Theoretical Computer Science 607 (2015) 435–445

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Kolmogorov structure functions for automatic complexity

Bjørn Kjos-Hanssen [1]

ARTICLE INFO

Article history:
Received 20 February 2015
Accepted 22 May 2015
Available online 6 June 2015

Keywords:
Kolmogorov structure function
Automatic complexity
Kolmogorov complexity

ABSTRACT

For a finite word w of length n we define and study the Kolmogorov structure function h_w for nondeterministic automatic complexity. We prove upper bounds on h_w that appear to be quite sharp, based on numerical evidence.

1. Introduction

Shallit and Wang [4] introduced automatic complexity as a computable alternative to Kolmogorov complexity. They considered deterministic automata, whereas Hyde and Kjos-Hanssen [3] studied the nondeterministic case, which in some ways behaves better. Unfortunately, even nondeterministic automatic complexity is somewhat inadequate. The string 00010000 has maximal nondeterministic complexity, even though intuitively it is quite simple. One way to remedy this situation is to consider a structure function analogous to that for Kolmogorov complexity.

The latter was introduced by Kolmogorov at a 1973 meeting in Tallinn and studied by Vereshchagin and Vitányi [6] and Staiger [5].

The Kolmogorov complexity of a finite word w is roughly speaking the length of the shortest description w^* of w in a fixed formal language. The description w^* can be thought of as an optimally compressed version of w. Motivated by the non-computability of Kolmogorov complexity, Shallit and Wang studied a deterministic finite automaton analogue.

Definition 1. (See Shallit and Wang [4].) The *automatic complexity* of a finite binary string $x = x_1 \dots x_n$ is the least number $A_D(x)$ of states of a deterministic finite automaton M such that x is the only string of length n in the language accepted by M.

Hyde and Kjos-Hanssen [3] defined a nondeterministic analogue:

Definition 2. The nondeterministic automatic complexity $A_N(w)$ of a word w is the minimum number of states of an NFA M, having no ϵ-transitions, accepting w such that there is only one accepting path in M of length $|w|$.

The minimum complexity $A_N(w) = 1$ is only achieved by words of the form a^n where a is a single letter.

E-mail address: bjoernkh@hawaii.edu.

[1] This work was partially supported by a grant from the Simons Foundation (#315188 to Bjørn Kjos-Hanssen). The author also acknowledges the support of the Institute for Mathematical Sciences of the National University of Singapore during the workshop on *Algorithmic Randomness*, June 2–30, 2014.

http://dx.doi.org/10.1016/j.tcs.2015.05.052
0304-3975/

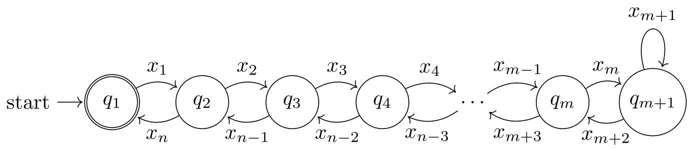

Fig. 1. A nondeterministic finite automaton that only accepts one string $x = x_1x_2x_3x_4 \dots x_n$ of length $n = 2m + 1$.

Definition 3. Let $n = 2m + 1$ be a positive odd number, $m \geq 0$. A finite automaton of the form given in Fig. 1 for some choice of symbols $x_1, \dots, x_n$ and states $q_1, \dots, q_{m+1}$ is called a *Kayleigh graph.*[2]

Theorem 4. *(See Hyde [2].) The nondeterministic automatic complexity $A_N(x)$ of a string x of length n satisfies*

$$A_N(x) \leq b(n) := \lfloor n/2 \rfloor + 1.$$

Proof. If n is odd, then a Kayleigh graph witnesses this inequality. If n is even, a slight modification suffices, see [2]. $\square$

The structure function of a string x is defined by $h_x(m) = \min\{k$: there is a k-state NFA M which accepts at most 2^m strings of length $|x|$ including $x\}$. In more detail:

Let

$$S_x = \{(q, m) \mid \exists\, q\text{-state NFA } M, x \in L(M) \cap \Sigma^n, |L(M) \cap \Sigma^n| \leq b^m\}.$$

Then S_x has the upward closure property

$$q \leq q', m \leq m', (q, m) \in S_x \implies (q', m') \in S_x.$$

From S_x we can define the structure function h_x and the dual structure function h_x^*.

Definition 5. (See Vereshchagin, personal communication, 2014, inspired by [6].) In an alphabet Σ containing b symbols, we define

$$h_x^*(m) = \min\{k : (k, m) \in S_x\} \quad \text{and}$$
$$h_x(k) = \min\{m : (k, m) \in S_x\}.$$

Remark 6. On the one hand, h mimics the structure function as defined by Kolmogorov. On the other hand, h^* has a natural domain $[0, n]$ whereas the domain of h is initially $[1, \infty)$, until some upper bound on the automatic complexity is proved, at which point it becomes $[1, \lfloor n/2 \rfloor + 1]$. One often prefers that a function has a simple domain and a complicated range rather than the other way around, e.g., consider the case of the range of a computable function on $\mathbb{N}$ (which is only computably enumerable).

History of the structure function Kolmogorov first introduced the structure function in a talk at The Third International Symposium on Information Theory, June 18–23, 1973, Tallinn, Estonia, Soviet Union. The meeting coincided with a Nixon/Brezhnev meeting in the U.S. Kolmogorov was born in 1903 hence 70 years old at the time. The results were not published until they appeared as an abstract of a talk for the Moscow Mathematical Society [1] in *Uspekhi Mat. Nauk* in the Communications of the Moscow Mathematical Society, page 155 (in the Russian edition, not translated into English). The talk was given on April 16, 1974 and was entitled “Complexity of algorithms and objective definition of randomness”.

2. Basic properties

Definition 7. The entropy function $\mathcal{H} : [0, 1] \to [0, 1]$ is given by

$$\mathcal{H}(p) = -p \log_2 p - (1 - p) \log_2(1 - p).$$

Remark 8. Throughout the paper, log (with no subscript) denotes either the natural logarithm $\ln = \log_e$, or $\log_b$ where the value of b is immaterial.

[2] The terminology is a nod to the more famous Cayley graphs as well as to Kayleigh Hyde's first name.