Electronic Colloquium on Computational Complexity, Report No. 15 (2000)

Multi-conditional Descriptions and Codes in Kolmogorov Complexity

Andrej Muchnik, Alexej Semenov*

January 27, 2000

It is well known, that Andrei Kolmogorov in Russia and R. Solomonoff in the USA discovered the basic notion and facts on complexity of finite objects independently in mid-60s. In 60s and 70s the research of the area was conducted mostly by students of Kolmogorov and members of the Moscow mathematical community based on Moscow State (Lomonosov) University: L. Levin, V. Vyugin, A. Zvonkin, N. Petri, P. Gács. Chaitin considerably contributed to popularity of the theory in the West; Solovay's manuscript influenced research in Russia as well. From the end of 70s till his death in 1987 Kolmogorov was the Chair of Department of Mathematical Logic. Now his former student Vladimir Uspensky, who also works in the field of Kolmogorov complexity, heads it. In the first steps of his work as the chair of our department Professor Kolmogorov proposed to the first author of this paper to start a seminar on complexity. (The seminar is called Kolmogorov Seminar now.) Starting from the very beginning the seminar included research reports and surveys on practical programming and computational complexity as well. In his first talks at the seminar Kolmogorov introduced a program of research of frequency approach to randomness and some other topics. This program was implemented in works of V. Vovk, A. Shen, N. Vereshagin, An. Muchnik and others. In the following years A. Shen and N. Vereshagin joined the leadership of the seminar and, after the death of Kolmogorov, became the major moving force of it. Professor Uspensky contributed a lot to its development. In Kolmogorov and Uspensky lecture at Bernoulli congress the results obtained by the Moscow group in the period of Kolmogorov presence are summarized. Our work presented here is the result of general cooperation and many specific discussions of the members of Kolmogorov seminar which is 20 years old now.

Let us outline now the main results of the paper. A major relation of the Kolmogorov complexity involve transformation of a condition y into an object x using a description (program) p. The (Kolmogorov) entropy $K(x|y)$ is the minimal size (length) of such p, $K(x)$ is $K(x|\text{a fixed object})$. Let us have two conditions y_1 and y_2 now. What is the size of program p, using which we can transform into x both conditions y_1 and y_2? It is easy to construct

*Both authors work at the Institute of New Technologies of Education, Moscow, N.Radischevskaya, 10.

ISSN 1433-8092

p of size $K(x|y_1) + K(x|y_2)$. It is also evident that p cannot be smaller then $\max\{K(x|y_1), K(x|y_2)\}$. We prove that this bound can be reached up to (usual for the theory) additive term $const \cdot \log K(x)$. Moreover, we can choose p which contains little information in addition to x: $K(p|x)$ is less then $const \cdot \log K(x)$. Let us consider now all conditions y with limited entropy $K(x|y)$. We can construct a set of programs of the size limited by the same bound such as the cardinality of the set is limited by a polynomial of $K(x)$. Some important constructions of the Kolmogorov theory can be interpreted in a calculus of problems. In particular, the construction of p by given y_1 and y_2 corresponds to a realization of formula $Y_1 \vee Y_2 \to X$. It is interesting to compare or result with a theorem from the paper [2], where a problem corresponding to the formula $X \leftrightarrow Y$ is considered. It is proved there that the entropy of p which transforms both x into y and y into x is close to maximum of entropies of programs that transform x into y and y into x. We can obtain a similar result from our theorem mentioned above considering problems of transforming x and y into $\langle x, y \rangle$. The key concept of our paper is the idea of code. We say that object r is a code of object x with a condition y provided $K(x|y, r)$ is small. Given a code r one can easily construct a description p of x relative to condition y with the size of p close to $K(r)$.

Minimal codes in the Kolmogorov entropy theory

Complete proofs for these results were presented in September, 1999 at the Kolmogorov Seminar at Moscow State University.

Define Kolmogorov entropy of x conditional to y (for constructible objects x and y) as length of the shortest program that outputs x given the input y.[1] We suppose that programs are written in some optimal programming language. Here we consider not only sizes of programs converting y into x, but also some other properties. We prove that for any y_1, y_2 and x there exists a program p satisfying the following conditions. First, $p(y_1) = p(y_2) = x$; second, the length of p is less than $\max\{K(x|y_1), K(x|y_2)\} + const \cdot \log K(x)$; third, $K(p|x) < const \cdot \log K(x)$.[2] The simplest program satisfying the first condition has length $K(x|y_1) + K(x|y_2)$. K. Ju. Gorbunov proved in [1] that the latter bound cannot be improved for some objects y_1, y_2 and x with absolute entropy exponential in $K(x|y_1) + K(x|y_2)$. Therefore the $const \cdot K(x)$ difference is inevitable. It also cannot be avoided in the following claim: the minimal information sufficient for finding x when y is known can be obtained from x alone. We stress that these bounds depend only on the entropy of x, not on the entropies of the conditions y. We have an interesting corollary: while there exist exponentially many programs of a given length, we can obtain x from all y's with given $K(x|y)$ using only polynomially many (in $K(x)$) programs of length $K(x|y)$. We will also give a lower bound for the cardinality of such "universal" set of programs.

[1] The results in this article are equally true for both *simple* and *prefix* entropy.

[2] By definition, absolute entropy $K(w) = K(w|\Lambda)$, where Λ is some fixed object (for example, an empty string).

Let us explain the informal meaning of the word "code" in this article's name. The object p is called a *code* for x with known y if $K(x|y,p)$ is sufficiently small. It is obvious that, given p, we can construct a program q that outputs x given an input y and that has length not much more than $K(p)$. It is often convenient to consider codes instead of programs.

It is interesting to compare the results for the formula $A \vee B \to C$, and the results for $A \leftrightarrow B$. In [2] it is proven that minimal entropy of program obtaining y from z and vice versa approximately equals the maximum of minimal entropy of program obtaining z from y and minimal entropy of program obtaining y from z. This result can be deduced from ours by considering a problem of obtaining $\langle y, z\rangle$ from y and z. However, this method is accurate only up to $\mathit{const} \cdot \log K(\langle y, z\rangle)$ while the proof in [2] is accurate up to *const*. There is another difference. There exists a minimal program p obtaining x from y and a minimal program q obtaining x from z such that $\langle p, q\rangle$ contains no information about any minimal program obtaining x from both y and z. A pair of minimal programs p and q obtaining x from y and y from x respectively, however, always contains much common information with a suitable minimal program obtaining both y from x and x from y.

Theorem 1 (An. A. Muchnik) *There exists a number c and a partial computable function $\lambda xl.F(x,l)$ such that the following is true: if $K(x) \leq l$ then*

i) $F(x,l)$ is a set with cardinality less than $\frac{cl}{\log l}$ consisting of strings with length l (this is a universal code set for x);

ii) for any y_1, y_2, there exist p_1, p_2 such that for some $p \in F(x,l)$ p's prefix of length $K(x|y_1)$ equals p_1, p's prefix of length $K(x|y_2)$ equals p_2 and

$$K(x|y_1,p_1) < c\log l, \qquad K(x|y_2,p_2) < c\log l.$$

From i) and ii) it follows that $K(p_1|x) < \mathit{const} \cdot \log l$ and $K(p_2|x) < \mathit{const} \cdot \log l$.

Proof. For every number l define a number r_l (we will have further $K(x) \leq l$, $r_l = \lceil \log F(x,l) \rceil$). We will give exact value for r_l later on; sometimes we will omit the subscript l. By L and R denote the sets of binary strings of length l and r respectively. Consider the space of functions $L \times R \to L$ and uniform probability distribution over this space. This distribution help us prove the existence of a function f satisfying the condition (1) (shown below) for every $n < l - \log l$. (We will have further $n = K(x|y)$.) Let M be the set of binary strings of length $m = n + \lceil \log l \rceil$. By $\varphi(z,\rho)$ denote the prefix of $f(z,\rho)$ of length m. Consider the condition:

$$\forall B \subset M \quad |B| \leq 2^n \Rightarrow \left|\{z\colon |\{\rho\colon \varphi(z,\rho) \in B\}| \geq 2^{r-1}\}\right| < |B|. \qquad (1)$$

Consider the probability of the following: (1) does not hold for a randomly chosen f. We want to bound this probability from above. First, fix some values of n, $z \in L$, $\rho \in R$ and $B \subset M$, where $|B| \leq 2^n$. Now $\varphi(z,\rho) \in B$ with probability $\frac{|B|}{|M|} \le \frac{2^n}{2^m} \le \frac{1}{l}$. Now fix values of n, z, B and R', where $|R'| \ge |R|/2$. Now the event $\forall \rho \in R' \quad \varphi(z,\rho) \in B$ has probability less than $l^{-|R|/2}$. So for fixed n, z and B we have

$$\Pr\left(|\{\rho\colon \varphi(z,\rho) \in B\}| \ge 2^{r-1}\right) \le 2^{|R|} \cdot l^{-|R|/2} < 2^{-\,const \,\cdot\, \log l \cdot |R|}.$$

Now fix n, $s \le 2^n$, $B \subset M$, where $|B| = s$, and $Z \subset L$, where $|Z| = s$, and get

$$\Pr\left(\forall z \in Z \quad |\{\rho\colon \varphi(z,\rho) \in B\}| \ge 2^{r-1}\right) \le 2^{-\,const \,\cdot\, \log l \cdot |R| \cdot s}.$$

Multiplying this by the number of possible pairs (B, Z), we get the upper bound on the probability of the following event not happening:

$$\forall B \subset M \quad |B| = s \Rightarrow \left|\left\{z\colon\ |\{\rho\colon \varphi(z,\rho) \in B\}| \ge 2^{r-1}\right\}\right| < s$$

(for fixed values of n and s). This upper bound is $2^{-\,const \,\cdot\, \log l \cdot |R| \cdot s} \cdot 2^{ms} \cdot 2^{ls} \le \left(2^{-\,const \,\cdot\, \log l \cdot |R| + 2l}\right)^s$. Letting $|R|$ be $\frac{cl}{\log l}$, we achieve

$$2^{-\,const \,\cdot\, \log l \cdot |R| + 2l} < \frac{1}{2l}.$$

Therefore,

$$\sum_s \left(2^{-\,const \,\cdot\, \log l \cdot |R| + 2l}\right)^s < 1/l.$$

This sum bounds the desired probability from above for a fixed value of n. We have $n < l$, so the probability of (1) being false for some n is less than 1. We defined $r_l = \log |R|$ and proved that for each l there exists a function f satisfying (1). Since we can effectively verify whether (1) is true or not, we can find f by trying every function possible.

By $\nu_l(u)$ denote the first program t of length l such that $t(\Lambda) = u$ (if no such program exist, $\nu_l(u)$ is undefined). We will write νu instead of $\nu_l(u)$ where l is clear from the context. Suppose we find, by enumerating the entropy K from above, that $K(x) \le l$. Let f be the function satisfying (1) with the parameter l. Define $F(x, l)$ as the set of strings $f(x, \rho)$ for all ρ. Let us check the condition ii) of the Theorem. Let y be one the objects y_1 and y_2; $n = K(x|y)$. Let d be an auxiliary parameter whose value we will specify later. We will use the numbers r and m, the sets R and M and the function φ defined earlier. Define

$$D = \{u\colon K(u) \le l \quad \& \quad K(u|y) \le n\}.$$

We know that $x \in D$. Clearly $|D| \le 2^{n+1}$. For every $q \in M$ define the following subset of D:

$$E^q = \{u\colon \exists \rho \in R \quad q = \varphi(\nu u, \rho)\}.$$

Define

$$G = \{q\colon |E^q| > 2^d\} \subset M.$$

Evidently $|G| < |D| \cdot |R|/2^d < 2^{n+1+r-d}$. Define

$$H = \left\{u\colon |\{\rho\colon \varphi(\nu u, \rho) \in G\}| \ge 2^{r-1}\right\}.$$

We will specify the parameter d to be greater than r, so $|G| < 2^n$. Using (1), we get $|H| < |G|$. Note that D is uniformly enumerable w.r.t. l, n and y, E^q is uniformly enumerable w.r.t. l, n, y and q, G is uniformly enumerable w.r.t. l, n, y and d, H is uniformly enumerable w.r.t. l, n, y and d. Suppose $x \in H$. Then $K(x|y)$ is less (up to an additive constant) than sum of the entropy of H's enumerator program (conditional to y) and the length of x's position in this enumeration. We have

$$K(x|y) < K(l) + K(n) + K(d) + \log_2 |H| + \mathit{const}$$
$$= K(l) + K(n) + K(d) + n + r - d + \mathit{const}\,.$$

Putting $d = \alpha \log l$ for a sufficiently large constant α, we get

$$K(l) + K(n) + K(d) + n + r - d + \mathit{const} < n.$$

However, $n = K(x|y)$. This contradiction shows that $x \notin H$.

Since we could take either y_1 or y_2 for y, we actually constructed two families of sets: D_1, E_1^q, G_1, H_1 for $y = y_1$ and D_2, E_2^q, G_2, H_2 for $y = y_2$. Since $x \notin H_1$ and $x \notin H_2$, we have

$$|\{\rho\colon \varphi_1(\nu x, \rho) \in G_1\}| < 2^{r-1} \qquad \text{and} \qquad |\{\rho\colon \varphi_2(\nu x, \rho) \in G_2\}| < 2^{r-1}.$$

Since $|R| = 2^r$, there exists $\rho_0 \in R$ such that $\varphi_1(\nu x, \rho_0) \notin G_1$ and $\varphi_2(\nu x, \rho_0) \notin G_2$. Therefore, $|E_1^{\varphi_1(\nu x, \rho_0)}| < 2^d$ and $|E_2^{\varphi_2(\nu x, \rho_0)}| < 2^d$. From the definition of E^q, we have $x \in E_1^{\varphi_1(\nu x, \rho_0)}$ and $x \in E_2^{\varphi_2(\nu x, \rho_0)}$. So, the entropy of x conditional to $\langle y_1, \varphi_1(\nu x, \rho_0)\rangle$ is less (up to an additive constant) than sum of the entropy of enumerator for $E_1^{\varphi_1(\nu x, \rho_0)}$ conditional to $\langle y_1, \varphi_1(\nu x, \rho_0)\rangle$, and the number of x in this enumeration. That is,

$$K(x|y_1, \varphi_1(\nu x, \rho_0)) < K(l) + K(n_1) + \log_2 |E_1^{\varphi_1(\nu x, \rho_0)}| + \mathit{const}$$
$$= K(l) + K(n_1) + d + \mathit{const}\,.$$

Just the same,

$$K(x|y_2, \varphi_2(\nu x, \rho_0)) < K(l) + K(n_2) + d + \mathit{const}\,.$$

Recall that $d = \alpha \log l$. Let p be $f(\nu x, \rho_0)$, p_1 be the p's prefix of length n_1, p_2 be the p's prefix of length n_2. Since $\varphi_1(\nu x, \rho_0)$ is the p's prefix of length m_1, $\varphi_2(\nu x, \rho_0)$ is the p's prefix of length m_2 and $m_1 = n_1 + \lceil \log l \rceil$, $m_2 = n_2 + \lceil \log l \rceil$, for some constant c we have $K(x|y_1, p_1) < c \log l$ and $K(x|y_2, p_2) < c \log l$. □

Note that the number of conditions y in the proven theorem can be made polynomial in $K(x)$ instead of two.

Interestingly, the semi-lattice introduced in [3] always contains the difference of two elements, but may not contain their intersection.

The following theorem shows that the code set constructed in the previous theorem cannot be made significantly smaller.

Theorem 2 (An. A. Muchnik) *For every number α there exists a number c such that for any binary string x and any code set P of cardinality less than $\frac{K(x)}{c\log K(x)}$ containing codes of length $\lceil K(x)/2\rceil$ there exists a string y such that the following conditions hold:*

i) $K(y) < cK(x)$;

ii) $K(x|y) < K(x)/2$;

iii) $\forall p \in P \quad K(x|y,p) > \alpha \log K(x)$.

Proof. Let c be a sufficiently large number. Suppose $P = \{p_1, \dots, p_j\}$ and $j < \frac{K(x)}{c\log K(x)}$. By v_i denote the prefix of p_i of length $c\log K(x)/3$. Let w be a concatenation of $v_1, \dots, v_j$. We have

$$K(w) < jc\log K(x)/3 + const = K(x)/3 + const \qquad \text{and}$$
$$K(x|w) > K(x) - K(w) - const \cdot \log K(x) > K(x)/2$$

for sufficiently large $K(x)$. Consider the value of $K(x|wz)$ where z is a prefix of x. When the length of z changes by 1 the value of $K(x|wz)$ changes by no more than a constant. Since $K(x|w\Lambda) = K(x|w) > K(x)/2$ and $K(x|wx) < const$, there exists a prefix z_0 of x such that $K(x)/2 > K(x|wz_0) > K(x)/2 - const$. Take wz_0 as y. Check the condition i):

$$K(y) < K(w) + K(z_0) < K(x)/3 + K(x) + const\,.$$

Check the condition ii):

$$K(x|y) = K(x|wz_0) < K(x)/2.$$

Check the condition iii). For each i the string y contains "much" information about p_i, namely,

$$K(p_i|y) < K(x)/2 - c\log K(x)/3 + const \cdot \log K(x).$$

Using this inequality, we obtain for all i

$$\begin{aligned} K(x|y,p_i) &> K(x|y) - K(p_i|y) - const \cdot \log K(x) \\ &> (K(x)/2 - const) - (K(x)/2 - c\log K(x)/3 + const \cdot \log K(x)) \\ &\qquad - const \cdot \log K(x). \end{aligned}$$

Let c be sufficiently large; the condition iii) now follows. □

References

[1] K. Yu. Gorbunov. On a complexity of the formula $((A \vee B) \to C)$. Theoretical Computer Science, v. 207, p. 383–386, 1998.

[2] C. H. Bennet, P. Gács, M. Li, P. M. B. Vitanyi, W. H. Zurek. Information distance. IEEE Transactions on Information Theory, v. 44, no. 4, p. 1407–1423, 1998.

[3] An. Muchnik, A. Romashchenko, A. Shen, N. Vereshchagin. Upper semilattice of binary strings with the relation "x is simple conditional to y". Theoretical Computer Science, to appear, 2000.