

Колмогоровская сложность

Будем рассматривать функции над множеством *двоичных слов* (слов из нулей и единиц) $\{0, 1\}^*$. Длину слова $x \in \{0, 1\}^*$ обозначим $|x|$.

Определение. Пусть фиксирована (частичная) вычислимая функция $D: \{0, 1\}^* \rightarrow \{0, 1\}^*$, называемая *способом описания*. *Колмогоровской сложностью* слова $x \in \{0, 1\}^*$ относительно способа описания D называется число

$$KS_D(x) = \min\{|y| \mid D(y) = x\}.$$

Если же такого y , что $D(y) = x$, не существует, то $KS_D(x) = +\infty$.

1. Докажите, что для любых двух способов описания D_1 и D_2 существует такой способ описания D_0 , что $KS_{D_0}(x) \leq KS_{D_1}(x) + c$ и $KS_{D_0}(x) \leq KS_{D_2}(x) + c$, где c — константа, не зависящая от x .

Теорема Соломонова – Колмогорова (об оптимальном способе описания). Существует такой способ описания D , что для любого другого способа описания D' существует константа c , такая что для всех $x \in \{0, 1\}^*$

$$KS_D(x) \leq KS_{D'}(x) + c.$$

В дальнейшем будем считать, что фиксирован некоторый оптимальный способ описания D , и колмогоровскую сложность будем обозначать просто $KS(x)$ (эта величина определена с точностью до аддитивной константы, зависящей от выбора оптимального способа описания).

2. а) Докажите, что существует константа c , такая что для любого слова x верно $KS(x) \leq |x| + c$.

б) Докажите, что $KS(\underbrace{111 \dots 111}_{n \text{ единиц}}) \leq \log_2 n + c$.

в) Докажите, что $KS(\underbrace{111 \dots 111}_{n \text{ единиц}}) \leq KS(n) + c$ (при вычислении $KS(n)$ считается, что число n представлено в двоичной записи).

г) Оцените сверху величину $KS(\underbrace{111 \dots 111}_{4^{2^n + 3^n} \text{ единиц}})$.

3. а) Докажите, что существует константа c , такая что для любых двух слов x и y верно

$$KS(xy) \leq KS(x) + KS(y) + 2 \log_2 KS(x) + c,$$

где xy — слово, получаемое приписыванием слова y к слову x .

б) Докажите более сильное неравенство

$$KS(xy) \leq KS(x) + KS(y) + \log_2 KS(x) + 2 \log_2 \log_2 KS(x) + c.$$

4. 2) Покажите, что если в качестве описаний использовать слова в четырёхбуквенном алфавите (скажем, конечные последовательности цифр 0, 1, 2, 3), то сложность (измеряемая как длина кратчайшего описания) будет равна половине обычной (с точностью до аддитивной константы).

5. Докажите, что для любого n существует «несжимаемое» слово x длины n , такое что $KS(x) \geq n$.

6. Чему асимптотически равно среднее арифметическое сложностей всех слов длины n ?

7. а) Докажите, что если у слова изменить последний бит, то его сложность изменится не более чем на константу.

б) Докажите, что при изменении одного бита слова его сложность может увеличиться на величину не более чем порядка логарифма его длины.

в) Докажите, что оценка в предыдущем пункте точная: бывает, что при изменении одного бита сложность слова увеличивается на величину порядка логарифма его длины.

8. Пусть D — оптимальный способ описания. Докажите, что D не является всюду определённой функцией.

9. Докажите, что если f вычислима, то для некоторой константы c неравенство $\text{KS}(f(x)) \leq \text{KS}(x) + c$ верно для всех x .

10. Существует ли вычислимая функция $f: \mathbb{N} \rightarrow \{0, 1\}^*$, такая что $\text{KS}(f(n)) \geq n$ для всех n ?

11. а) Докажите, что функция $\text{KS}: x \mapsto \text{KS}(x)$ невычислима. б) Докажите, что у функции KS не существует нетривиальных вычислимых нижних оценок: если вычислимая функция $k: \{0, 1\}^* \rightarrow \mathbb{N}$ такова, что для всех x из области определения k верно $k(x) \leq \text{KS}(x)$, то k ограничена сверху константой.