Computational Complexity and Computability

Lecture 9 - Information & Kolmogorov Complexity

Koushik Pal

University of Toronto

February 3, 2021

Information

Two fundamental concepts in computer science:

- Algorithm
- Information

Unlike algorithms, there is no universally accepted comprehensive definition for information.

One definition of information is via computability theory.

Information

Question: Can we quantify how much information is contained in a string?

Example

- $x = 110110110110110110110110110110110110110110110110$

- $y = 10110111011110111110111111011111110111111110$
- $z = 110001110101010011101110001000110100101011110$

Idea: The more we can "compress" a string, the less "information" it contains.

Thesis: The amount of information in a string is equivalent to the shortest way of describing that string.

Information

Question: How to describe strings?

Answer: Use Turing machines with inputs.

To be more specific, describe a string x as $\langle M, w \rangle$ such that M is a TM that, on input w, halts with only x on its tape.

Kolmogorov Complexity

Definition

Let $x \in \{0, 1\}^*$.

The minimal description of x, denoted $d(x)$, is the lexicographically shortest string $\langle M, w \rangle$ such that M is a TM that on input w halts with only x on its tape.

The descriptive complexity (also known as Kolmogorov complexity) of x, denoted $K(x)$, is the length of the minimal description of x, i.e.,

$$K(x) = |d(x)|.$$

Encoding

Problem: How to figure out where M ends and w starts in the encoding $\langle M, w\rangle$?

There are many possible solutions. Here are two examples. Assume the alphabet is $\{0, 1\}$.

- Write each bit of $\langle M\rangle$ twice, i.e., 0 as 00 and 1 as 11, and use 01 as a delimiter between $\langle M\rangle$ and w; e.g.

$$\langle M, w\rangle = \underbrace{11001111001100\cdots1100}_{\langle M\rangle}\,01\,\underbrace{0110101001010}_{w}.$$

  In this case, $|\langle M, w\rangle| = 2|\langle M\rangle| + |w| + 2$.

- If $\langle M\rangle = z_1 z_2 \ldots z_k \in \{0, 1\}^*$ and $w = w_1 w_2 \ldots w_n \in \{0, 1\}^*$, let

$$\langle M, w\rangle = 0z_1 0z_2 0 \ldots 0z_k 1 w_1 w_2 \ldots w_n.$$

  In this case, $|\langle M, w\rangle| = 2|\langle M\rangle| + |w| + 1$.

Properties of Kolmogorov Complexity

Property 1

The amount of "information" in x isn't much more than $|x|$.

Theorem

There is a constant c such that for all $x \in \{0, 1\}^*$,

$$K(x) \leq |x| + c.$$

Proof.

Define the TM M_{id}, which on any input w, immediately halts, thereby leaving w on the tape.

Clearly, $\langle M_{id}, x \rangle$ is a description of x, and hence

$$K(x) \leq |\langle M_{id}, x \rangle| = 2|\langle M_{id} \rangle| + |x| + 1 = |x| + c,$$

where $c = 2|\langle M_{id} \rangle| + 1$. $\square$

Properties of Kolmogorov Complexity

Property 2

The amount of "information" in xx isn't much more than in x, i.e., repetitive strings have low information.

Theorem

There is a constant c such that for all $x \in \{0, 1\}^*$,

$$K(xx) \leq K(x) + c.$$

Proof.

Define a TM N as follows:

On input $\langle M, w \rangle$:

- *Simulate M on w, let s be the result.*
- *Output ss.*

Let $\langle M, w \rangle$ be the minimal description of x. Then $\langle N, \langle M, w \rangle \rangle$ is a description of xx. Hence,

$$K(xx) \leq |\langle N, \langle M, w \rangle \rangle| = 2|\langle N \rangle| + K(x) + 1 = K(x) + c,$$

where $c = 2|\langle N \rangle| + 1$. $\square$

Properties of Kolmogorov Complexity

Corollary

There is a constant c such that for all $n \geq 2$ and $x \in \{0,1\}^*$,

$$K(x^n) \leq K(x) + c \log n.$$

In particular, $K((01)^n) = \mathcal{O}(\log n)$.

Proof

Define a TM T as follows:

On input $\langle n, \langle M, w\rangle\rangle$:

- *Simulate M on w; let s be the result.*
- *Print s for n times.*

Let $\langle M, w\rangle$ be the minimal description of x. Then $\langle T, \langle n, \langle M, w\rangle\rangle\rangle$ is a description of x^n. Hence,

$$\begin{aligned} K(x^n) \leq |\langle T, \langle n, \langle M, w\rangle\rangle\rangle| &\leq 2|\langle T\rangle| + 2\lceil \log n\rceil + K(x) + 2 \\ &\leq K(x) + \mathcal{O}(\log n). \end{aligned}$$

□

Does the model matter?

Question: TMs are programming languages. If we used another programming language, could we get significantly shorter descriptions?

Answer: No!

Does the model matter?

Definition

An interpreter is a semi-computable function $p : \{0,1\}^* \to \{0,1\}^*$ (which takes programs as inputs and prints their outputs).

The minimal description of x under p, denoted $d_p(x)$, is the lexicographically shortest string s for which $p(s) = x$. (For example, $d_{Python}(x)$ is the shortest binary encoding of a Python program that outputs x.)

Finally, define $K_p(x) = |d_p(x)|$ as the descriptive complexity of x under p.

Does the model matter?

Theorem

For every interpreter p, there is a constant c such that for all $x \in \{0,1\}^*$,

$$K(x) \leq K_p(x) + c.$$

(In other words, using any other programming language would only change $K(x)$ by some constant.)

Proof.

Define the TM M_p, which on any input w, outputs $p(w)$.

Then $\langle M_p, d_p(x)\rangle$ is a description of x. Hence,

$$K(x) \leq |\langle M_p, d_p(x)\rangle| = 2|\langle M_p\rangle| + K_p(x) + 1 = K_p(x) + c,$$

where $c = 2|\langle M_p\rangle| + 1$. $\square$