О взаимной информации двоичного слова и его энтропии

Андрей Мучник*

Наши рассмотрения будут относиться к двум типам энтропии (и условной энтропии) двоичных слов. А именно, простая энтропия (определённая Колмогоровым), которую мы будем обозначать KS, и префиксная энтропия (определённая Левиным), которую мы будем обозначать KP. Когда какое-то утверждение может быть отнесено одновременно к обоим типам энтропии, мы будем писать K вместо KS и KP. Мы предполагаем фиксированными вычислимые изоморфизмы между множеством двоичных слов и множеством натуральных чисел, множеством пар двоичных слов и так далее. Все логарифмы будут браться по основанию 2. Через $\ell(x)$ будет обозначаться длина слова x.

Как известно, обе рассматриваемые энтропии перечислимы сверху и очень близки по величине:

$$KS(x|y) \leqslant KP(x|y) + O(1)\,,$$
$$KP(x|y) \leqslant KS(x|y) + O(\log KS(x|y))\,.$$

Неудивительно, что KS и KP обладают многими одинаковыми свойствами. Однако в некоторых случаях они ведут себя по-разному. Пример свойства из общей теории алгоритмов, существенно различающего KS и KP, приведён в [1]. В настоящей работе исследуются количественные свойства, относительно которых KS и KP ведут себя противоположным образом. Нас будут интересовать следующие вопросы: помогает ли знание слова найти его энтропию и помогает ли знание энтропии слова

*Институт новых технологий, 109004, Москва, ул. Нижняя Радищевская, 10. E-mail: muchnik@lpcs.math.msu.ru, fax: (095)9156963. Работа выполнена при поддержке Российского Фонда Фундаментальных Исследований, гранты N 01-01-00505, N 02-01-22001.

найти само слово? Более точно, мы хотим описать поведение функций $I(x : K(x)) = K(K(x)) - K(K(x)|x)$ и $I(K(x) : x) = K(x) - K(x|K(x))$ (вместо I будем писать IS, когда $K = KS$, и IP, когда $K = KP$). По теореме Колмогорова–Левина о симметрии взаимной информации $K(x) - K(x|y) \approx K(y) - K(y|x) \approx K(x) + K(y) - K(\langle x, y\rangle)$. Если $y = K(x)$, то, как показал Гач,

$$K(\langle x, y\rangle) = K(\langle x, K(x)\rangle) = K(x) + O(1)$$

(минимальный код слова x даёт x, а его длина даёт $K(x)$). Однако взаимная информация симметрична только с точностью до $O\big(\log K(x) + \log K(y)\big)$ (для KS это доказано Колмогоровым и Левиным, а для KP — Гачем). Поскольку $K(K(x)) \leqslant O(\log K(x))$, логарифмическая точность оказывается недостаточной.

Основные результаты, доказательства которых приведены в теоремах 1–4, можно изобразить в виде таблицы.

	KS	KP
$I(K(x) : x)$	Взаимная информация ограничена	Взаимная информация стремится к ∞
$I(x : K(x))$	Взаимная информация не ограничена, но не стремится к ∞	Взаимная информация стремится к ∞

Мы видим, что для KP знание слова помогает найти его энтропию и знание энтропии слова помогает найти само слово. Для KS знание энтропии слова не помогает найти само слово, а знание слова иногда помогает найти его энтропию, а иногда — нет.

Теорема 1. *Функция $IS(KS(x) : x)$ ограничена.*

Доказательство. Рассмотрим достаточно большое натуральное число D. Предположим, что $KS(x|KS(x)) < KS(x) - D$. Пусть y — минимальный код x при известном $KS(x)$. Обозначим через n двоичную запись числа $N = KS(x) - \ell(y)$, в которой каждая цифра повторена дважды. По слову $n01y$ можно эффективно найти x. Действительно, первое слева вхождение 01 отделяет y от n; по n находим $KS(x) - \ell(y)$; зная y, находим $KS(x)$; по y и $KS(x)$ находим x. Следовательно, $KS(x) \leqslant \ell(n01y) + C \leqslant 2\log N + 4 + \ell(y) + C$, где C зависит только от выбора оптимального языка

программирования. То есть, $N = KS(x) - \ell(y) \leqslant 2\log N + C + 4$. Поскольку $N > D$, при больших D получается противоречие. Таким образом, при достаточно большом D имеем $KS(x) \leqslant KS(x|KS(x)) + D$. $\square$

Возможно, читатель удивится, почему то же самое рассуждение нельзя применить к функции $IP(KP(x) : x)$. В этом случае y — минимальный *префиксный* код x при известном $KP(x)$, $n01$ — *префиксный* код N. Казалось бы, $n01y$ — префиксный код x, и получается аналогичный результат. В действительности y является префиксным кодом x только при фиксированном условии. Совокупность же всех кодов, соответствующих функции $KP(\cdot|\cdot)$ при всех условиях, не является префиксным множеством.

Теорема 2. *Функция $IP(KP(x) : x)$ стремится к бесконечности.*

Доказательство. Как известно, перечислимая снизу функция $\mu(x|y) = 2^{-KP(x|y)}$ для каждого y задаёт полумеру на области определения x и является наибольшей среди таких функций с точностью до мультипликативной константы. Соответственно $\mu(x) = 2^{-KP(x)}$. Через μ^t обозначим результат шага t перечисления снизу функции μ. При каждом t область, на которой $\mu^t \neq 0$, конечна.

Определим вычислимую функцию $\nu_k^t(x|n)$ от натуральных аргументов x, n, t и $k > 0$ с рациональными значениями. Для $t = 0$ функция ν тождественно равна нулю, при увеличении t она не убывает и для всех t, n, k выполнено $\sum\limits_x \nu_k^t(x|n) \leqslant 2^{-k}$.

Положим $\nu_k^{t+1}(x|n) = 2^k \cdot \min\{\mu^t(x), 2^{-n}\}$, если $\sum\limits_x \min\{\mu^t(x), 2^{-n}\} \leqslant 2^{-2k}$, и $\nu_k^{t+1}(x|n) = \nu_k^t(x|n)$ в противном случае.

Функция $\nu(x|n) = \sum\limits_k \max\limits_t \nu_k^t(x|n)$ перечислима снизу и $\sum\limits_x \nu(x|n) \leqslant 1$. Поэтому $\nu(x|n) \leqslant O(\mu(x|n))$.

Для любого k существует такое x_0, что $\sum\limits_{x \geqslant x_0} \mu(x) \leqslant 2^{-2k-1}$, и существует такое n_0, что для любого $n \geqslant n_0$ имеет место $\sum\limits_{x < x_0} 2^{-n} \leqslant 2^{-2k-1}$. Отсюда для $n \geqslant n_0$ имеем $\sum\limits_x \min\{\mu^t(x), 2^{-n}\} \leqslant 2^{-2k}$, и поэтому $\nu(x|n) \geqslant 2^k \cdot \min\{\mu(x), 2^{-n}\}$. Для всех достаточно больших x значение $\mu(x)$ не превышает 2^{-n_0}, следовательно, при $n = -\log \mu(x) = KP(x)$ получается $\nu(x|n) \geqslant 2^k \cdot \mu(x)$. Так как $\mu(x|n) \geqslant 2^{-C} \cdot \nu(x|n) \geqslant 2^{k-C} \cdot \mu(x)$, то

$KP(x|KP(x)) \leqslant KP(x) - k + C$, где C зависит только от выбора оптимального языка программирования. $\square$

Теорема 3. *Функция $IP(x : KP(x))$ стремится к бесконечности.*

Доказательство. Доказательство этой теоремы похоже на доказательство предыдущей теоремы.

Определим вычислимую функцию $\nu_k^t(n|x)$ от натуральных аргументов n, x, t и $k > 0$ с рациональными значениями. Для $t = 0$ функция ν тождественно равна нулю, при увеличении t она не убывает и для всех t, x, k выполнено $\sum\limits_n \nu_k^t(n|x) \leqslant 2^{-k}$.

Для определения $\nu_k^{t+1}(n|x)$ введём вспомогательную величину

$$\alpha_k(n|x) = \begin{cases} 2^k \cdot \mu^t(n), & \text{при } \mu^t(x) = 2^{-n}, \\ \nu_k^t(n|x), & \text{иначе.} \end{cases}$$

Если $\sum\limits_n \alpha_k(n|x) \leqslant 2^{-k}$, положим $\nu_k^{t+1}(n|x) = \alpha_k(n|x)$, в противном случае $\nu_k^{t+1}(n|x) = \nu_k^t(n|x)$.

Функция $\nu(n|x) = \sum\limits_k \max\limits_t \nu_k^t(n|x)$ перечислима снизу и $\sum\limits_n \nu(n|x) \leqslant 1$. Поэтому $\nu(n|x) \leqslant O(\mu(n|x))$.

Для любого k существует такое n_0, что $\sum\limits_{n \geqslant n_0} 2^k \mu(n) \leqslant 2^{-k}$, и существует такое x_0, что для любого $x \geqslant x_0$ имеет место $\mu(x) \leqslant 2^{-n_0}$. Отсюда для $x \geqslant x_0$ имеем $\nu(n|x) = 0$ при $n < n_0$ и неравенство $\sum\limits_n \alpha_k(n|x) \leqslant 2^{-k}$ из определения функции ν будет выполнено при всех t. Следовательно, при $n = -\log \mu(x) = KP(x)$ получается $\nu(n|x) \geqslant 2^k \cdot \mu(n)$. Так как $\mu(n|x) \geqslant 2^{-C} \cdot \nu(n|x) \geqslant 2^{k-C} \cdot \mu(n)$, то $KP(KP(x)|x) \leqslant KP(x) - k + C$, где C зависит только от выбора оптимального языка программирования.

$\square$

Теорема 4. *Функция $IS(x : KS(x))$ неограничена, но не стремится к бесконечности.*

Доказательство. Неограниченность функции $IS(x : KS(x))$ почти очевидна. Пусть x — случайное слово длины n (то есть $KS(x) = n + O(1)$). Тогда $KS(KS(x)|x) = O(1)$, в то время как $KS(KS(x)) \to \infty$.

Теперь наша цель — доказать, что функция $IS(x : KS(x))$, ограниченная на некоторое бесконечное множество, не превосходит константы.

Поскольку для любого слова x длины n простая энтропия x не превышает $n + O(1)$, то $KS(KS(x)) \leqslant \log n + O(1)$. Мы покажем, что

$$\forall n\ \exists x \quad \ell(x) = n \wedge KS(KS(x)|x) \geqslant \log n - O(1)\,, \qquad (*)$$

откуда сразу следует утверждение теоремы.

Назовём *полуперечислителем* вычислимую функцию $f(x,t)$ (где x — двоичное слово, t и значение f — натуральные числа), которая монотонно невозрастает по t и $\lim\limits_{t\to\infty} f(x,t) \geqslant KS(x)$. Если $\forall x \lim\limits_{t\to\infty} f(x,t) = KS(x)$, то f называется *перечислителем*. Значения f мы будем называть *гипотезами* об энтропии x.

Пусть f_0 — какой-нибудь перечислитель, C — достаточно большое натуральное число. Построим по ним полуперечислитель f_1. Для каждого x он переходит к новой гипотезе $f_1(x,t)$, только когда за время t она была выдвинута перечислителем f_0 и было обнаружено, что $KS(f_1(x,t)|x) < \lfloor \log \ell(x) \rfloor - C$. Количество гипотез f_1 для каждого x меньше $\ell(x)/2^C$. Простая энтропия программы вычисления f_1 не больше $\log C + O(1)$. Предположим, что утверждение $(*)$ не выполнено, тогда $\exists n\, \forall x\, [\ell(x) = n \ \Rightarrow \lim\limits_{t\to\infty} f(x,t) = KS(x)]$. Осталось доказать следующую лемму.

Лемма. *Пусть число c достаточно велико и полуперечислитель f задаётся программой с простой энтропией $d > 0$. Если $\forall x\, [\ell(x) = n \ \Rightarrow \lim\limits_{t\to\infty} f(x,t) = KS(x)]$, то на некотором x длины n количество гипотез f больше $n/(cd)$.*

Эта лемма усиливает теоремы 2 и 4 из работы [2] (автор доказал эту лемму после прочтения [2]).

Доказательство. Без ограничения общности предположим, что $f(x,0) = \ell(x) + O(1) > \ell(x)$. Построим вспомогательную частичную вычислимую функцию $g(y,a)$ (где y — двоичное слово, a и значение g — натуральные числа). На рис. 1 приведён алгоритм перечисления графика функции g (в дальнейшем мы будем ссылаться на пункты этого алгоритма).

Очевидно, что $\forall a, x\ \ KS(x) \leqslant \min\{\ell(y) \mid g(y,a) = x\} + O(d + \log a)$. Пусть b — достаточно большое число (его величина зависит от выбора оптимального языка программирования). Фиксируем $a = bd$. В пункте 5 мы всегда надём $t' > t$, для которого $f(x,t') < f(x,t)$, потому что $a > O(d + \log a)$. Для любых n, m есть не более одного y длины m, для

0. *Параллельно для всех натуральных* a *и* $n > 5$ *делаем следующее.*
1. *Последовательно для каждого* x *длины* n*:*
2. *Положим* $t = 0$.
3. *Если* $f(x,t) \leqslant n/2 + a$,
то завершаем процесс для данного n.
4. *Ищем такое слово* y *длины* $f(x,t) - a$,
что $g(y,a)$ *ещё не определено;*
если оно найдено,
то полагаем $g(y,a) = x$,
иначе завершаем процесс для данного n.
5. *Ждём такое* $t' > t$, *что* $f(x,t') < f(x,t)$.
6. *Если* $f(x,t') < f(x,t) - a - 1$,
то переходим к следующему x *из пункта* 1,
иначе полагаем $t = t'$ *и переходим к пункту* 3.

Рис. 1. Алгоритм перечисления графика функции $g(y,a)$.

которого $g(y,a) = x$, $\ell(x) = n$ и $KS(x) \geqslant m - 1$ (такие y будем называть *исключениями*). При $\ell(y) = m$ значение $g(y,a)$ могло быть определено только при рассмотрении в пункте 0 чисел $n < 2m$. Мощность $\{x \mid KS(x) < m-1\}$ меньше $2^m/2$, количество исключений длины m меньше $2m$; поэтому если бы в пункте 4 не удалось найти y, то $2^m/2 < 2m$, что противоречит ограничению $n > 5$ из пункта 0.

При $n < 4a$ утверждение леммы будет выполнено для $c \geqslant 4b$. Пусть $n \geqslant 4a$.

Если для всех x длины n количество гипотез f не больше $n/(cd)$, то при некотором t условие в пункте 6 обязательно будет выполнено. Действительно, при уменьшении $f(x,t)$ от $n + O(1)$ до $n/2 + a$ было меньше $n/(cd)$ прыжков, следовательно, один из них был больше $(n/2 - a) : (n/(cd)) \geqslant cd/4 > a + 1$ (если $c > 4b + 4$). Получается, что для каждого x длины n некоторая гипотеза f строго меньше n. Но мощность $\{x \mid KS(x) < n\}$ строго меньше 2^n, а количество слов длины n равно 2^n. Противоречие. □

Теорема доказана. □

В теоремах 2 и 3 не было дано никакой эффективной оценки на ско-

рость роста функций $IP(KP(x) : x)$ и $IP(x : KP(x))$. Это неудивительно: как известно, даже для функций $KP(x)$ и $KP(KP(x))$ не существует частично вычислимых неограниченных нижних оценок. Тем не менее, следующие две теоремы показывают, что для каждой из функций $IP(KP(x) : x)$ и $IP(x : KP(x))$ существует бесконечное (и не слишком редкое) множество, на котором функция увеличивается достаточно быстро.

Теорема 5. *Для любого натурального k существует двоичное слово z длины не больше k, для которого $IP(KP(z) : z) \geqslant \log k - O(1)$.*

Доказательство. Будем считать число k достаточно большим (для малых k утверждение теоремы тривиально).

Определим вычислимую функцию $\nu^t(x|n)$ от натуральных аргументов x, n и t с рациональными значениями. Для $t = 0$ функция ν тождественно равна нулю, при увеличении t она не убывает и для всех t, n выполнено $\sum\limits_x \nu^t(x|n) \leqslant 1$.

Положим $\nu^{t+1}(x|n) = (n/4)\cdot 2^{-n}$, если $\mu^t(x) \geqslant 2^{-n}$ и $\left|\{x|\mu^t(x) \geqslant 2^{-n}\}\right| \leqslant 2^{n+2}/n$, в противном случае $\nu^{t+1}(x|n) = \nu^t(x|n)$.

Функция $\nu(x|n) = \max\limits_t \nu^t(x|n)$ перечислима снизу и $\sum\limits_x \nu(x|n) \leqslant 1$. Поэтому $\nu(x|n) \leqslant O(\mu(x|n))$.

Покажем, что для одного из $n \in [k/2, k]$ при всех t выполнено $\left|\{x|\mu^t(x) \geqslant 2^{-n}\}\right| \leqslant 2^{n+2}/n$. Если бы это было не так, то

$$\forall n \in [k/2, k] \quad \left|\{x|\mu(x) \geqslant 2^{-n}\}\right| \cdot 2^{-n} > 4/n\,.$$

Суммируя, выводим

$$\sum_{n=k/2}^{k} \left|\{x|\mu(x) \geqslant 2^{-n}\}\right| \cdot 2^{-n} > \sum_{n=k/2}^{k} 4/n > 2\,.$$

С другой стороны

$$2 \geqslant \sum_x 2\mu(x) \geqslant \sum_{n=k/2}^{k} \left|\{x|\mu(x) \geqslant 2^{-n}\}\right| \cdot 2^{-n}\,,$$

противоречие.

Фиксируем найденное n.

Рассмотрим случайное слово y длины k. Его простая, а тем более префиксная энтропия больше $k - O(1)$. Пусть y_i — начало y длины i. Понятно, что в последовательности $KP(y_k), KP(y_{k-1}), \ldots$ соседние числа отличаются не более чем на константу. Если j — наименьшее натуральное число, для которого $KP(y_{k-j}) \leqslant n$, то $KP(y_{k-j}) \geqslant n - C$, где C зависит только от выбора оптимального языка программирования. Положим $z = y_{k-j}$, тогда $\mu(z) \geqslant 2^{-n}$ и $\mu^t(z) \geqslant 2^{-n}$ начиная с некоторого t. Начиная с этого t будет $\nu^t(z|n) = (n/4) \cdot 2^{-n}$ (за счёт выбора n). Отсюда $\mu(z|n) \geqslant 2^{-O(1)} \cdot (n/4) \cdot 2^{-n}$. Логарифмируя, получаем $KP(z|n) \leqslant n - \log n + O(1)$. Поскольку $|KP(z) - n| \leqslant C$, то $KP(z|KP(z)) \leqslant KP(z|n) + O(1)$. Следовательно, $KP(z) - KP(z|KP(z)) \geqslant (n - C) - (n - \log n + O(1)) \geqslant \log k - O(1)$. Теорема доказана. □

Теорема 6. *Для любого натурального k существует двоичное слово z длины не больше k, для которого $IP(z : KP(z)) \geqslant \log k - O(\log\log k)$.*

Доказательство. Будем считать число k достаточно большим (для малых k утверждение теоремы тривиально).

Рассмотрим натуральное число $n \leqslant k$, двоичная запись которого является случайным словом длины $\lfloor \log k \rfloor$, то есть $KS(n) = \lfloor \log k \rfloor + O(1)$. Пусть z — случайное слово длины n, то есть $KS(z) = n + O(1)$. Понятно, что $KS(n) - O(1) \leqslant KP(n) \leqslant KP(KP(z)) + KP(KP(z) - n) + O(1) \leqslant KP(KP(z)) + KP(O(\log n)) + O(1) \leqslant KP(KP(z)) + O(\log\log n)$. Поэтому $KP(KP(z)) \geqslant \log k - O(\log\log k)$. С другой стороны, $KP(KP(z)|z) \leqslant KP(n|z) + KP(KP(z) - n|z) + O(1) \leqslant O(1) + KP(O(\log n)|z) \leqslant O(\log\log k)$. Следовательно, $KP(KP(z)) - KP(KP(z)|z) \geqslant \log k - O(\log\log k)$. Теорема доказана. □

Автору было полезно знакомство с работой [2] и замечание анонимного рецензента о связи этой работы со статьёй [3]. Большую помощь оказал Алексей Вячеславович Чернов при подготовке текста к публикации, за что автор ему очень благодарен.

Литература

[1] An. A. Muchnik, S. Ye. Positselsky. Kolmogorov entropy in the context of computability theory. *Theoretical Computer Science*, 2002, v. 271, pp. 15–35.

[2] R. Beigel, H. Buhrman, P. Fejer, L. Fortnow, L. Longpré, F. Stephan, L. Torenvliet. Enumerators of the Kolmogorov Function. Technical Report 2001-004, NEC Research Institute, 2001.

[3] П. Гач. О симметрии алгоритмической информации. *Доклады АН СССР*, 1974, т. 218, N. 6, с. 1265–1267.