + c$$
$$f \overset{+}{=} g \iff f \overset{+}{\leq} g \ \wedge \ g \overset{+}{\leq} f$$

Description distance

Proposition 1 (The Invariance Theorem)

There is a universal interpreter: a 2-tape Turing Machine U such that for all 2-tape Turing Machines M holds

$$K_U(y \mid x) \quad \overset{+}{\leq} \quad K_M(y \mid x)$$

Description distance

Proof.

If U is universal machine for all 2-tape machines, i.e.

$$U(\ulcorner M \urcorner, x, y) \quad = \quad M(x, y)$$

then define

$$U'(J(x, y), z) \quad = \quad U(x, y, z)$$

where J is Cantor's pairing function, so that

$$U'(J(\ulcorner M \urcorner, p), x) \quad = \quad M(p, x)$$

and thus

$$K_U(y \mid x) \quad \leq \quad K_M(y \mid x) + \ell(\ulcorner M \urcorner) + 6$$

$\sqcap$

Description distance

Scholium

For any two universal interpreters U, V holds

$$K_U(y|x) \quad \stackrel{+}{=} \quad K_V(y|x)$$

Description distance

Definition

The *description distance* from a string x to a string y (or *Kolmogorov complexity* of y relative to x) is the description distance with respect to a universal interpreter

$$K(y \mid x) \quad = \quad \bigwedge_{\{p\}(x)=y} \ell(p)$$

Description distance

Definition

The *description complexity* (or *Kolmogorov complexity*) of a string y is the description distance from the empty string $\langle\rangle$ to y

$$K(y) \quad = \quad \bigwedge_{\{p\}()=y} \ell(p)$$

Properties of description complexity

- $\#\{x \in \{0,1\}^\ell \mid K(x) < \ell - m\} \leq 2^{\ell - m}$
- $m(x) \leq K(x) \leq \ell(x)$ for $m(x) = \bigwedge_{x \leq y} K(y)$
- $\lim_{x\to\infty} m(x) = \infty$
- $\forall \phi \in RPF.\ \lim_{x\to\infty} \phi(x) = \infty \Rightarrow m(x) < \phi(x)$
- $\exists \chi \in RTF.\ \lim_{t\to\infty} \chi(t, x) = K(x)$
- $\exists h.\ |K(x+h) - K(x)| < 2h$

Compression and randomness

Question

- Temperature is the average of kinetic energies.
- What is the entropy average of?

Compression and randomness

Theorem (Schack 1997)

Kolmogorov complexity can be defined with optimal encoding and then

$$H(q) \quad \approx \quad \int_{i \in I} K(q_i)$$

Compression and randomness

Upshot

The most informative strings cannot be compressed:

$$K(x) = \ell(x)$$

Compression and randomness

Upshot

The most informative strings cannot be compressed:

$$K(x) = \ell(x)$$

This is Kolmogorov's definition of *randomness*.

Outline

Entropy encoding: more likely $\rightsquigarrow$ shorter

Shannon's Noiseless Coding

Let $\mu : \Sigma \to [0,1]$ be the frequency distribution of the alphabet Σ. Then the optimal word length for representing a symbol $s \in \Sigma$ is

$$\ell(s) = -\log \mu(s)$$

Entropy encoding: more likely $\rightsquigarrow$ shorter

Shannon's Noiseless Coding

Let $\mu : \Sigma \to [0,1]$ be the frequency distribution of the alphabet Σ. Then the optimal word length for representing a symbol $s \in \Sigma$ is

$$\ell(s) = -\log \mu(s)$$

Remark

The Shannon entropy is the average word length.

Occam's Razor: shorter $\rightsquigarrow$ more likely

Solomonoff's Algorithmic Probability

Let U be a *self-delimiting* universal Turing machine, i.e.

$$U(p)\downarrow \implies \forall q.\ \neg U(p :: q)\downarrow$$

Then the *a priori* probability distribution of data from a dataset Σ is

$$\mu(y) = \sum_{U(p)=y} 2^{-\ell(p)}$$

Occam's Razor: shorter $\rightsquigarrow$ more likely

Solomonoff's Algorithmic Probability

Let U be a *self-delimiting* universal Turing machine, i.e.

$$U(p)\downarrow \implies \forall q.\ \neg U(p :: q)\downarrow$$

Then the *a priori* probability distribution of data from a dataset Σ is

$$\mu(y) = \sum_{U(p)=y} 2^{-\ell(p)}$$

Remark

Self-delimiting $\implies \sum_{y\in\Sigma}\mu(y) \leq 1$.

Occam's Razor: shorter $\rightsquigarrow$ more likely

Solomonoff's Algorithmic Probability

Let U be a *self-delimiting* universal Turing machine, i.e.

$$U(p,x)\downarrow \implies \forall q.\ \neg U(p :: q, x)\downarrow$$

Then the *a priori* conditional probability distribution of data from a dataset Σ is

$$\mu(y \mid x) = \sum_{U(p,x)=y} 2^{-\ell(p)}$$

Remark

Self-delimiting $\implies \sum_{y\in\Sigma}\mu(y) \leq 1$.

Inductive inference

Solomonoff's model of theory formation

- phenomena are modeled as bitstrings $d, h \ldots \in \{0,1\}^*$
- d — observed data
- h — hypothetic data
- $\{p\}(d) = h$ — explanation of causality
- $\Pr(h \mid d) = \bigvee_{\{p\}(d)=h} 2^{-\ell(p)}$ — the simplest explanation is the best

Inductive inference

Inferring the most effective explanations

- The goal is to maximize

$$\Pr(h \mid d) = \frac{\Pr(d \mid h)\cdot \Pr(h)}{\Pr(d)}$$

Inductive inference

Inferring the most effective explanations

- The goal is to maximize

$$\Pr(h \mid d) = \frac{\Pr(d \mid h)\cdot \Pr(h)}{\Pr(d)}$$

- Since $\Pr(x) = 2^{-K(x)}$, this is equivalent to minimizing

$$K(h \mid d) = K(d \mid h) + K(h) - K(d)$$

Inductive inference

Minimum Description Length Principle (MDLP)

Given an observation d, the best explanation minimizes

$$\ell(p) + \ell(q)$$

where

$$\{p\}() = h \qquad \{q\}(h) = d$$

Outline

What did we achieve?

Kolmogorov: Algorithmic information

Solomonoff: Algorithmic probability

Chaitin: The number of wisdom

Bennett: Logical depth

Complex descriptions

The halting number

$$\kappa = .\kappa_1\kappa_2\kappa_3\cdots\kappa_p\cdots$$

$$\kappa_p = \begin{cases} 1 & \text{if } U(p)\downarrow \\ 0 & \text{otherwise} \end{cases}$$

Complex descriptions

The halting number

$$\kappa = .\kappa_1\kappa_2\kappa_3\cdots\kappa_p\cdots$$

$$\kappa_p = \begin{cases} 1 & \text{if } U(p)\downarrow \\ 0 & \text{otherwise} \end{cases}$$

Remark

The value of κ depend on a fixed Universal Turing Machine U. Its crucial properties do not depend on the choice of U.

Complex descriptions

Fact

κ is highly compressible!

Complex descriptions

Definition

$$K^t_M(y) = \bigwedge_{\substack{\{p\}()=y \\ time(p,y)\le t(|y|)}} \ell(p)$$

Complex descriptions

Proposition (Barzdin)

Suppose that $Y = \{i \in \mathbb{N} \mid y_i = 1\}$ is recursively enumerable. Then for any $c > 0$ there is a total recursive function t such that

$$K^t_M(y) \le c \log y$$

Complex descriptions

Proposition 2

For every n there is a program of length $\leq 2\log n$ that outputs the first n digits of κ.

Complex descriptions

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Question

What might this compressed κ look like?

Complex descriptions

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Chaitin's number

$$\Omega = \sum_{s\in\mathbb{N}} \mu(s) = \sum_{U(p)\downarrow} 2^{-\ell(p)}$$

Complex descriptions

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Chaitin's number

$$\Omega = \sum_{s\in\mathbb{N}} \mu(s) = \sum_{U(p)\downarrow} 2^{-\ell(p)}$$

Interpretations

- probability that a hypothesis will be formed a priori
- probability that a randomly chosen program will halt

Complex descriptions

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Proposition 3

Ω is incompressible.

Complex descriptions

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Proposition 4

$\Omega_{1..n}$ decides halting of all programs of length up to n.

Complex descriptions

Interpretation

- Many open problems can be formulated as the questions whether certain search programs will halt:
  - Riemann Hypothesis
  - $P = NP$, $AP = ANP$, one-way, trapdoor, DDH…
  - Twin Primes…

Complex descriptions

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Interpretation

- Many open problems can be formulated as the questions whether certain search programs will halt:
  - Riemann Hypothesis
  - $P = NP$, $AP = ANP$, one-way, trapdoor, DDH…
  - Twin Primes…
- These programs are shorter than 5000 characters.
- Knowing $\Omega_{1..5000}$ would resolve most of the open problems of mathematics.

Outline

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Time bounded algorithmic probability

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Definition

$$\mu^t(y) \quad = \sum_{\substack{\{p\}()=y \\ time(p,y)\le t(|y|)}} 2^{\ell(p)}$$

Logical depth

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Definition (Bennett, Adleman…)

$$L_\epsilon(y) \quad = \quad \bigwedge \left\{ t \le REC \,\middle|\, \frac{\mu^t(y)}{\mu(y)} \ge \epsilon \right\}$$

Logical depth

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Definition (Bennett, Adleman…)

$$L_\epsilon(y) \quad = \quad \bigwedge \left\{ t \le REC \,\middle|\, \frac{\mu^t(y)}{\mu(y)} \ge \epsilon \right\}$$

Interpretation

The shortest time in which that one of the minimal programs that output y will halt, with the conditional probability $\ge \epsilon$.

Logical depth

Proposition 5 (Adleman)

The predicate $L_\epsilon(y) \leq O(n^k)$ is decidable in polynomial time.

Logical depth

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Proposition 6 (Bennett)

Deep strings cannot be quickly computed from shallow ones.

Logical depth

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Bennett's interpretation

"A structure is deep if most of its algorithmic probability is contributed by slow-running processes."

Logical mutual information

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Levin's Law of Conservation of Information

Let

- d be a stream of randomized observations, and
- h a stream generated by a deterministic mathematical model.

Then the probability that the mutual information is $I(D : H) > m$ is less than 2^{-m}.

Logical depth

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Levin's comment

"Following Church's Thesis, Theorem 3 precludes the increase of information through computational processes. Theorem 2 precludes the increase of information through a combination of computational and randomized processes."

Logical depth

Complexity 6: AIT
Dusko Pavlovic
What
Kolmogorov
Solomonoff
Chaitin
Bennett

Levin's interpretation

"Our results contradict the assertion of some mathematicians that the truth of any valid proposition can be verified in the course of scientific progress through informal methods. (To do so by formal methods has been proven impossible by Gödel.)"