Kolmogorov complexity

Cheng Shichao

Institute of Software, Chinese Academy of Sciences, Beijing

August 28, 2021

Background of KC

The notion of algorithmic complexity (algorithmic entropy) appeared in the 1960s in between the theory of computation, probability theory, and information theory.

Kolmogorov complexity:

For a string x consider all its descriptions, the length of the shortest string y among them is called the Kolmogorov complexity of x with respect to D: $C_D(x) = \min\{l(y) \mid D(y) = x\}$.

Martin-Löf random

X is Martin-Löf random iff there is a constant c such that $K(X{\upharpoonright}_n) \geq n - c$ for all n, where $X{\upharpoonright}_n$ denotes the initial segment of X of length n and K denotes prefix-free Kolmogorov complexity.

Two examples

Formal language theory:

Show $L = \{a^k b^k | k > 0\}$ not regular.

Combinatorics:

There is a tournament (complete directed graph) T of n players that contains no large transitive subtournaments ($> 1 + 2\log n$).

Pumping lemma

Let L be a regular language. Then there exists an integer $p \geq 1$ depending only on L such that every string w in L of length at least p (p is called the "pumping length") can be written as $w = xyz$ (i.e., w can be divided into three substrings), satisfying the following conditions:

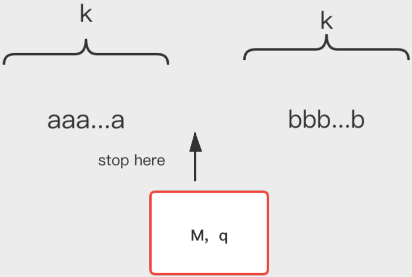
- $|y| \geq 1$
- $|xy| \leq p$
- $\forall n \geq 0,\ xy^n z \in L$

Note:

The pumping lemma is a necessary condition of regular language, which is often used to prove that a particular language is non-regular.

Proof using KC

By contradiction, assume that DFA M accepts L. Choose k so that $C(k) \gg 2|M|$. Simulate M.



$C(k) < |M| + q + O(1) < 2|M|$. Contradiction.

Tournament proof by KC

There is a tournament (complete directed graph) T of n players that contains no large transitive subtournaments ($> 1 + 2\log n$).

- Choose a random T.
- One bit codes an edge. $C(T) \geq n(n-1)/2$.
- If there is a large transitive subtournament, then a large number of edges are given for free!
- $C(T) < n(n-1)/2 - (\text{subgraph\_edges}) + \text{overhead}$

Let k represent the number of nodes in this subgraph. We know that subgraph_edges $= k(k-1)/2$ and overhead $= k * \log n$, so to avoid contradiction, $k < 2\log n + 1$.

Current research

The goal is to give an answer to the problem appears in Miller and Nies (2006).

For each X there exists a Martin-Löf random Y such that $X \leq_K Y$?

Miller and Nies (2006) and Miller and Yu (2011) also asked:

Is there a $\leq_K$-maximal and/or a $\ll_K$-maximal real?

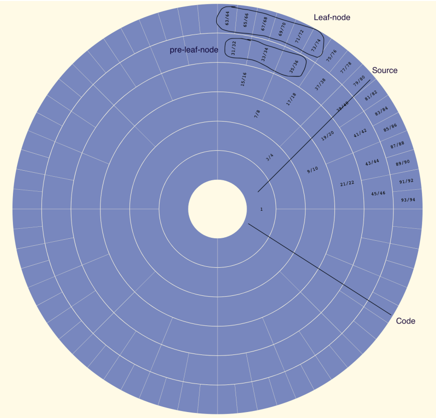
Definition:

$A \leq_{rK} B$ if $\exists d \forall n.\ K(A{\restriction}_n \mid B{\restriction}_n) < d$.

$A \leq_K B$ if $\exists d \forall n.\ K(A{\restriction}_n) \leq K(B{\restriction}_n) + d$.

$A \ll_K B$ if $\lim_n K(A{\restriction}_n) - K(B{\restriction}_n) = \infty$.

Current result



There exists a constructing strategy for spoiler to defeat the greedy coder with no more than $n(n-1)+1+\frac{(n-4)(n-3)(n-2)}{3}$ hits in the tree of height n.

- 踏实勤奋
- 集思广益
- 劳逸结合

…

博学笃志，格物明德