Обзор: рез-тов Колмог. сем. МГУ

Русский вариант Мучник-Семенов

~~Анализ~~ минимальных кодов в теории Колмогоровской энтропии.

Полные доказательства приводимых результатов были изложены в сентябре 1999 года на Колмогоровском семинаре Московского Государственного университета.

Для конструктивных объектов x и y Колмогоровская энтропия x при условии y (обозначаемая $K(x|y)$) определяется как длина кратчайшей программы, которая на входе y даёт выход x.[1] Причём предполагается, что программы пишутся в некотором оптимальном языке программирования. В данной работе нас будут интересовать не только размеры программ, переводящих y в x, но и другие свойства этих программ. Мы докажем, что для любых y_1, y_2 и x существует программа p со следующими тремя свойствами. Во-первых, $p(y_1) = p(y_2) = x$; во-вторых, длина p превышает $\max\{K(x|y_1), K(x|y_2)\}$ не более, чем на $const \cdot \log K(x)$; в-третьих, $K(p|x)$ меньше, чем $const \cdot \log K(x)$.[2] Простейшая программа, удовлетворяющая первому свойству, имеет длину $K(x|y_1) + K(x|y_2)$. К.Ю.Горбунов доказал в [1], что последня оценка не улучшаема для некоторых объектов y_1, y_2 и x, абсолютная энтропия которых экспоненциальна относительно $K(x|y_1) + K(x|y_2)$. Поэтому погрешность $const \cdot \log K(x)$ в нашем результате неизбежна. Оказывается, эта погрешность неизбежна и в утверждении о том, что минимальную информацию, достаточную для нахождения x при известном y, можно извлечь из самого x (а не взять со стороны). Подчеркнём, что наши оценки зависят только от энтропии x, но не от энтропии условий y. Получается интересное следствие: хотя всех программ одного размера экспоненциально много, для нахождения x по всевозможным y с фиксированным значением $K(x|y)$ достаточно полиномиального от $K(x)$ количества программ размера $K(x|y)$. Будет дана и нижняя оценка на мощность такого “универсального„ множества программ. Если отождествлять программы p и q, когда $K(p|q)$ и $K(q|p)$ достаточно малы, то последняя нижняя оценка оказывается точной.

Объясним неформальный термин ‘код’, употреблённый в названии доклада. Объект p называется кодом объекта x при известном объекте y, когда величина $K(x|y,p)$ достаточно мала. Просто понять, что по коду p эффективно строится программа q, выдающая x на входе y и имеющая длину не намного превышающую K(p). Рассуждать о кодах во многих случаях удобнее, чем о программах.

Интересно сравнить результаты, относящиеся к формуле $(A \vee B) \to C$, с результатами, относящимися к формуле $A \leftrightarrow B$. В [2] доказано, что минимальная энтропия программы, переводящей объекты y и z друг в друга, примерно равна максимуму минимальной энтропии программы, переводящей y в z, и минимальной энтропии программы, переводящей z в y. Этот результат может быть получен из нашего, если рассмотреть задачу перевода объектов y и z в объект $<y,z>$. Однако такой способ даёт погрешность $const \cdot \log K(<y,z>)$, тогда как доказательство из [2] даёт погрешность $const$. Укажем ещё одно различие. Бывает, что для минимальной программы p, переводящей y в x, и минимальной программы q, переводящей z в x, пара $<p,q>$ содержит нулевую информацию о любой минимальной программе, одновременно переводящей y в x

---

[1]Результаты этой работы в равной степени относятся к *простой* и к *префиксной* энтропии.

[2]По определению, абсолютная энтропия $K(w) = K(w|\Lambda)$, где Λ - некоторый заранее фиксированный объект (например, пустое слово).

и z в x. Наоборот, пара минимальных программ p и q, переводящих соответственно x в y и y в x, всегда имеет большую общую информацию с подходящей минимальной программой, одновременно переводящей x в y и y в x.

Теорема 1 (Ан. А. Мучник).
Существуют число c и частичная вычислимая функция $\lambda xl.F(x,l)$, для которых выполнено следующее: если $K(x) \le l$, то
i) значением $F(x,l)$ является множество мощности меньше $cl/\log l$, состоящее из двоичных слов длины l (универсальное множество кодов для x);
ii) для любых y_1, y_2 существуют такие p_1, p_2, что для некоторого $p \in F(x,l)$ p_1- начало p длины равной $K(x|y_1)$, p_2- начало p длины равной $K(x|y_2)$, $K(x|y_1,p_1) < c\log l$, $K(x|y_2,p_2) < c\log l$.
(Отметим, что из i) и ii) следует $K(p_1|x) < const \cdot \log l$ и $K(p_2|x) < const \cdot \log l$).

Доказательство.

Каждому числу l сопоставим число r_l (содержательный смысл: $K(x) \le l$, $r_l = \lceil \log F(x,l) \rceil$). Значение r_l будет определено позже, индекс l иногда будет опускаться. Обозначим через L и R множества двоичных слов длины l и r соответственно. Рассмотрим пространство функций из $L \times R$ в L и равномерное распределение вероятностей на этом пространстве. Распределение вероятностей понадобится, чтобы найти функцию f, которая при каждом $n < l - \log l$ удовлетворяет приводимому ниже требованию $(*)$.
(Содержательный смысл: $n = K(x|y)$). Обозначим через M множество двоичных слов длины $m = n + \lceil \log l \rceil$. Обозначим через $\varphi(z,\rho)$ начало $f(z,\rho)$ длины m.

$$\forall B \subset M \quad |B| \le 2^n \to \left|\left\{z : \left|\{\rho : \varphi(z,\rho) \in B\}\right| \ge 2^{r-1}\right\}\right| < |B| \qquad (*)$$

Оценим сверху вероятность невыполнения $(*)$. Сначала фиксируем n, $z \in L, \rho \in R$ и $B \subset M, |B| \le 2^n$. Вероятность того, что $\varphi(z,\rho)$ принадлежит B, равна $|B|/|M| \le 2^n / 2^m \le 1/l$. Теперь фиксируем n, z, B и $R' \subset R, |R'| \ge |R|/2$. Вероятность того, что $\forall \rho \in R' \quad \varphi(z,\rho) \in B$, не превышает $l^{-|R|/2}$. Получается, что при фиксированных n, z, и B вероятность события $\left|\{\rho : \varphi(z,\rho) \in B\}\right| \ge 2^{r-1}$ не превышает $2^{|R|} \cdot l^{-|R|/2} < 2^{-const \cdot \log l \cdot |R|}$. Теперь фиксируем n, $s \le 2^n$, $B \subset M, |B| = s$ и $Z \subset L, |Z| = s$. Вероятность события $\forall z \in Z \quad \left|\{\rho : \varphi(z,\rho) \in B\}\right| \ge 2^{r-1}$ не превышает $2^{-const \cdot \log l \cdot |R| \cdot s}$. После умножения на количество возможноных пар B и Z получим верхнюю оценку на вероятность невыполнения события $\forall B \subset M \quad |B| = s \to \left|\left\{z : \left|\{\rho : \varphi(z,\rho) \in B\}\right| \ge 2^{r-1}\right\}\right| < s$ при фиксированных n и s, а именно $2^{-const \cdot \log l \cdot |R| \cdot s} \cdot 2^{m \cdot s} \cdot 2^{l \cdot s} \le \left(2^{-const \cdot \log l \cdot |R| + 2l}\right)^s$. Если $|R| = c \cdot l / \log l$ для достаточно большой константы c, то число $2^{-const \cdot \log l \cdot |R| + 2l}$ меньше $1/2l$, а следовательно $\sum_s \left(2^{-const \cdot \log l \cdot |R| + 2l}\right)^s < 1/l$. Указанная сумма по s оценивает сверху вероятность невыполнения $(*)$ при фиксированном n. Поскольку $n<l$, то вероятность невыполнения хотя бы при одном n требования $(*)$ строго меньше 1. Итак, мы определили $r_l = \log|R|$ и доказали, что для каждого l существует функция f со свойством $(*)$. Так как выполненность $(*)$ эффективно проверяема, мы можем перебором находить требуемую функцию.

Будем обозначать через $\nu_l(u)$ первую программу t длины l, для которой обнаружится $t(\Lambda)=u$ (если такой программы нет, то $\nu_l(u)$ не определено). Когда понятно о каком l идёт речь, мы будем писать νu вместо $\nu_l(u)$. Пусть, в процессе перечисления сверху энтропии K обнаружилось $K(x)\le l$. Пусть f - функция со свойством $(*)$, соответствующая параметру l. Определим $F(x,l)$ как множество слов вида $f(x,\rho)$. Проверим пункт ii) из формулировки теоремы. Пусть, y - один из объектов y_1,y_2; $n=K(x|y)$; d - вспомогательный параметр, значение которого будет определено позже. Мы будем использовать определённые в предыдущем абзаце числа r и m, множества R и M и функцию φ. Сейчас будут построены ещё некоторые вспомогательные множества. Мы знаем, что x принадлежит множеству $D=\{u{:}K(u)\le l\,\&\,K(u|y)\le n\}$. Ясно, что $|D|<2^{n+1}$. Для $q\in M$ определим в D подмножество

$E^q=\{u{:}\exists\rho\in R\quad q=\varphi(\nu u,\rho)\}$. Определим в M подмножество $G=\{q{:}|E^q|>2^d\}$. Понятно, что $|G|<|D|\cdot|R|/2^d<2^{n+1+r-d}$. Определим в D подмножество

$H=\{u{:}\left|\{\rho{:}\varphi(\nu u,\rho)\in G\}\right|\ge 2^{r-1}\}$. Параметр d будет определён так, что $d>r$, поэтому $|G|<2^n$. Благодаря свойству $(*)$, имеем $|H|<|G|$. Обратим внимание, что множество D равномерно перечислимо по l, n, y; множество E^q равномерно перечислимо по l, n, y, q; множество G равномерно перечислимо по l, n, y, d; множество H равномерно перечислимо по l, n, y, d. Предположим, что $x\in H$. Тогда энтропия x при условии y с точностью до аддитивной константы не превышает суммы энтропии программы перечисления H при условии y и длины записи номера x в этом перечислении. То есть

$K(x|y)<K(l)+K(n)+K(d)+\log_2|H|+const=K(l)+K(n)+K(d)+n+r-d+const$. Если положить $d=\alpha\cdot\log l$ для достаточно большой константы α, то окажется $K(l)+K(n)+K(d)+n+r-d+const<n$. Это противоречит тому, что $n=K(x|y)$. Итак, $x\notin H$. Напомним, что в качестве у можно взять y_1 или y_2. То есть, фактичести, были построены два семейства множеств - D_1,E_1^q,G_1,H_1 и D_2,E_2^q,G_2,H_2. Из того, что $x\notin H_1$ и $x\notin H_2$, следует $\left|\{\rho{:}\varphi_1(\nu x,\rho)\in G_1\}\right|<2^{r-1}$ и $\left|\{\rho{:}\varphi_2(\nu x,\rho)\in G_2\}\right|<2^{r-1}$. Поскольку $|R|=2^r$, существует $\rho_0\in R$, для которого $\varphi_1(\nu x,\rho_0)\notin G_1$ и $\varphi_2(\nu x,\rho_0)\notin G_2$. Это значит, что $|E_1^{\varphi_1(\nu x,\rho_0)}|\le 2^d$ и $|E_2^{\varphi_2(\nu x,\rho_0)}|\le 2^d$. Из определения множества E^q следует, что $x\in E_1^{\varphi_1(\nu x,\rho_0)}$ и $x\in E_2^{\varphi_2(\nu x,\rho_0)}$. Поэтому энтропия x при условии $\langle y_1,\varphi_1(\nu x,\rho_0)\rangle$ с точностью до аддитивной константы не превышает суммы энтропии программы перечисления $E_1^{\varphi_1(\nu x,\rho_0)}$ при условии $\langle y_1,\varphi_1(\nu x,\rho_0)\rangle$ и длины записи номера x в этом перечислении. То есть

$K(x|y_1,\varphi_1(\nu x,\rho_0))<K(l)+K(n_1)+\log_2|E_1^{\varphi_1(\nu x,\rho_0)}|+const=K(l)+K(n_1)+d+const$. Аналогично $K(x|y_2,\varphi_2(\nu x,\rho_0))<K(l)+K(n_2)+d+const$. Напомним, что $d=\alpha\cdot\log l$. В качестве p возьмём $f(\nu x,\rho_0)$. Пусть, p_1 - начало p длины равной n_1, p_2 - начало p длины равной n_2. Так как $\varphi_1(\nu x,\rho_0)$ является началом p длины равной m_1, $\varphi_2(\nu x,\rho_0)$ является началом p длины равной m_2 и $m_1=n_1+\lceil\log l\rceil$, $m_2=n_2+\lceil\log l\rceil$, то для некоторого c получается $K(x|y_1,p_1)<c\log l$ и $K(x|y_2,p_2)<c\log l$. ◊

Количество условий y в доказанной теореме может быть увеличено с двух до полинома от $K(x)$.

Интересно, что в полурешётке введённой в [3], не всегда есть пересечение двух элементов, но всегда есть разность.

Следующая теорема показывает, что множество кодов построенное в предыдущей теореме не может быть значительно уменьшено.

Теорема 2 (Ан. А. Мучник).
Каждому числу α соответствует такое число c, что
для любого двоичного слова x и для любого множества P мощности меньше $K(x)/c\log K(x)$, состоящего из двоичных кодов длины равной $\lceil K(x)/2 \rceil$, найдётся условие y, для которого:
i) $K(y) < cK(x)$;
ii) $K(x|y) < K(x)/2$;
iii) $\forall p \in P \quad K(x|y,p) > \alpha \log K(x)$.

Доказательство.
Пусть c - достаточно большое число. Предположим, что $P = \{p_1,\dots,p_j\}$ и $j < K(x)/c\log K(x)$. Обозначим через v_i начало p_i длины равной $c\log K(x)/3$. Пусть, w - конкатенация слов $v_1,\dots,v_j$. Тогда
$K(w) < j \cdot c\log K(x)/3 + const = K(x)/3 + const$ и
$K(x|w) > K(x) - K(w) - const \cdot \log K(x) > K(x)/2$ при достаточно большом $K(x)$.
Рассмотрим значения величины $K(x|wz)$ для слов z, пробегающих начала слова x. Когда длина z меняется на 1, значение $K(x|wz)$ меняется не более, чем на константу. Поскольку $K(x|w\Lambda) = K(x|w) > K(x)/2$ и $K(x|wx) < const$, то среди начал x существует такое z_0, что $K(x)/2 > K(x|wz_0) > K(x)/2 - const$.
Возьмём в качестве y слово wz_0. Проверим пункт i):
$K(y) < K(w) + K(z_0) < K(x)/3 + K(x) + const$. Проверим пункт ii):
$K(x|y) = K(x|wz_0) < K(x)/2$. Проверим пункт iii). Для каждого i слово y содержит “много” информации о p_i, а именно
$K(p_i|y) < K(x)/2 - c\log K(x)/3 + const \cdot \log K(x)$. Используя последнее неравенство, выводим: $\forall i \quad K(x|y,p_i) > K(x|y) - K(p_i|y) - const \cdot \log K(x) >$
$> (K(x)/2 - const) - (K(x)/2 - c\log K(x)/3 + const \cdot \log K(x)) - const \cdot \log K(x)$.
При достаточно большом c пункт iii) выполнен. ◊

[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.Gacs, 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 semi-lattice of binary strings with the relation “x is simple conditional to y”. Theoretical Computer Science, to appear, 2000.