Kolmogorov Complexity

Fangzhou Zhai

2014.12.5

Outline

- A. N. Kolmogorov.
- Kolmogorov Complexity
- Kolmogorov Complexity and Data Compression

Andrey Nikolaevich Kolmogorov

Andrey Nikolaevich Kolmogorov

- Kolmogorov (1903-1987) made significant contributions to probability theory and statistics, the theory of dynamical systems, intuitionistic logic, geometry and topology, the theory of functions and functional analysis, classical mechanics, the theory of turbulence, and information theory.

Andrey Nikolaevich Kolmogorov

- Kolmogorov (1903-1987) made significant contributions to probability theory and statistics, the theory of dynamical systems, intuitionistic logic, geometry and topology, the theory of functions and functional analysis, classical mechanics, the theory of turbulence, and information theory.
- Information theory own him a lot for his work in the axiomatization of probability theory and stochastic processes.

Andrey Nikolaevich Kolmogorov

- Kolmogorov (1903-1987) made significant contributions to probability theory and statistics, the theory of dynamical systems, intuitionistic logic, geometry and topology, the theory of functions and functional analysis, classical mechanics, the theory of turbulence, and information theory.
- Information theory own him a lot for his work in the axiomatization of probability theory and stochastic processes.
- J. Wolfowits, 1963: *I came to the USSR with the specific purpose of finding out whether Andrey Nikolaevich Kolmogorov is an individual or an institution.*

Andrey Nikolaevich Kolmogorov

- Kolmogorov (1903-1987) made significant contributions to probability theory and statistics, the theory of dynamical systems, intuitionistic logic, geometry and topology, the theory of functions and functional analysis, classical mechanics, the theory of turbulence, and information theory.
- Information theory own him a lot for his work in the axiomatization of probability theory and stochastic processes.
- J. Wolfowits, 1963: *I came to the USSR with the specific purpose of finding out whether Andrey Nikolaevich Kolmogorov is an individual or an institution.*

A. N. Kolmogorov

Universal Turing Machine

Universal Turing Machine

A universal Turing Machine $\mathcal{U}$ is a computing device that can execute unambiguous instructions (or program, or input) p, and possibly yields an output string $\mathcal{U}(p)$.

Universal Turing Machine

A universal Turing Machine $\mathcal{U}$ is a computing device that can execute unambiguous instructions (or program, or input) p, and possibly yields an output string $\mathcal{U}(p)$.
A universal Turing Machine may not have an output, and may even never halt.

Universal Turing Machine

A universal Turing Machine $\mathcal{U}$ is a computing device that can execute unambiguous instructions (or program, or input) p, and possibly yields an output string $\mathcal{U}(p)$.
A universal Turing Machine may not have an output, and may even never halt.Consider instructions "output all natural numbers"

Universal Turing Machine

A universal Turing Machine $\mathcal{U}$ is a computing device that can execute unambiguous instructions (or program, or input) p, and possibly yields an output string $\mathcal{U}(p)$.
A universal Turing Machine may not have an output, and may even never halt.Consider instructions "output all natural numbers" and "list all natural numbers then output '3'".

The Notion of Information

The Notion of Information

Definition (Entropy)

For any random variable X, the entropy of X is defined as

$$H(X) := \mathbb{E}[log \frac{1}{p(x)}]$$

The Notion of Information

Definition (Entropy)

For any random variable X, the entropy of X is defined as

$$H(X) := \mathbb{E}[log \frac{1}{p(x)}]$$

Entropy is defined for random variables and relies on the distribution.

Definition (Entropy)

For any random variable X, the entropy of X is defined as

$$H(X) := \mathbb{E}[log \frac{1}{p(x)}]$$

Entropy is defined for random variables and relies on the distribution. However, the notion of information is also meaningful at the instance level.

The Notion of Information

Definition (Entropy)

For any random variable X, the entropy of X is defined as

$$H(X) := \mathbb{E}[log \frac{1}{p(x)}]$$

Entropy is defined for random variables and relies on the distribution. However, the notion of information is also meaningful at the instance level.

Example

$X = 111111111111111111111111111111111111111111$

The Notion of Information

Definition (Entropy)

For any random variable X, the entropy of X is defined as

$$H(X) := \mathbb{E}[log \frac{1}{p(x)}]$$

Entropy is defined for random variables and relies on the distribution. However, the notion of information is also meaningful at the instance level.

Example

$X = 111111111111111111111111111111111111111111$
$Y = 100101011101010011001000101010010111101010$

X carries as much information as "42 1s" while there is no immediate short description of Y.

Kolmogorov Complexity I

A description wraps all the information of a string. The minimal description length of a string is thus a measurement of its complexity, or the information it carries.

Kolmogorov Complexity I

A description wraps all the information of a string. The minimal description length of a string is thus a measurement of its complexity, or the information it carries.

Definition (Kolmogorov Complexity)

The Kolmogorov Complexity of $x \in \{0,1\}^*$ is defined as

$$K(x) := min_{p:\mathcal{U}(p)=x} l(p)$$

where $l(p)$ denotes the length of program p.

Kolmogorov Complexity I

A description wraps all the information of a string. The minimal description length of a string is thus a measurement of its complexity, or the information it carries.

Definition (Kolmogorov Complexity)

The Kolmogorov Complexity of $x \in \{0,1\}^*$ is defined as

$$K(x) := min_{p:\mathcal{U}(p)=x} l(p)$$

where $l(p)$ denotes the length of program p.

Kolmogorov complexity is also referred to as "absolute information" in comparison with entropy.

Kolmogorov Complexity II

Kolmogorov Complexity II

- Any initial segment of an algebraic number has low Kolmogorov complexity.

Kolmogorov Complexity II

- Any initial segment of an algebraic number has low Kolmogorov complexity. e.g., first million digits of the greatest root of $x^2 = 2$.

Kolmogorov Complexity II

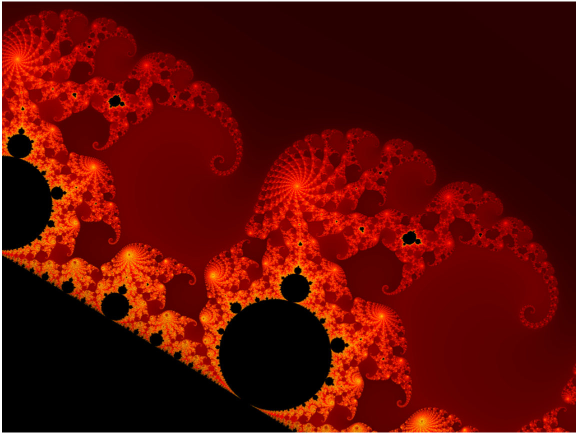
- Any initial segment of an algebraic number has low Kolmogorov complexity. e.g., first million digits of the greatest root of $x^2 = 2$.
- The Kolmogorov Complexity of a string that consists of n ones is at most $\log n + c$.

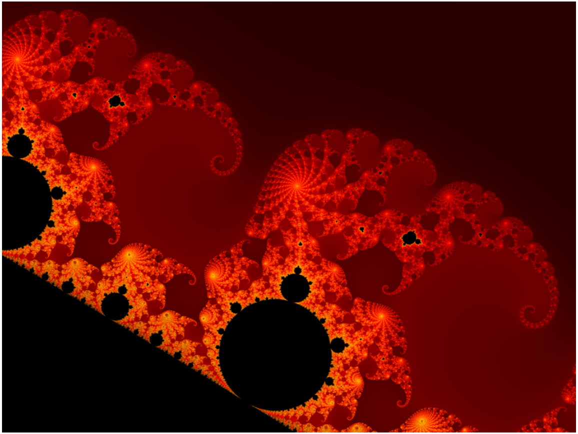
Kolmogorov Complexity II

- Any initial segment of an algebraic number has low Kolmogorov complexity. e.g., first million digits of the greatest root of $x^2 = 2$.
- The Kolmogorov Complexity of a string that consists of n ones is at most $\log n + c$.
- Losslessly compressed data can be seen as a description.

Kolmogorov Complexity III

Kolmogorov Complexity III





This 2.36MB fractal picture is generated by one complex polynomial. The description length of this picture is not much larger than that of the polynomial.

Upper Bound

Upper Bound

Theorem

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

$$K(x) \leq l(x) + c$$

Upper Bound

Theorem

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

$$K(x) \leq l(x) + c$$

Proof.

Consider program "print string x". □

Kolmogorov Complexity and Data Compression I

The number of strings that have short descriptions are bounded:

The number of strings that have short descriptions are bounded:

Theorem (Lower Bound)

$$|\{x|K(x) < k\}| < 2^k$$

The number of strings that have short descriptions are bounded:

Theorem (Lower Bound)

$$|\{x|K(x) < k\}| < 2^k$$

Proof.

There are only $2^k - 1$ descriptions of length less than k. □

Corollary

For each n, at least one string of length n is of Kolmogorov Complexity at least n.

Corollary

*For each n, at least one string of length n is of Kolmogorov Complexity at least n. This string is **not compressible**.*

Corollary

*For each n, at least one string of length n is of Kolmogorov Complexity at least n. This string is **not compressible**.*

A string is random, if it cannot be compressed. Intuitively, 111111111111111111 is not random, while 100101001011110110 is sort of random.

Randomness

Corollary

*For each n, at least one string of length n is of Kolmogorov Complexity at least n. This string is **not compressible**.*

A string is random, if it cannot be compressed. Intuitively, 1111111111111111111 is not random, while 10010100101111011010 is sort of random.

Definition (Random String)

x is **random** if $K(x) \geq l(x)$.

Kolmogorov Complexity and Data Compression II

The probability that a uniformly random binary string has low Kolmogorov complexity is fairly small:

The probability that a uniformly random binary string has low Kolmogorov complexity is fairly small:

Theorem

Let $X_i \sim$ *iid* $B(\frac{1}{2})$, *then*

$$P[K(X^n) < n - k] < 2^{-k}$$

Kolmogorov Complexity and Data Compression II

The probability that a uniformly random binary string has low Kolmogorov complexity is fairly small:

Theorem

Let $X_i \sim$ iid $B(\frac{1}{2})$, then

$$P[K(X^n) < n - k] < 2^{-k}$$

Proof.

$$P[K(X^n) < n - k] = \sum_{x : K(x) < n-k} P(x) =$$

Kolmogorov Complexity and Data Compression II

The probability that a uniformly random binary string has low Kolmogorov complexity is fairly small:

Theorem

Let $X_i \sim$ *iid* $B(\frac{1}{2})$, *then*

$$P[K(X^n) < n - k] < 2^{-k}$$

Proof.

$$P[K(X^n) < n - k] = \sum_{x:K(x)<n-k} P(x) = |\{x|K(x) \leq n-k\}| \cdot 2^{-n}$$

Kolmogorov Complexity and Data Compression II

The probability that a uniformly random binary string has low Kolmogorov complexity is fairly small:

Theorem

Let $X_i \sim iid\ B(\frac{1}{2})$, then

$$P[K(X^n) < n - k] < 2^{-k}$$

Proof.

$$P[K(X^n) < n - k] = \sum_{x:K(x)<n-k} P(x) = |\{x|K(x) \leq n-k\}| \cdot 2^{-n}$$

$$< 2^{n-k} \cdot 2^{-n} =$$

The probability that a uniformly random binary string has low Kolmogorov complexity is fairly small:

Theorem

Let $X_i \sim$ iid $B(\frac{1}{2})$, then

$$P[K(X^n) < n - k] < 2^{-k}$$

Proof.

$$P[K(X^n) < n - k] = \sum_{x:K(x)<n-k} P(x) = |\{x|K(x) \leq n - k\}| \cdot 2^{-n}$$

$$< 2^{n-k} \cdot 2^{-n} = 2^{-k}$$

□

Kolmogorov Complexity and Data Compression III

Corollary

For files of size n, only a 2^{-k} fragment can be potentially compressed by k bits.

Corollary

For files of size n, only a 2^{-k} fragment can be potentially compressed by k bits.

This is really a tiny fragment: for files of size $100kB$,

Corollary

For files of size n, only a 2^{-k} fragment can be potentially compressed by k bits.

This is really a tiny fragment: for files of size $100kB$, only 2^{-8192} of them can be possibly compressed by $1kB$,

Corollary

For files of size n, only a 2^{-k} fragment can be potentially compressed by k bits.

This is really a tiny fragment: for files of size $100kB$, only 2^{-8192} of them can be possibly compressed by $1kB$, and $2^{-409600}$ of them can be possibly compressed to half its original size.

Corollary

For files of size n, only a 2^{-k} fragment can be potentially compressed by k bits.

This is really a tiny fragment: for files of size $100kB$, only 2^{-8192} of them can be possibly compressed by $1kB$, and $2^{-409600}$ of them can be possibly compressed to half its original size. Compression softwares are, nevertheless, running happily, because the files we use them to compress, luckily, belongs to this tiny fragment.

Summary

Summary

- The Kolmogorov Complexity of a string x is defined as the length of minimal description, and is thus upper bounded by the length of x.

Summary

- The Kolmogorov Complexity of a string x is defined as the length of minimal description, and is thus upper bounded by the length of x.
- By far the majority of the strings have their Kolmogorov complexity close to their length, thus cannot be compressed much. Compression is possible in practice because our files lie in the tiny bit that is compressible.

Further Readings

Further Readings

- Kolmogorov Complexity and Entropy.
  http://homepages.cwi.nl/~paulv/papers/info.pdf

Further Readings

- Kolmogorov Complexity and Entropy.
  http://homepages.cwi.nl/~paulv/papers/info.pdf
- The undecidability of $K(x)$.

Further Readings

- Kolmogorov Complexity and Entropy.
  http://homepages.cwi.nl/~paulv/papers/info.pdf
- The undecidability of $K(x)$.
- Time bounded Kolmogorov complexity $K^t(x)$.

Further Readings

- Kolmogorov Complexity and Entropy.
  http://homepages.cwi.nl/~paulv/papers/info.pdf
- The undecidability of $K(x)$.
- Time bounded Kolmogorov complexity $K^t(x)$.
- The magic number $\Omega = \sum_{p:\mathcal{U}(p) halts} 2^{-l(p)}$.

Further Readings

- Kolmogorov Complexity and Entropy.
  http://homepages.cwi.nl/~paulv/papers/info.pdf
- The undecidability of $K(x)$.
- Time bounded Kolmogorov complexity $K^t(x)$.
- The magic number $\Omega = \sum_{p:\mathcal{U}(p) halts} 2^{-l(p)}$.
- Philosophical thoughts. Occam's Razor.

Further Readings

- Kolmogorov Complexity and Entropy. http://homepages.cwi.nl/~paulv/papers/info.pdf
- The undecidability of $K(x)$.
- Time bounded Kolmogorov complexity $K^t(x)$.
- The magic number $\Omega = \sum_{p:\mathcal{U}(p) halts} 2^{-l(p)}$.
- Philosophical thoughts. Occam's Razor.
- Worship Andrey N. Kolmogorov. http://www.kolmogorov.com/. http://theor.jinr.ru/~kuzemsky/ankolmogbio.html.

Thanks for your attention.