О совместной условной сложности (энтропии).

Н.К. Верещагин* Ан.А Мучник†

Аннотация

Колмогоровская сложность слова b при известном a определяется как минимальная длина программы, перерабатывающей a в b. Мы обобщаем это понятие на четверки слов a, b, c, d: их совместной условной сложностью $K((a \to c) \wedge (b \to d))$ называется минимальная длина программы, перерабатывающей a в c, а b в d. В работе доказано, что совместная условная сложность не выражается через обычные условные и безусловные сложности. Вопрос о существовании задачи о переработке информации, сложность которой не выражается через обычные условные и безусловные сложности, был поставлен А. Шенем на одном из заседаний колмогоровского семинара в МГУ в 1994 году. Наш результат дает положительный ответ на этот вопрос.

Кроме того, мы доказываем аналогичные утверждения и для классической шенноновской энтропии. Мы приводим два различных доказательства обоих результатов — "эффективное" и "квази-эффективное". В заключение мы приводим квази-эффективное доказательство усиленного варианта известного результата о существовании слов с невыделяемой общей информацией. Ранее было известно только неэффективное доказательство этого утверждения.

Предварительная публикация результатов данной статьи, относящихся к колмогоровской сложности, появилась в сборнике докладов конференции Conference on Computational Ccomplexity в 2001 году. [A. Muchnik and N. Vereshchagin. "Logical operations and Kolmogorov complexity. II". Proc. of 16th Annual IEEE Conference on Computational Complexity, Chicago, June 2001, pp. 256–265.]

1. Введение

Пусть дана некоторая частичная вычислимая функция $V(x, y)$, аргументами и значениями которой являются двоичные слова. *Условной колмогоровской сложностью* $K_V(b|a)$ называется наименьшая длина такого p, что $V(p, b) = b$. Если выполнено это равенство, то говорят, что p является *описанием b при известном a относительно способа описания V*. По теореме Соломонова–Колмогорова, существует способ описания U, который не хуже любого другого способа описания V в следующем смысле. Для всех V существует c такое, что $K_U(b|a) \leqslant K_V(b|a) + c$ для всех a, b. Функции U, удовлетворяющие теореме Соломонова–Колмогорова, называются *оптимальными*. Зафиксируем любой оптимальный способ описания U и будем обозначать $K(b|a) = K_U(b|a)$, опуская индекс. При переходе к другому оптимальному способу описания $K(b|a)$ может измениться, но не более, чем на слагаемое, не зависящее от a, b.

*Кафедра математической логики и теории алгоритмов, Московский государственный университет им. М.В. Ломоносова, e-mail: ver@mccme.ru. Работа выполнена при поддержке гранта РФФИ 09-01-00709

†Московский институт новых технологий образования, Москва 109004, Нижняя Радищевская 10.

Просто *колмогоровской сложностью* $K(a)$ называется $K(a|\text{пустое слово})$. Если $V(p, \text{пустое слово}) = a$, то говорят, что p является *описанием* a *относительно способа описания* V. Колмогоровская сложность любого слова конечна, более того, она не превосходит длины слова плюс некоторая константа. Слова, у которых сложность примерно равна длине, называются случайными.

Колмогоровская сложность не увеличивается при вычислимых преобразованиях: для любой вычислимой частичной функции f найдется константа c такая, что

$$K(f(x)) \leqslant K(x) + c$$

для всех x из области определения f.

Колмогоровская сложность пары слов $\langle p, q\rangle$ определяется как сложность слова $[p, q]$, где $\langle x, y\rangle \mapsto [x, y]$ некоторая вычислимая биекция между парами двоичных слов и двоичными словами. При смене вычислимой биекции на другую, $K([p, q])$ изменяется не более, чем на аддитивную константу, поскольку сложность не увеличивается при вычислимых преобразованиях. Поэтому нам неважно, какую именно биекцию выбрать. В дальнейшем мы будем сокращать $K(\langle p, q\rangle) = K([p, q])$, как $K(p, q)$.

Интуиция подсказывает, что сложность пары не превосходит суммы сложностей: $K(p, q) \leqslant K(p) + K(q)$. Однако это неравенство верно только с точностью до логарифмического слагаемого: $K(p, q) \leqslant K(p) + K(q) + O(\log K(q))$. Проблема в том, что мы не можем просто соединить кратчайшие описания p и q в одно описание пары $\langle p, q\rangle$ — требуется еще информация, где заканчивается первое описание и начинается второе. Эту информацию можно закодировать словом размера $O(\log K(q))$ и приписать в начало конкатенации кратчайших описаний p и q. Причем код можно выбрать "самоограничивающим", то есть таким, что читая его слева направо, мы можем понять, где он заканчивается.

Языком программирования назовем вычислимую частичную функцию $P(x, y)$, аргументами и значениями которой являются двоичные слова, и такую, что для любой другой вычислимой частичной функции $Q(x, y)$ найдется тотальная вычислимая функция s, для которой $P(s(x), y) = Q(x, y)$ для всех x, y. Функция s в этом контексте называется транслятором Q-программ в P-программы. Если $P(x, y) = z$, то говорят, что *программа* x *на входе* y *выдает* z, или *перерабатывает* y *в* z. Зафиксируем некоторый язык программирования P и будем обозначать через $a \to b$ множество всех программ переработки a в b. Будем также обозначать $P(x, y)$ через $[x](y)$.

Колмогоровской сложностью непустого множества слов A называется минимальная сложность его элементов: $K(A) = \min\{K(a) \mid a \in A\}$. В качестве A мы обычно будем рассматривать множество программ, решающих некоторую алгоритмическую задачу. В этом случае само A мы будем называть "задачей", а его элементы — её решениями. Например, $a \to b$ есть задача преобразования a в b, а $K(a \to b)$ — сложность этой задачи.

Поскольку колмогоровская сложность не увеличивается при вычислимых преобразованиях при изменении языка прогаммирования $K(a \to b)$ меняется не более чем на слагаемое, не зависящее от a, b. Кроме того, нетрудно понять, что $K(a \to b)$ совпадает с $K(b|a)$ с точностью до постоянного слагаемого. Это верно для любого выбора языка программирования P в определении множества $a \to b$, что также доказывает независимость $K(a \to b)$ от выбора языка программирования.

Сложность задачи совместной переработки a в c и b в d можно определять двумя естественными эквивалентными способами:

- $K((a \to c) \wedge (b \to d))$ есть минимальная сложность программы, перерабатывающей $0a$ в c и $1b$ в d, то есть сложность пересечения множеств $0a \to c$ и $1b \to d$;

- $K((a \to c) \wedge (b \to d))$ есть минимальная сложность пары программ $\langle p, q\rangle$, где p перерабатывает a в c, а q перерабатывает b в d, то есть, сложность декартова произведения множеств $a \to c$ и $b \to d$.

Все вышеприведенные понятия имеют свои аналоги в классической теории информации Шеннона. Аналогом понятия колмогоровской сложности слова является шенноновская энтропия случайной величины. Напомним определение последней. Пусть α случайная величина с возможными значениями $a_1, \ldots, a_n$. Пусть p_i обозначает вероятность события $\alpha = a_i$. *Шенноновская энтропия* α определяется формулой

$$H(\alpha) = -\sum_i p_i \log p_i$$

(Все логарифмы в работе берутся по основанию 2.) Из вогнутости логарифмической функции следует, что энтропия Шеннона не превосходит логарифма количества возможных значений и равна ему тогда и только тогда, когда случайная величина имеет равномерное распределение. Кроме того, она неотрицательна и равна нулю тогда и только тогда, когда все вероятности $p_1, \ldots, p_n$, кроме одной, нулевые.

Пусть β — другая случайная величина с возможными значениями $b_1, \ldots, b_k$, совместно распределенная с α (это означает, что заданы вероятности каждого из событий $\alpha = a_i, \beta = b_j$). Величина $H(\alpha|\beta = b_j)$ (для любого фиксированного j) определяется аналогично $H(\alpha)$ с той разницей, что p_i в определении нужно заменить на условную вероятность $\Pr[\alpha = a_i|\beta = b_j]$. *Условная энтропия Шеннона* определяетя формулой

$$H(\alpha|\beta) = \sum_j \Pr[\beta = b_j] \cdot H(\alpha|\beta = b_j).$$

Можно показать, что условная энтропия не превышает безусловной: $H(\alpha|\beta) \leqslant H(\alpha)$. Для колмогоровской сложности аналогичное неравенство очевидно (разумеется, оно выполнено с точностью до постоянного слагаемого).

Непосредственно из определения следует равенство

$$H(\alpha, \beta) = H(\alpha|\beta) + H(\beta).$$

Аналогичное равенство для колмогоровской сложности

$$K(a, b) = K(a|b) + K(b),$$

выполнено только с точностью до слагаемого порядка $O(\log K(a, b))$ и было доказано Колмогоровым и Левиным (см. [7]). Из препоследнего равенства и неравенства $H(\alpha|\beta) \leqslant H(\alpha)$ следует неравенство

$$H(\alpha, \beta) \leqslant H(\alpha) + H(\beta),$$

аналог которого для колмогоровской сложности мы уже отмечали.

Ясно, что условная энтропия равна нулю, только если обнуляются все слагаемые $H(\alpha|\beta = b_j)$, для которых $\Pr[\beta = b_j]$ положительно. Иными словами, $H(\alpha|\beta) = 0$ тогда и только тогда, когда для некоторой функции g с вероятностью 1 выполнено $\alpha = g(\beta)$.

По аналогии с колмогоровской сложностью, для которой $K(a|b)$ определяется как минимальная длина программы, перерабатывающей b в a, хотелось бы, чтобы $H(\alpha|\beta)$ было равным минимальной энтропии случайной величины γ (совместно распределенной с α, β), для которой для некоторой функции g с вероятностью 1 выполнено $\alpha = g(\beta, \gamma)$. Это верно с точностью до логарифмических добавок. Действительно, с одной стороны имеет место следующая

Теорема 1 ([4]). *Для любых случайных величин α, β найдется случайная величина γ, для которой $H(\gamma) \leqslant H(\alpha|\beta) + O(\log H(\alpha|\beta))$ и $H(\alpha|\beta,\gamma) = 0$.*

(Здесь мы применили сокращение $H(\alpha|\beta,\gamma) = H(\alpha|\langle\beta,\gamma\rangle)$, которого будем придерживаться и в дальнейшем.)

А с другой стороны, для любых трех случайных величин справедливо неравенство

$$H(\alpha|\beta) \leqslant H(\alpha|\beta,\gamma) + H(\gamma)$$

Поэтому энтропия любой случайной величины γ, для которой $H(\alpha|\beta,\gamma) = 0$, не может быть меньше $H(\alpha|\beta)$. Последнее неравенство следует из равенства

$$H(\alpha,\gamma|\beta) = H(\alpha|\beta,\gamma) + H(\gamma|\beta),$$

которое является условной версией известного нам равенства $H(\alpha,\gamma) = H(\alpha|\gamma) + H(\gamma)$ и выводится из последнего с помощью усреднения по всем исходам β. (Как обычно, мы сокращаем $H(\alpha,\gamma|\beta) = H(\langle\alpha,\gamma\rangle|\beta)$.)

По аналогии с $K((a \to c) \wedge (b \to d))$, определим

$$H((\alpha \to \gamma) \wedge (\beta \to \delta)) = \min\{H(\rho) \mid H(\gamma|\alpha,\rho) = H(\delta|\beta,\rho) = 0\}.$$

Здесь α, γ, β и δ произвольные совместно распределенные случайные величины, а ρ пробегает случайные величины, совместно распределенные с исходными случайными величинами. (Другими словами, мы имеем право определить совместное распределение всех пяти случайных величин так, чтобы проекция этого распределения на первые четыре координаты была равна данному нам распределениию.)

Возникает естественный вопрос: являются ли величины $K((a \to c) \wedge (b \to d))$ и $H((\alpha \to \gamma) \wedge (\beta \to \delta))$ чем-то существенно новым? То есть, выражаются ли они через обычные сложности (соответственно, энтропии) a, b, c, d (соответственно, $\alpha, \gamma, \beta, \delta$) и взаимные условные сложности (энтропии) кортежей, составленных из них? Чтобы аккуратнее поставить оба вопроса, введем понятие сложностного (энтропийного) вектора: сложностным вектором кортежа слов $\langle a_1, \ldots, a_k\rangle$ числовой вектор длины $2^k - 1$, состоящий из сложностей самих слов $a_1, \ldots, a_k$, их пар, троек и т.д. Например, сложностной вектор тройки слов $\langle a, b, c\rangle$ равен $\langle K(a), K(b), K(c), K(a,b), K(a,c), K(b,c), K(a,b,c)\rangle$. Заменив здесь сложность на энтропию и бинарные слова на случайные величины, получим понятие энтропийного вектора. Заметим, что сложностной (энтропийный) вектор кортежа слов (случайных величин) определяет и все взаимные условные сложности (энтропии) кортежей, составленные из них. Для колмогоровской сложности это верно с логарифмической точностью по вышеупомянутой теореме Колмогорова–Левина ($K(\langle a, b\rangle) = K(a|b) + K(b) + O(\log K(\langle a, b\rangle))$, а для энтропии из равенства $H(\alpha,\beta) = H(\beta) + H(\alpha|\beta)$.

Ответ на оба вопроса даётся следующей теоремой.

Теорема 2. *Существуют две последовательности четверок слов $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n, \tilde{d}_n$ и $\bar{a}_n, \bar{b}_n, \bar{c}_n, \bar{d}_n$ линейной (от n) сложности, у которых соответствующие компоненты сложностного вектора отличаются всего на $O(1)$, но разность $K((\tilde{a}_n \to \tilde{c}_n) \wedge (\tilde{b}_n \to \tilde{d}_n)) - K((\bar{a}_n \to \bar{c}_n) \wedge (\bar{b}_n \to \bar{d}_n))$ растёт линейно (как функция от n). Аналогичное утверждение верно для шенноновской энтропии.*

Эта теорема дает отрицательный ответ на вопрос, поставленный А. Шенем на одном из заседаний колмогоровского семинара в МГУ в 1994 году: верно ли, что колмогоровская

сложность любого множества, полученного из синглетонов $\{a_1\}, \ldots, \{a_k\}$ с помощью операций $\vee, \wedge, \to$ выражается с логарифмической точностью через компоненты сложностного вектора $a_1, \ldots, a_k$? Операции над множествами понимаются в следующем смысле:

- $A \wedge B = \{\langle a, b\rangle \mid a \in A, b \in B\}$
- $A \vee B = \{0a \mid a \in A\} \cup \{1b \mid b \in B\}$
- $A \to B = \{p \mid [p](x) \in B$ для всякого $x \in A\}$

Для всех изученных к тому времени (1994 год) конкретных выражений из операций $\vee, \wedge, \to$ ответ на вопрос Шеня положителен. Вот список тех из них, для которых этот вопрос не тривиален, с указанием сложности соответствующего множества. Все равенства и неравенства справедливы с точностью до слагаемого $O(\log K(a, b, \ldots))$. (Для упрощения обозначений мы пишем a вместо $\{a\}$ для обозначения множества, состоящего из единственного слова a.)

- $K((a \to b) \wedge c) = K(c) + K(b|a, c)$.
  Верхняя оценка $K((a \to b) \wedge c) \leqslant K(c) + K(b|a, c)$ очевидна. Нижняя оценка получается следующим образом. Пусть p — программа, преобразующая a в b, для которой $K(p, c)$ минимально. Докажем, что $K(p, c) \geqslant K(c) + K(b|a, c)$. Ясно, что $K(b|a, c) \leqslant K(p|c)$ (любую программу q, преобразующую c в p можно использовать и для преобразования пары $\langle a, c\rangle$ в b: сначала найдем p, применив q к c, затем применим p к a). Пользуясь равенством $K(p|c) = K(p, c) - K(c)$, получим $K(p, c) \geqslant K(c) + K(b|a, c)$, что и требовалось доказать.
- $K((a \to b) \wedge (b \to a)) = \max\{K(b|a), K(a|b)\}$.
  Нижняя оценка $K((a \to b) \wedge (b \to a)) \geqslant \max\{K(b|a), K(a|b)\}$ здесь очевидна, а верхняя была получена в работе [1].
- $K((a \to c) \wedge (b \to c)) = \max\{K(c|a), K(c|b)\}$.
  Здесь также нижняя оценка $K((a \to c) \wedge (b \to c)) \geqslant \max\{K(c|a), K(c|b)\}$ очевидна, а верхняя была получена в работе [5].
- $K((a \to b) \wedge (b \to c)) = \max\{K(b, c|a), K(c|b)\}$.
  Любое решение $\langle p, q\rangle$ этой задачи можно вычислимым образом переделать в элемент множества $a \to (b \wedge c)$, поскольку $c = [q]([p](a))$. Поэтому это множество эквивалентно множеству $(a \to (b \wedge c)) \wedge (b \to c)$. (Эквивалентность двух множеств означает, что имеется алгоритм, который для всех a, b, c по любому элементу первого множества даёт некоторый элемент второго множества, и наоборот. Ясно что сложности эквивалентных множеств отличаются на $O(1)$.) Последнее множество эквивалентно множеству $(a \to (b \wedge c)) \wedge (b \to (b \wedge c))$. По предыдущему пункту его сложность равна $\max\{K(b, c|a), K(c|b)\}$.
- $K((a \to b) \to c) = \min\{K(c), K(a) + K(c|a, b)\}$.
  Верхняя оценка здесь очевидна, а нижняя может быть доказана методом из работы [6] (где доказано более слабое неравенство $K((a \to b) \to c) \geqslant \min\{K(c), K(a)\}$).

2. Сложность задачи $(a \to c) \wedge (b \to d)$

Какие имеются оценки сложности задачи $(a \to c) \wedge (b \to d)$? Наилучшие известные авторам оценки даются в следующей теореме. (В ее формулировке и далее мы игнорируем слагаемые порядка логарифма от сложности кортежа, состоящего из всех слов, входящих в формулировку.)

Теорема 3.

$$K((a \to c) \wedge (b \to d)) \leqslant \min\{K(c|a) + K(d|b), K(d|b,c) + K(c), K(c|a,d) + K(d)\},$$
$$K((a \to c) \wedge (b \to d)) \geqslant \max\{K(b,c,d|a) - K(b|a,c), K(a,c,d|b) - K(a|b,d)\}.$$

Доказательство. По определению, решением нашей задачи является любая пара программ $\langle p, q\rangle$ такая, что p преобразует a в c, а q преобразует b в d. Взяв в качестве p и q кратчайшие такие программы, мы получим верхнюю оценку $K(c|a) + K(d|b)$ для её сложности. Взяв в качестве p программу, выдающую c на любом входе, а в качестве q программу, спаривающую данный вход с c и применяющую к полученной паре минимальную программу преобразования пары $\langle b, c\rangle$ в d, получим верхнюю оценку $K(d|b,c) + K(c)$. Верхняя оценка $K(c|a,d) + K(d)$ доказывается аналогичным образом.

Чтобы получить нижнюю оценку, рассмотрим тройку $\langle p, q, r\rangle$, где r — кратчайшая программа преобразования пары $\langle a, c\rangle$ в b. Имея эту тройку и a, можно найти b, c, d: применив p к a, находим c, применив r к $\langle a, c\rangle$, находим b, применив q к b, находим d. Поэтому

$$K(b,c,d|a) \leqslant K(p,q) + K(r) = K(p,q) + K(b|a,c),$$

следовательно,

$$K(b,c,d|a) - K(b|a,c) \leqslant K(p,q).$$

Аналогично доказывается оценка

$$K(a,c,d|b) - K(a|b,d) \leqslant K(p,q). \quad \square$$

Рассуждения, примененные в доказательстве, переносятся на энтропию Шеннона, поэтому аналог теоремы 3 справедлив для энтропии Шеннона (также с логарифмической точностью). Для доказательства первого неравенства

$$H((\alpha \to \gamma) \wedge (\beta \to \delta)) \leqslant H(\gamma|\alpha) + H(\delta|\beta) + O(\log H(\alpha,\beta,\gamma,\delta))$$

рассмотрим случайные величины ρ, σ, где ρ, σ случайные величины минимальной энтропии, для которых $H(\gamma|\alpha,\rho) = H(\delta|\beta,\sigma) = 0$. По теореме 2 их энтропии близки, соответственно, к $H(\gamma|\alpha)$, $H(\delta|\beta)$. Применение этой теоремы дает только отдельно совместное распределение γ, α, ρ и совместное распределение δ, β, σ. Очевидно, эти распределения можно продолжить до совместного распределения всех шести случайных величин так, чтобы совместное распределение исходных четырех случайных величин не изменилось. Тогда случайная величина $\langle \rho, \sigma\rangle$ удостоверяет искомое неравенство.

Аналогичным образом доказывается неравенство

$$H((\alpha \to \gamma) \wedge (\beta \to \delta)) \leqslant H(\delta|\beta,\gamma) + H(\gamma) + O(\log H(\alpha,\beta,\gamma,\delta))$$

и симметричное ему. Неравенство

$$H((\alpha \to \gamma) \wedge (\beta \to \delta)) + H(\beta|\alpha,\gamma) \geqslant H(\beta,\gamma,\delta|\alpha)$$

(и симметричное ему) верно безо всяких поправочных членов. Действительно, рассмотрим случайную величину ρ, реализующую определение совместной условной энтропии. Заменим в доказываемом неравенстве совместную энтропию на $H(\rho)$ и добавим в его левую часть (нулевые) слагаемые $H(\gamma|\rho,\alpha)$ и $H(\delta|\rho,\beta)$. В результате, как нетрудно проверить, получится неравенство, верное для любых случайных величин.

Пример 1. Пусть a, b, c, d получены разрезанием случайного слова длины $4n$ на четыре блока длины n. Тогда верхняя и нижняя оценки из теоремы равны $2n$, и, следовательно, такова сложность нашей задачи в этом случае. Для шенноновского случая аналогичный пример состоит их независимых случайных величин $\alpha, \beta, \gamma, \delta$, равномерно распределенных среди двоичных слов длины n.

Пример 2. Нетрудно предъявить слова a, b, c, d, для которых верхняя оценка в теореме 3 больше нижней, и при этом совместная сложность равна нижней оценке. Например, возьмём случайное слово длины $2n$, разрежем его на два блока x, y длины n и положим $a = d = x$, $b = c = y$. Для этих a, b, c, d нижняя оценка равна n, а верхняя равна $2n$. При этом сложность задачи $(a \to c) \wedge (b \to d)$ равна n, поскольку можно взять $p = q = x \oplus y$ Для шенноновского случая аналогичный пример состоит из независимых случайных величин α, β равномерно распределенных среди слов длины n и $\gamma = \beta$, $\delta = \alpha$.

Доказательство теоремы 2

Будем называть четверки, построенные способом, применённым в предыдущих примерах, "стандартными". Точнее, стандартными называются четверки слов, составленые из фиксированного числа блоков некоторого одного случайного слова. Можно доказать, что если четверка слов a, b, c, d стандартна, то $K((a \to c) \wedge (b \to d))$ равно нижней оценке теоремы 3 (с точностью до константы, зависящей от количества блоков, на которые разбито исходное случайное слово). Поэтому хотя бы одна из четверок слов теоремы 2 должна быть нестандартной, и для нее $K((a \to c) \wedge (b \to d))$ должно быть существенно больше нижней оценки теоремы 3. Этой четверкой будет $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$. Для ее построения мы применим метод из работы [3], основанный на линейной алгебре над конечными полями. Вторая четверка $\bar{a}, \bar{b}, \bar{c}, \bar{d}$ будет стандартной.

Сначала мы определим стандартную четверку. Для этого возьмём случайное бинарное слово длины $7n$ и разрежем его на 7 блоков u, v, w, p, q, r, s длины n. Положим $\bar{a} = uvws$, $\bar{b} = pqrs$, $\bar{c} = ups$, $\bar{d} = vqs$. Нетрудно проверить, что сложностной вектор у этой четверки равен:

$$K(\bar{a}) = K(\bar{b}) = 4n, \quad K(\bar{c}) = K(\bar{d}) = 3n,$$
$$K(\bar{a}, \bar{b}) = 7n, \quad K(\bar{a}, \bar{c}) = K(\bar{a}, \bar{d}) = K(\bar{b}, \bar{c}) = K(\bar{b}, \bar{d}) = K(\bar{c}, \bar{d}) = 5n,$$
$$K(\bar{a}, \bar{c}, \bar{d}) = K(\bar{b}, \bar{c}, \bar{d}) = 6n, \quad K(\bar{a}, \bar{b}, \bar{c}) = K(\bar{a}, \bar{b}, \bar{d}) = K(\bar{a}, \bar{b}, \bar{c}, \bar{d}) = 7n.$$

Отсюда следует, что нижняя оценка для $K((\bar{a} \to \bar{c}) \wedge (\bar{b} \to \bar{d}))$ из теоремы 3 есть $n + O(\log n)$. При этом $K((\bar{a} \to \bar{c}) \wedge (\bar{b} \to \bar{d})) \leqslant n + O(1)$, поскольку, зная побитовую сумму слов p и v (по модулю 2), можно $\bar{a}$ преобразовать в $\bar{c}$, а $\bar{b}$ в $\bar{d}$.

Теперь построим нестандартную четверку. Обозначим через F_n, поле, состоящее из 2^n элементов. Мы будем рассматривать точки, прямые и плоскости в трёхмерном афинном пространстве над полем F_n. Всего в трёхмерном афинном пространстве над полем F_n имеется 2^{3n} точек, $2^{4n+o(1)}$ прямых (количество прямых в точности равно $\frac{2^{3n}(2^{3n}-1)}{2^n(2^n-1)}$) и $2^{3n+o(1)}$ плоскостей. В качестве $\langle \tilde{a}, \tilde{b} \rangle$ мы возьмём "случайную" пару различных пересекающихся прямых, $\tilde{c}$ будет их точкой пересечения, а $\tilde{d}$ — плоскостью, в которой обе они лежат (мы называем прямые пересекающимися, если они имеют единственную общую точку). Чтобы определить понятие случайной пары пересекающихся прямых, заметим, что общее количество пар пересекающихся прямых равно $2^{7n+o(1)}$. Выберем любую целую константу ε такую, что $2^{7n-\varepsilon}$ не превосходит этого количества и назовём *случайной* любую пару, сложность которой при

известном n не меньше $7n-\varepsilon$. Такая пара обязательно найдется, поскольку количество слов длины строго меньше $7n-\varepsilon$ равно $1+2+\dots+2^{7n-\varepsilon-1}<2^{7n-\varepsilon}$.

Нетрудно подсчитать, что сложностной вектор четверки $\tilde a,\tilde b,\tilde c,\tilde d$ тот же, что и у четверки $\bar a,\bar b,\bar c,\bar d$ (с точностью до постоянного слагаемого). Нам осталось установить, что $K((\tilde a\to\tilde c)\wedge(\tilde b\to\tilde d))$ больше cn для некоторого $c>1$. (Напомним, что $K((\bar a\to\bar c)\wedge(\bar b\to\bar d))\leqslant n+O(1)$.)

Для шенноновского случая пример еще проще: случайная величина $\langle\tilde\alpha,\tilde\beta\rangle$ равномерно распределена среди всех пар пересекающихся прямых, $\tilde\gamma$ есть их общая точка, а $\tilde\delta$ — плоскость, в которой они обе лежат. Итак, нам достаточно доказать следующую теорему.

Теорема 4. *$K((\tilde a\to\tilde c)\wedge(\tilde b\to\tilde d))\geqslant 1.5n-O(\log n)$ и $H((\tilde\alpha\to\tilde\gamma)\wedge(\tilde\beta\to\tilde\delta))\geqslant 1.5n-O(1)$.*

Доказательство. Докажем первое неравенство. Пусть p принадлежит множеству $(\tilde a\to\tilde c)\wedge(\tilde b\to\tilde d)$, то есть p — это пара программ $\langle q,r\rangle$, первая из которых преобразует $\tilde a$ в $\tilde c$, а вторая $\tilde b$ в $\tilde d$. Рассмотрим множество S, состоящее из всех пар различных прямых a,b таких, что программа p преобразует a в общую точку a и b, а программа q преобразует b в плоскость, содержащую a и b. При известных q, r и n мы можем перечислять все элементы S. Так как пара $\langle\tilde a,\tilde b\rangle$ принадлежит S, мы можем заключить, что

$$7n-O(1)\leqslant K(\tilde a,\tilde b)\leqslant K(p)+\log|S|+O(\log\log|S|).$$

Таким образом, нам достаточно доказать, что количество пар в S не превосходит $2^{5.5n+O(1)}$. Это непосредственно следует из следующей комбинаторной леммы.

Лемма 1. *Пусть f — некоторая функция, сопоставляющая каждой прямой некоторую точку на этой прямой, а g — некоторая функция, сопоставляющая каждой прямой некоторую плоскость, содержащую эту прямую. Пусть множество S состоит из всех пар прямых $\langle a,b\rangle$ таких, что $f(a)$ принадлежит b, а плоскость $g(b)$ содержит a. Тогда $|S|$ не превосходит $2^{5.5n+O(1)}$.*

Доказательство. Сначала посмотрим, какая оценка получится, если не прибегать к хитростям. Для каждой прямой b существует не более $2^{2n+o(1)}$ прямых a в плоскости $g(b)$, поэтому мощность S не более, чем в $2^{2n+o(1)}$ раз, превосходит количество прямых $(2^{4n+o(1)})$, что даёт верхнюю оценку $|S|\leqslant 2^{6n+o(1)}$. Точно такую же оценку мы получим, если при каждом a подсчитаем количество b, проходящих через $f(a)$. Заметим, что в первом подсчёте мы не использовали того, что прямая b должна содержать $f(a)$, а во втором — того, что прямая a должна лежать в плоскости $g(b)$. Наш план таков: мы модифицируем первое из рассуждений, показав, что в S в среднем на каждую прямую b приходится не более $2^{1.5n+O(1)}$ прямых a. При этом мы уже будем учитывать условие $f(a)\in b$. (Можно действовать и симметричным образом — показать, что в S в среднем на каждую прямую a приходится не более $2^{1.5n+O(1)}$ прямых b.)

Разобьём S на слои, поместив в один слой пары $\langle a,b\rangle$ с одинаковым значением $g(b)$. Мы ограничим сверху количество пар в каждом слое, а затем просуммируем полученные оценки. Итак, фиксируем плоскость d и оценим сверху количество пар $\langle a,b\rangle\in S$, для которых $g(b)=d$.

Для этого рассмотрим произвольную точку c на плоскости d и обозначим через A_c множество прямых a в плоскости d с $f(a)=c$, а через B_c множество прямых b, содержащих точку c, для которых $g(b)=d$. Ясно, что количество пар $\langle a,b\rangle\in S$, для которых $g(b)=d$, не превосходит

$$\sum_c|A_c||B_c|\leqslant\sqrt{\sum_c|A_c|^2\sum_c|B_c|^2}.$$

Обе суммы в правой части этого неравенства легко ограничить сверху, поскольку они имеют ясный смысл. А именно, $\sum_c |A_c|^2$ задаёт вероятность того, что для двух равномерно и независимо выбранных прямых a', a'' в плоскости d будет выполнено $f(a') = f(a'')$. Точнее, пусть N обозначает общее количество прямых в плоскости d (примерно равное 2^{2n}). Тогда

$$\begin{aligned}\operatorname{Prob}[f(a') = f(a'')] &= \sum_c \operatorname{Prob}[f(a') = f(a'') = c] \\ &= \sum_c \operatorname{Prob}[f(a') = c] \operatorname{Prob}[f(a'') = c] = \sum_c |A_c|^2/N^2.\end{aligned}$$

При любом фиксированном a' вероятность события $f(a') = f(a'')$ не превосходит вероятности того, что прямая a'' проходит через точку $f(a')$. Последняя вероятность равна $2^{-n+o(1)}$, поэтому

$$\operatorname{Prob}[f(a') = f(a'')] \leqslant 2^{-n+o(1)} \implies \sum_c |A_c|^2 \leqslant N^2 2^{-n+o(1)} = 2^{3n+o(1)}.$$

Вторая сумма $\sum_c |B_c|^2$ задаёт среднее количество общих точек у двух равномерно и независимо выбранных прямых b', b'' в множестве M_d, состоящем из всех прямых b, для которых $g(b) = d$ (все они лежат в плоскости d). Точнее,

$$\begin{aligned}\mathbf{E}\,|b' \cap b''| &= \sum_c \operatorname{Prob}[c \in b' \cap b''] \\ &= \sum_c \operatorname{Prob}[c \in b'] \operatorname{Prob}[c \in b''] = \sum_c |B_c|^2/|M_d|^2.\end{aligned}$$

Любые две различных прямых имеют не более одной общей точки, а равные прямые имеют 2^n общих точек, поэтому

$$\mathbf{E}\,|b' \cap b''| \leqslant 1 + 2^n \operatorname{Prob}[b' = b''] = 1 + 2^n/|M_d|,$$

следовательно,

$$\sum_c |B_c|^2 = \mathbf{E}\,|b' \cap b''||M_d|^2 \leqslant |M_d|^2 + |M_d|2^n.$$

Напомним, что количество $\langle a, b\rangle \in S$, для которых $g(b) = d$, не превосходит $\sqrt{\sum_c |A_c|^2 \sum_c |B_c|^2}$, поэтому оно не больше чем

$$\sqrt{2^{3n+o(1)}(|M_d|^2 + |M_d|2^n)} \leqslant 2^{1.5n+o(1)}(|M_d| + 2^n)$$

(последнее неравенство устанавливается простым возведением в квадрат). Осталось просуммировать полученные оценки по всем d:

$$|S| \leqslant 2^{1.5n+o(1)} \sum_d (|M_d| + 2^n).$$

Семейства прямых M_d образуют разбиение множества всех прямых, следовательно сумма мощностей M_d равна общему количеству прямых $2^{4n+o(1)}$. Кроме того, количество различных d равно $2^{3n+o(1)}$, поэтому сумма $2^{n+o(1)}$ по всем d также равна $2^{4n+o(1)}$. Следовательно,

$$|S| \leqslant 2^{1.5n+o(1)}(2^{4n+o(1)} + 2^{4n+o(1)}) = 2^{5.5n+1+o(1)}. \quad \square$$

Итак, первое неравенство теоремы доказано. Второе неравенство также следует из леммы 1. Пусть имеется случайная величина ρ, для которой $H(\tilde{\gamma}|\rho,\tilde{\alpha}) = H(\tilde{\delta}|\rho,\tilde{\beta}) = 0$. Это означает, что для некоторых функций F, G в вероятностью 1 выполнены равенства $\tilde{\gamma} = F(\rho,\tilde{\alpha})$, $\tilde{\delta} = G(\rho,\tilde{\beta})$. Нам нужно доказать, что $H(\rho) \geqslant 1.5n - O(1)$.

Для этого зафиксируем произвольное возможное значение p случайной величины ρ. Рассмотрим функции $f(a) = F(p,a)$ и $g(b) = F(p,b)$. Нам дано, что при условии $\rho = p$ с вероятностью 1 выполнено $\tilde{\gamma} = f(\tilde{\alpha})$, $\tilde{\delta} = g(\tilde{\beta})$ (без ограничения общности мы можем считать что все возможные значения ρ имеют положительную вероятность). Значит оба равенства выполнены для всех пар a, b пересекающихся прямых, имеющих положительную вероятность при условии $\rho = p$. По лемме 1 таких пар не больше $2^{5.5n+O(1)}$. По свойству энтропии, мы можем заключить, что $H(\tilde{\alpha},\tilde{\beta}|\rho = p) \leqslant 5.5n + O(1)$. Отсюда следует, что $H(\tilde{\alpha},\tilde{\beta}|\rho)$ также не превосходит $5.5n + O(1)$. Но поскольку $\langle\tilde{\alpha},\tilde{\beta}\rangle$ выбирается с равномерным распределением мы получаем

$$7n - O(1) \leqslant H(\tilde{\alpha},\tilde{\beta}) \leqslant H(\tilde{\alpha},\tilde{\beta}|\rho) + H(\rho) \leqslant 5.5n + O(1) + H(\rho),$$

откуда следует требуемая оценка. Теорема доказана. □

Заметим, что изученное нами выражениие $(p \to q) \wedge (r \to s)$, как и приведённые ранее примеры, имеет глубину вложенности 2. Имеется ещё одно выражение глубины 2, для которого при подставлении синглетонов сложность полученного множества не выражается через сложностной вектор подставляемых слов. Это выражение есть $(p \vee q) \to (r \vee s)$. Любая программа, решающая эту задачу, должна по любому из слов $0p, 1q$ дать одно из слов r, s. Сделаем в эту формулу подстановку $p = a$, $q = b$, $r = \langle a, c\rangle$, $s = \langle b, d\rangle$, где a, b, c, d какие-нибудь слова с таким же сложностным вектором, что у двух построенных четвёрок. При таком сложностном векторе по a явно проще найти $\langle a, c\rangle$, чем $\langle b, d\rangle$, а по b явно проще найти $\langle b, d\rangle$, чем $\langle a, c\rangle$. Поэтому сложность получившегося множества та же, что и у множества $(a \to \langle a, c\rangle) \wedge (b \to \langle b, d\rangle)$. Последнее же эквивалентно множеству $(a \to c) \wedge (b \to d)$, и подставляя вместо a, b, c, d только что построенные четвёрки, мы получаем задачи существенно разной сложности.

Простым перебором можно установить, что для всех остальных выражений глубины 2 ответ на вопрос Шеня положительный.

2.1. Усиление теоремы 4

В примере, использованном в доказательстве теоремы 4, имеется разрыв примерно $0.5n$ между полученной нижней оценкой для $K((a \to c) \wedge (b \to d))$ и верхней оценкой теоремы 3. Возникает естественный вопрос: бывает ли так, что $K((a \to c) \wedge (b \to d))$ равно верхней оценке теоремы 3, причем эта верхняя оценка значительно меньше, чем нижняя оценка той же теоремы? Следующая теорема даёт утвердительный ответ и на этот вопрос.

Теорема 5. *Для каждого n существует четвёрка слов $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$ такая, что сложности всех слов $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$ равны n, сложности всех пар слов $\langle\tilde{a},\tilde{b}\rangle$, $\langle\tilde{a},\tilde{c}\rangle$, $\ldots$, $\langle\tilde{c},\tilde{d}\rangle$ равны $2n$, сложности всех троек $\langle\tilde{a},\tilde{b},\tilde{c}\rangle$, $\ldots$ равны $3n$, сложность четвёрки $\langle\tilde{a},\tilde{b},\tilde{c},\tilde{d}\rangle$ равна также $3n$ и сложность задачи $(\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})$ равна $2n$ (все равенства верны с точностью до слагаемого $O(\log n)$).*

Аналогичное утверждение (с той же точностью) выполнено и для шенноновской энтропии.

Нетрудно подсчитать, что для такой четвёрки нижняя оценка теоремы 3 равна n, а верхняя оценка той же теоремы равна $2n$, так что мы действительно получаем требуемый пример.

Доказательство. Фиксируем n. Назовём множество S четвёрок слов длины n *равномерным*, если для каждой тройки слов $\langle a, b, c\rangle$ слов длины n существует ровно одно d, для которого четвёрка $\langle a, b, c, d\rangle$ принадлежит S. Для каждого n мы определим некоторое равномерное множество S, которое можно будет найти эффективно по данному n. В качестве $\langle \tilde a, \tilde b, \tilde c, \tilde d\rangle$ мы возьмём произвольный случайный элемент S, то есть любую четверку из S, для которой $K(\tilde a, \tilde b, \tilde c, \tilde d) \geqslant 3n$. Такая четверка обязательно найдется, поскольку количество слов длины строго меньше $3n$ равно $1 + 2 + \dots + 2^{3n-1} < 2^{3n}$. Этим мы обеспечим равенства $K(\tilde a, \tilde b, \tilde c, \tilde d) = K(\tilde a, \tilde b, \tilde c) = 3n$, а следовательно и равенства $K(\tilde a) = K(\tilde b) = K(\tilde c) = n$, $K(\tilde a, \tilde b) = K(\tilde a, \tilde c) = K(\tilde b, \tilde c) = 2n$. Чтобы было выполнено неравенство $K((\tilde a \to \tilde c) \wedge (\tilde b \to \tilde d)) \geqslant 2n$, мы сделаем так, что оно было выполнено для большинства четвёрок $\langle a, b, c, d\rangle$ из S (точнее, доля четвёрок, для которых это не выполнено, будет $O(1/n)$). Тогда это неравенство будет выполнено и для любой случайной четвёрки. Чтобы были выполнены остальные требования на сложностной вектор, достаточно чтобы и тройки $\langle \tilde a, \tilde b, \tilde d\rangle$, $\langle \tilde a, \tilde c, \tilde d\rangle$ и $\langle \tilde b, \tilde c, \tilde d\rangle$ были случайными. Заметим, что о случайности последней тройки можно специально не беспокоиться, так это гарантируется неравенством $K((\tilde a \to \tilde c) \wedge (\tilde b \to \tilde d)) \geqslant 2n$. Действительно, из него следует

$$2n \leqslant K((\tilde a \to \tilde c) \wedge (\tilde b \to \tilde d)) \leqslant K(\tilde d|\tilde b, \tilde c) + K(\tilde c) = K(\tilde d|\tilde b, \tilde c) + n \leqslant 2n,$$

откуда

$$K(\tilde d|\tilde b, \tilde c) = n \Longrightarrow K(\tilde b, \tilde c, \tilde d) = K(\tilde b, \tilde c) + K(\tilde d|\tilde b, \tilde c) = 2n + n = 3n.$$

Случайность остальных двух троек $\langle \tilde a, \tilde b, \tilde d\rangle$, $\langle \tilde a, \tilde c, \tilde d\rangle$ мы обеспечим точно так же. А именно мы сделаем так, чтобы не только сложность задачи $(\tilde a \to \tilde c) \wedge (\tilde b \to \tilde d)$ была больше $2n$ для большинства всех элементов S, но и сложность каждой из задач $(\tilde c \to \tilde b) \wedge (\tilde a \to \tilde d)$ и $(\tilde b \to \tilde a) \wedge (\tilde c \to \tilde d)$ была не меньше $2n$ для большинства элементов S. Симметричными рассуждениями отсюда выводится случайность троек $\langle \tilde a, \tilde b, \tilde d\rangle$ и $\langle \tilde a, \tilde c, \tilde d\rangle$ для любой случайной четвёрки $\langle \tilde a, \tilde b, \tilde c, \tilde d\rangle$ из S.

Итак, положим $k = O(\log n)$, а константу в $O(\log n)$ подберём позже. Нам достаточно при данном n построить равномерное множество S такое, что каждое из неравенств

$$\begin{aligned} &K((a \to c) \wedge (b \to d)) \geqslant 2n - k,\\ &K((c \to b) \wedge (a \to d)) \geqslant 2n - k,\\ &K((b \to a) \wedge (c \to d)) \geqslant 2n - k, \end{aligned} \tag{1}$$

выполнено для по крайней мере $(1 - O(1/n))$-ой части всех $\langle a, b, c, d\rangle$ из S.

Что означает первое из неравенств (1)? Оно означает, что нет пары программ $\langle p, q\rangle$ сложности меньше $2n - k$, для которой $[p](a) = c$, $[q](b) = d$. Иными словами, у нас имеется некоторое небольшое множество M пар функций (содержащее менее 2^{2n-k} элементов). Нам надо, чтобы для большинства $\langle a, b, c, d\rangle$ из S не существовало такой пары $\langle f, g\rangle$ из M, что $f(a) = c$ и $g(b) = d$. Поскольку множество пар программ сложности меньше $2n - k$ нам неизвестно, мы построим множество S такое, что любое малое множество пар функций M обладает требуемым свойством. Далее через M будем обозначать произвольное множество, состоящее из 2^{2n-k} пар всюду определённых функций, отображающих слова длины n в слова длины n.

Лемма 2. *При подходящем выборе $k = O(\log n)$ для всех n существует равномерное множество S такое, что для каждого M ни одно из множеств*

$$\begin{aligned} &\{\langle a, b, c, d\rangle \in S \mid \exists \langle f, g\rangle \in M \;\; f(a) = c,\;\; g(b) = d\},\\ &\{\langle a, b, c, d\rangle \in S \mid \exists \langle f, g\rangle \in M \;\; f(b) = a,\;\; g(c) = d\},\\ &\{\langle a, b, c, d\rangle \in S \mid \exists \langle f, g\rangle \in M \;\; f(c) = b,\;\; g(a) = d\} \end{aligned}$$

не содержит более $O(1/n)$-ой части всех четвёрок из S.

Доказательство. Мы докажем, что случайно выбранное равномерное множество S с большой вероятностью удовлетворяет утверждению леммы. Ввиду симметрии, достаточно доказать, что с вероятностью, близкой к 1, первое из указанных множеств не содержит более $O(1/n)$-ой части всех четвёрок из S ни для одного M.

Скажем, что множество M обслуживает четвёрку $\langle a,b,c,d\rangle$, если $f(a)=c$ и $g(b)=d$ для некоторой пары $\langle f,g\rangle \in M$. Фиксируем M и оценим сверху вероятность того, что для случайного S множество M обслуживает более $O(1/n)$-ой части элементов S.

Назовём тройку $\langle a,b,c\rangle$ *плохой*, если при случайном выборе пары $\langle f,g\rangle$ из M вероятность события $f(a)=c$ больше $n\cdot 2^{-n}$ (это свойство не зависит от b). В противном случае назовём тройку *хорошей*. Среднее значение вероятности $\mathrm{Prob}[f(a)=c]$, когда $\langle a,b,c\rangle$ выбирается случайно, равно 2^{-n}. Поэтому доля троек, для которых $\mathrm{Prob}[f(a)=c]>n\cdot 2^{-n}$, меньше $1/n$. Следовательно, плохих троек менее $(1/n)2^{3n}$, а хороших более $(1-1/n)2^{3n}$.

Для любой хорошей тройки $\langle a,b,c\rangle$ вероятность того, что для случайно выбранного d множество M обслуживает четвёрку $\langle a,b,c,d\rangle$ не превосходит $O(1/n)$. Действительно, если M обслуживает четвёрку $\langle a,b,c,d\rangle$, то d принадлежит множеству $\{g(b)\mid \langle f,g\rangle\in M,\ f(a)=c\}$. Поскольку тройка $\langle a,b,c\rangle$ хорошая, это множество имеет менее $n2^{-n}|M| = 2^{\log n-n+2n-k} < 2^n/n$ элементов (если выбрать k так, чтобы $k>2\log n$). Следовательно, вероятность того, что d принадлежит этому множеству, меньше $O(1/n)$.

Теперь воспользуемся оценкой Чернова [2]: если проводить N независимых испытаний, вероятность успеха в каждом из которых равна p, то для любого $0<\varepsilon\leqslant p(1-p)$ доля успешных испытаний будет меньше $p+\varepsilon$ с вероятностью не меньше $1-2^{-\varepsilon^2 N/(2p)}$. Мы будем применять следующее его очевидное следствие, получаемое при $\varepsilon=p/2$: если вероятность успеха в отдельном испытании не превосходит p, то вероятность того, что доля успешных испытаний больше $3p/2$, меньше $2^{-pN/8}$. В нашем случае испытания нумеруются хорошими тройками $\langle a,b,c\rangle$, так что $N\geqslant 2^{3n-O(1)}$. Испытание с номером $\langle a,b,c\rangle$ заключается в случайном выборе d, а успехом является то, что четвёрка $\langle a,b,c,d\rangle$ обслуживается множеством M. Вероятность успеха p есть $O(1/n)$. Поэтому для случайно выбранного S с вероятностью не менее $1-2^{-\Omega(N/n)}$ для не более чем $O(N/n)$ хороших троек $\langle a,b,c\rangle$ четверка $\langle a,b,c,d\rangle$ из S обслужена M. Количество обслуженных четверок из S не превосходит суммы количества хороших обслуженных четверок и количества плохих четверок (четверка $\langle a,b,c,d\rangle$ называется хорошей, если хороша тройка $\langle a,b,c\rangle$). Первое слагаемое не превосходит $O(N/n)=O(2^{3n}/n)$ с вероятностью не менее $1-2^{-\Omega(2^{3n}/n)}$, а второе всегда меньше $2^{3n}/n$. Поэтому с указанной вероятностью количество обслуженных четвёрок из S не больше $2^{3n}/n+O(2^{3n}/n)=O(2^{3n}/n)$.

Количество различных M не превосходит

$$((2^n\cdot 2^n)^{2^n})^{2^{2n-k}} = 2^{2^{3n-k+\log 2n}}.$$

Следовательно, с вероятностью не менее

$$1-2^{2^{3n-k+\log 2n}-2^{3n-\log n-O(1)}}$$

любое M обслуживает не более $O(1/n)$-ой части всех четвёрок из S. При подходящем выборе k эта вероятность стремится к 1 при n стремящемся к бесконечности. □

Свойства множества S, сформулированные в лемме, алгоритмически разрешимы. Поэтому по данному n мы можем перебором найти первое множество S, удовлетворяющее утверждению леммы. Как уже объяснялось, любая случайная четвёрка из S удовлетворяет условию теоремы.

Осталось доказать аналогичное утверждение для энтропии Шеннона. Выберем произольное множество S, удовлетворяющее лемме 2. В качестве $\tilde\alpha, \tilde\beta, \tilde\gamma$ возьмем независимые равномерно распределенные слова длины n, а $\tilde\delta$ положим равным такому слову d, для которого четверка $\tilde\alpha, \tilde\beta, \tilde\gamma, d$ принадлежит S.

Докажем, что $H((\tilde\alpha \to \tilde\gamma) \wedge (\tilde\beta \to \tilde\delta)) \geqslant 2n - O(\log n)$. Пусть имеется случайная величина ρ, для которой $H(\tilde\gamma|\rho, \tilde\alpha) = H(\tilde\delta|\rho, \tilde\beta) = 0$. Это означает, что для некоторых функций F, G в вероятностью 1 выполнены равенства $\tilde\gamma = F(\rho, \tilde\alpha)$, $\tilde\delta = G(\rho, \tilde\beta)$. Нам нужно доказать, что $H(\rho) \geqslant 2n - O(\log n)$.

Вспомним, что $H(\rho)$ определяется как среднее значение минус логарифма вероятности исхода p по всем исходам p случайной величины ρ. Поэтому достаточно доказать, что лишь небольшая доля исходов ρ имеет вероятность менее 2^{-2n+k} (будем такие исходы называть *легкими*, а остальные — *тяжелыми*). Точнее, с помощью леммы 2 мы докажем, что суммарная вероятность всех тяжелых исходов не больше $O(1/n)$. Это даст оценку

$$H(\rho) \geqslant (1 - O(1/n))(2n - k) = (1 - O(1/n))(2n - O(\log n)) \geqslant 2n - O(\log n).$$

Для этого зафиксируем произвольное возможное значение p случайной величины ρ. Рассмотрим функции $f_p(a) = F(p, a)$ и $g_p(b) = F(p, b)$. Нам дано, что при условии $\rho = p$ с вероятностью 1 выполнено $\tilde\gamma = f(\tilde\alpha)$, $\tilde\delta = g(\tilde\beta)$ (без ограничения общности мы можем считать что все возможные значения ρ имеют положительную вероятность). Значит оба равенства выполнены для всех четверок a, b, c, d из S, имеющих положительную вероятность при условии $\rho = p$.

Количество тяжелых исходов не превосходит 2^{2n-k} (так как их суммарная вероятность не больше 1). Поэтому мы можем применить лемму 2 к множеству M, состоящему из пар функций (f_p, g_p) для всех тяжелых p. По лемме 2 для не более чем $O(1/n)$-ой доли четверок a, b, c, d из S найдется тяжелое p, для которого вероятность пятерки a, b, c, d, p положительна. Поэтому суммарная вероятность всех тяжелых исходов не больше $O(1/n)$, что и требовалось доказать.

Аналогичным образом устанавливаются неравенства $H((\tilde\gamma \to \tilde\beta) \wedge (\tilde\alpha \to \tilde\delta)) \geqslant 2n - O(\log n)$ и $H((\tilde\beta \to \tilde\alpha) \wedge (\tilde\gamma \to \tilde\delta)) \geqslant 2n - O(\log n)$. Отсюда, точно так же, как и для случая колмогоровской сложности, следует, что четверка $\tilde\alpha, \tilde\beta, \tilde\gamma, \tilde\delta$ имеет нужный энтропийный вектор. □

Существует другая четвёрка с тем же сложностным вектором, что и в условии только что доказанной теоремы, но для которой $K((a \to c) \wedge (b \to d)) = n$. А именно, возьмём случайное слово длины $3n$, разрежем его на три части длины n. Это будут a, b и c; затем положим $d = a \oplus b \oplus c$. Зная $a \oplus c$, мы можем по a найти c, а по b найти d, следовательно $K((a \to c) \wedge (b \to d)) = n$. Таким образом, мы получаем второе доказательство теоремы 4.

Интересно сравнить это новое доказательство со старым. В обоих случаях мы доказываем существование некоторой четвёрки слов со специфическими свойствами (а именно четвёрки с данным сложностным вектором и с данной нижней оценкой на сложность задачи $((a \to b) \wedge (c \to d))$. При этом свойства четвёрки во втором случае более сильные (нижняя оценка равна $2n$, а не $1.5n$, как в первом случае). Но зато доказательство менее эффективно. В обоих случаях по данному n мы эффективно строим некоторое множество четверок слов S, а затем берём произвольную случайную четвёрку из S. Но в первом случае множество S предъявляется (это множество четверок, состоящих из двух пересекающихся прямых, их общей точки и общей плоскости), а во втором нет. Доказательства первого типа (мы говорим о доказательствах теорем существования объектов с некоторыми свойствами) в теории колмогоровской сложности принято называть эффективными, а второго типа — квазиэффективными. В обоих случаях существует вероятностный алгоритм, который по данному n

с вероятностью, близкой к 1, порождает некоторый объект с требуемыми свойствами. Но разница состоит в том, что в первом случае этот алгоритм работает полиномиальное от n время (сложение, умножение и деление в поле из 2^n элементов выполнимы за полиномиальное время), а во втором времени требуется значительно больше. В этой связи интересно было бы понять, нельзя ли построить множество S, удовлетворяющее лемме 2, таким образом, чтобы был полиномиальный алгоритм, который по a, b, c находит такое d, что четвёрка $\langle a, b, c, d\rangle$ принадлежит S.

Эффективные и квазиэффективные доказательства можно противопоставить неэффективным — таким доказательствам, в которых для каждого n доказывается лишь существование объекта с требуемыми свойствами, и не даётся никакого вероятностного алгоритма их порождения. Обычно такое доказательство проще придумать, чем (квази)эффективное. Настоящая статья представляет собой одно из исключений: мы не знаем более простого неэффективного доказательства существования a, b, c, d для которых сложность задачи $(a \to c) \wedge (b \to d)$ была бы больше нижней оценки теоремы 3.

В заключение мы приведём еще один пример, когда неожиданно оказывается, что квазиэффективное доказательство примерно такое же по сложности, как и неэффективное.

3. Построение пары слов, имеющих большую взаимную информацию, но не имеющих ничего общего

В работе [3] доказано что для каждого n существуют слова a_n, b_n, для которых $K(a_n) = K(b_n) = 2n$, $K(a_n, b_n) = 3n$ (так что взаимная информация $K(a_n) + K(b_n) - K(a_n, b_n)$ слов a_n и b_n равна n), но при этом a_n, b_n не имеют ничего общего в следующем смысле. Будем пытаться указать слово c небольшой сложности, относительно которого a и b имеют небольшую сложность. Точнее рассмотрим множество троек неотрицательных чисел u, v, w, для которых для почти всех n существует слово c сложности $un + O(\log n)$, для которого $K(a_n|c) \leqslant vn + O(\log n)$ и $K(b_n|c) \leqslant wn + O(\log n)$. Это множество называется положительным профилем a и b и обозначается через M^+_{ab}. Например, если a и b — случайные слова длины $2n$, имеющие общее подслово длины n, то можно взять это подслово в качестве c, и профиль будет содержать тройку $\langle 1, 1, 1\rangle$. Это соответствует максимально возможному количеству общего у a и b (при условии $K(a) = K(b) = 2n$, $K(a, b) = 3n$). Чем больше троек содержит профиль a, b, тем более общего у a, b. Мы говорим, что a и b *не имеют ничего общего*, если их профиль минимальный среди всех профилей пар слов, для которых $K(a) = K(b) = 2n$, $K(a, b) = 3n$. Каков же этот минимальный профиль? Он обозначается $M^+_{\min}$ и состоит из всех троек $\langle u, v, w\rangle$, для которых выполнены неравенства

$$u + v \geqslant 2, \quad u + w \geqslant 2, \quad u + v + w \geqslant 3$$

и **хотя бы одно** из неравенств

$$u + v \geqslant 3, \quad u + w \geqslant 3, \quad u + v + w \geqslant 4.$$

Нетрудно установить, что профиль любых a, b с $K(a) = K(b) = 2n$, $K(a, b) = 3n$ включает в себя $M^+_{\min}$. Для этого надо включать в c куски минимальных описаний a и b, и минимальных описаний a при известном b и b при известном a, и ничего больше. (Мы опускаем подробное доказательство.)

Так вот, в работе [3] приведено неэффективное доказательство следующей теоремы

Теорема 6. *Существуют последовательности a_n, b_n слов, для которых $K(a_n) = K(b_n) = 2n$, $K(a_n, b_n) = 3n$ и профиль которых равен M^+_{min} (равенства понимаются с точностью до слагаемого $O(\log n)$).*

Авторы [3] также указывают, что им неизвестно её квазиэффективное доказательство. Заметим, что конкретные сложности a, b и их пары выбраны в теореме 6 для упрощения формулировки. Аналогичную теорему можно сформулировать и доказать и для произвольных соотношений между сложностями a, b и их пары (при этом иножество $M^+_{\min}$, разумеется, изменится). Приводя ниже квазиэффективное доказательство, мы тоже ограничиваемся для простоты случаем $K(a_n) = K(b_n) = 2n$, $K(a_n, b_n) = 3n$.

Квазиэффективное доказательство теоремы 6. То что, профиль a, b равен $M^+_{\min}$ означает следующее. Любая тройка u, v, w, для которой $u + v < 2$, или $u + w < 2$, или $u + v + w < 3$, или одновременно

$$u + v < 3, \quad u + w < 3, \quad u + v + w < 4$$

не принадлежит профилю a, b. Если выполнено хотя бы одно из неравенств $u+v < 2$, $u+w < 2$, $u + v + w < 3$, то профиль любых a, b с $K(a_n) = K(b_n) = 2n$, $K(a_n, b_n) = 3n$ не содержит тройки $\langle u, v, w\rangle$. Это следует из неравенств $K(a) \leqslant K(c) + K(a|c)$, $K(b) \leqslant K(c) + K(b|c)$, $K(a, b) \leqslant K(c) + K(a|c) + K(b|c)$, соответственно.

Ясно, что достаточно построить последовательности a_n, b_n, слов длины $2n$, для которых $K(a_n, b_n) = 3n + O(\log n)$, $K(a_n) = K(b_n) = 2n + O(\log n)$ с таким свойством. Для для любых U, V, W, удовлетворяющих неравенствам

$$U + \max\{V, W\} < 3n - k, \quad U + V + W < 4n - k, \tag{2}$$

не существует такого c, что

$$K(c) < U, \quad K(a_n|c) < V, \quad K(b_n|c) < W,$$

где k есть некоторая линейная функция от $\log n$, которую мы выберем позднее. Для этого мы по данному n мы построим ориентированный граф, вершинами которого будут слова длины $2n$, а количество рёбер будет равно $2^{3n+O(1)}$. В качестве пары a, b мы возьмём любое случайное ребро графа. Тогда $K(a, b)$ будет примерно равно $3n$. Какие ещё свойства графа нам потребуются? Первое свойство состоит в том, что для любого множества слов M мощности меньше 2^{2n-k}, лишь ничтожная доля всех рёбер графа имеет хотя бы один конец в M. Применяя это свойство к множеству слов длины $2n$ и сложности меньше $2n - k$, мы получим, что для любого случайного ребра $\langle a, b\rangle$ сложности $K(a)$ и $K(b)$ не меньше $2n - k = 2n - O(\log n)$. Кроме того, нам нужно, чтобы большинство рёбер графа не попадало в множество

$$\{\langle a, b\rangle \mid \text{существуют } U, W, W, \text{ удовлетворяющие неравенствам (2), и слово } c \text{ такие, что } K(c) < U, \quad K(a_n|c) < V, \quad K(b_n|c) < W\}.$$

Поскольку это множество нам неизвестно, нам нужно определить некоторое алгоритмически проверяемое комбинаторное свойство этого множества и обеспечить, чтобы для любого множества M с таким свойством большинство рёбер графа не попадало в M. Вот в чём состоит это свойство. Наше множество является объединением по всем U, V, W, удовлетворяющим неравенствам (2), некоторых множеств M_{UVW}, каждое из которых есть объединение 2^U множеств вида $A \times B$, где $|A| < 2^V$, $|B| < 2^W$. Так что, назовём множество пар слов длины $2n$ *специальным*, если оно является объединением по всем U, V, W, удовлетворяющим неравенствам (2), множеств M_{UVW}, каждое из которых есть объединение 2^U множеств вида $A \times B$, где $|A| < 2^V$, $|B| < 2^W$. Таким образом, нам достаточно доказать следующую лемму.

Лемма 3. *Для некоторой функции $k = k_n = O(\log n)$ для всех n существует неориентированный граф с петлями, вершинами которого являются слова длины $2n$, для которого выполнено следующее. 1. Количество рёбер не больше 2^{3n} и не меньше 2^{3n-2}. 2. Для любого множества M мощности меньше 2^{2n-k}, не более $O(2^{3n}/n)$ рёбер графа имеет хотя бы один конец в M. 3. Любое специальное множество содержит не более $O(2^{3n}/n)$ рёбер графа.*

Доказательство. Докажем, что с большой вероятностью случайный граф удовлетворяет требуемым свойствам. Под случайным выбором графа мы понимаем следующее: 2^{3n} раз мы случайным образом выбираем упорядоченную пару вершин (разрешая петли). Ребро принадлежит построенному графу, если оно было выбрано хотя бы один раз. Нам нужно подсчитать вероятность двух указанных в лемме свойств.

1. Для любого фиксированного множества из 2^{3n-2} рёбер вероятность того, что случайно выбранное ребро попадает в это множество, равна 2^{-n-2}, а вероятность того, что все рёбра попадают туда, равна $2^{-(n+2)2^{3n}}$. Количество способов выбрать 2^{3n-2} рёбер не превосходит $2^{4n\cdot 2^{3n-2}} = 2^{n\cdot 2^{3n}}$. Таким образом, вероятность того, что все рёбра попадут в некоторое множество мощности 2^{3n-2} не превосходит $2^{-(n+2)2^{3n}+n\cdot 2^{3n}} = 2^{-2\cdot 2^{3n}} \ll 1$.

2. Вероятность того, что хотя бы один конец случайно выбранного ребра попадёт в фиксированное множество M мощности 2^{2n-k} не превосходит $1/n$ (если $k > \log n$). По неравенству Чернова, вероятность того, что это случится для более чем $O(1/n)$-ой части от 2^{3n} случайно выбранных рёбер, не превосходит $2^{-\Omega(2^{3n}/n)} = 2^{-2^{3n-\log n-O(1)}}$. Количество различных M не превосходит $2^{2^{2n}}$. Таким образом, с вероятностью не менее

$$1 - 2^{2^{2n} - 2^{3n-\log n-O(1)}}$$

для любого множества M мощности 2^{2n-k} не более $O(2^{3n}/n)$ рёбер случайного графа инцидентны M. Эта вероятность стремится к 1.

3. Докажем, что мощность любого специального множества не превосходит $O(2^{4n}/n)$. Действительно, количество различных троек U, V, W не превосходит $O(n^3)$. При фиксированных U, V, W мощность любого объединения 2^U множеств вида $A\times B$, где $|A| < 2^V$, $|B| < 2^W$ не превосходит $2^{U+V+W} \leqslant 2^{4n-k}$. Умножая это число на $O(n^3)$ получаем $2^{4n-k+3\log n+O(1)} = O(2^{4n}/n)$ (если $k > 3\log n$). Следовательно вероятность того, что случайное ребро попадает в фиксированное специальное множество M не превосходит $O(1/n)$. По неравенству Чернова с вероятностью $1 - 2^{-\Omega(2^{3n}/n)}$ не более $O(2^{3n}/n)$ рёбер графа попадают в M. Теперь оценим количество различных специальных множеств. При фиксированных U, V, W их количество не превосходит

$$(2^{2n\cdot 2^V}\cdot 2^{2n\cdot 2^W})^{2^U} = 2^{2n(2^V+2^W)2^U} < 2^{2^{3n-k+\log n+O(1)}}$$

Возведя это число в степень, равную количеству различных троек U, V, W получим верхнюю оценку количества специальных множеств: $2^{2^{3n-k+4\log n+O(1)}}$ Вспомним, что любое специальное множество содержит более $O(2^{3n}/n)$ рёбер с вероятностью меньшей $2^{-2^{3n-\log n-O(1)}}$. При $k = 5\log n + O(1)$ эта вероятность остается малой даже после умножения на количество специальных множеств. Поэтому с вероятностью близкой к 1 все специальные множества имеют $O(2^{3n}/n)$ рёбер. □

Благодарность. Авторы выражают благодарность А. Шеню за постановку вопроса.

Список литературы

[1] C.H. Bennett, P. Gács, M. Li, P. Vitanyi and W. Zurek. *Information Distance, IEEE transactions on Information Theory*, Vol. 44, No. 4, 1407–1423.

[2] H. Chernoff. A measure of assymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–509, 1952.

[3] An. Muchnik, A. Romashchenko, A. Shen, and N. Vereshchagin. Upper semi-lattice of binary strings with the relation "x is simple conditional to y". *Proc. of 14th Annual IEEE Conference on Computational Complexity*, Atlanta, May 4-6, 1999, 114–122.

[4] An. Muchnik and N. Vereshchagin. "Shannon Entropy vs. Kolmogorov Complexity". Computer Science — Theory and Applications: First International Computer Science Symposium in Russia, CSR 2006, St. Petersburg, Russia, June 8-12. 2006. Proceedings. Editors: Dima Grigoriev, John Harrison, Edward A. Hirsch Lecture Notes in Computer Science, vol. 3967 / 2006, pages 281–291.

[5] Muchnik An. A. Conditional complexity and codes, *Theoretical Computer Science*, v. 271 (2002), issues 1–2, p. 97–109.

[6] A. Shen, N. Vereshchagin. Logical operations and Kolmogorov complexity. To be published in *Theoretical Computer Science* 271 (2002) 125–129..

[7] Звонкин А. К., Левин Л. А., Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов. *Успехи математических наук*, т. 25, вып. 6 (156), с. 85–127.