

ФИВТ МФТИ, весна 2010.

Краткий конспект¹ лекций курса *Введение в теорию информации* (А. Ромащенко, весна 2010).

1 Лекция 1, 24 февраля.

1.1 Определение информации по Хартли

Определим *количество информации* в конечном множестве $A \subset \{0, 1\}^*$ как $H(A) = \lceil \log |A| \rceil$. Для многомерных множеств определяем количество информации в каждой его проекции. Например, если $A \subset \{0, 1\}^* \times \{0, 1\}^*$, то

$$H(A) = \lceil \log |A| \rceil, \quad H_1(A) = \lceil \log \pi_1 |A| \rceil, \quad H_2(A) = \lceil \log \pi_2 |A| \rceil.$$

Заметим, что $H(A) \leq H_1(A) + H_2(A)$, причём равенство достигается только если A есть в точности прямое произведение его проекций на первую и вторую координаты (т. е., значения проекций A на первую и вторую координату в естественном смысле независимы).

Определим количество информации в первой компоненте (проекции) A при известной второй компоненте как логарифм максимального количества значений первой координаты A , соответствующих некоторому фиксированному значению второй координаты. Для 2-мерных A это значит, что мы рассматриваем всевозможные вертикальные сечения, и берём логарифм от самого большого из них:

$$H_{1|2}(A) = \max_{a \in \pi_2(A)} \lceil |A_a| \rceil$$

Аналогично можно определить “количество информации” и для большего количества координат. Нетрудно заметить, что

$$H(A) \leq H_1(A) + H_{1|2}(A)$$

(по существу, это утверждение о том, что размер множества не больше, чем произведение размера проекции на первую координату и максимального размера вертикального сечения).

Упражнение 1 Докажите, что для любого 3-мерного множества A выполняется неравенство

$$2H(A) \leq H_{12}(A) + H_{23}(A) + H_{13}(A)$$

¹С исправлениями А. Чулкина.

1.2 Детские задачи.

Задача 1. Сколько нужно задать вопросов, подразумевающих ответ *да* или *нет*, чтобы отгадать задуманное число из интервала $1..100$? Тот же вопрос для числа из интервала $1..N$ для произвольного N . Можно ли не задавать вопросы один за другим, а прислать список сразу со всеми вопросами?

Те же вопросы для отгадывания двух различных задуманных чисел из интервала $1..N$.

Задача 2. Сколько нужно задать вопросов, чтобы отгадать число из интервала $1..100$, если отвечающему разрешается солгать в ответ на один из вопросов.

Упражнение 2 Найдите минимальное число вопросов, необходимое в этой задаче.

Задача 3. Имеется N шаров, из которых один радиоактивен. Есть прибор, который за один тест позволяет проверить, находится ли радиоактивный шар среди некоторого выбранного множества шаров.

(а) Сколько нужно провести тестов, чтобы гарантированно найти радиоактивный шар.

(б) Тот же вопрос, если среди N шаров есть два радиоактивных (требуется найти оба).

Упражнение 3 Найдите точный ответ на вопрос (б).

Задача 4. Имеется 25 монет одинаковых на вид. Одна из монет фальшивая. Все настоящие монеты имеют одинаковый вес, фальшивая легче. Есть чашечные весы без гирь. Сколько нужно произвести взвешиваний, чтобы найти фальшивую монету? Тот же вопрос для N монет (среди которых есть одна фальшивая).

Задача 5. Имеется 12 монет, одна из них фальшивая. Все настоящие монеты имеют одинаковый вес, а фальшивая легче или тяжелее. Есть чашечные весы без гирь. Сколько нужно произвести взвешиваний, чтобы найти фальшивую монету и определить легче она или тяжелее?

Упражнение 4 Решите аналогичную задачу для 13 монет.

Задача 6. Имеется 15 монет, одна из них фальшивая. Все настоящие монеты имеют одинаковый вес, а фальшивая легче или тяжелее. Есть чашечные весы без гирь. Сколько нужно произвести взвешиваний, чтобы найти фальшивую монету (не требуется определять, легче она или тяжелее) ?

Упражнение 5 *Решите аналогичную задачу для 14 монет.*

Задача 7. Имеется 100 серебряных монет, упорядоченных по весу, а также 101 золотая монета, упорядоченная по весу. Требуется найти 101-ую по весу монету среди всех этих монет.

Упражнение 6 *Найдите точный ответ на вопрос задачи.*

Задача 7. Имеется N шаров разного веса и чашечные весы, которые позволяют сравнить веса любых двух шаров. Сколько нужно взвешиваний, чтобы гарантированно упорядочить шары по весу?

Упражнение 7 *Найдите ответ на вопрос задачи для $N = 5$.*

Задача 9. 1000 мудрецов подвергаются следующему испытанию. Им надевают на голову колпаки белого и чёрного цвета и выстраивают в колонну. Каждый мудрец видит цвета колпаков у всех впереди стоящих, но не видит ни своего собственного, ни колпаков мудрецов позади него. У мудрецов по очереди спрашивают про цвет его колпака, начиная с самого первого (с того, кто видит все колпаки кроме своего собственного). При этом ответы каждого мудреца слышны всем остальным.

Мудрецам требуется договориться о совместной стратегии, которая позволит минимизировать число неправильных ответов (в самом неблагоприятном для них случае). Сколькими неправильными ответами можно обойтись?

Упражнение 8 *Решите ту же задачу, если один из мудрецов может ошибиться и отступить от заранее оговорённой стратегии.*

Упражнение 9 *Пусть колонна из мудрецов бесконечна. Каждый из них видит цвета колпаков всех мудрецов, стоящих впереди него (т. е., i -ый мудрец знает цвета всех колпаков кроме первых i штук). Существует ли стратегия, которая позволяет обойтись конечным числом неправильных ответов?*

2 Лекция 2, 1 марта: энтропия Шеннона

Определение энтропии Шеннона для распределения вероятностей на конечном множестве. Если случайная функция α принимает некоторые n значений с вероятностями p_1, \dots, p_n ($\sum p_i = 1$), то ей сопоставляется величина $H(\alpha) = -\sum p_i \log p_i$.

Простейшие свойства:

- $H(\alpha) \geq 0$, причём равенство достигается если и только если величина детерминированная (одна из вероятностей равна единице, остальные нулю).
- $H(\alpha) \leq \log n$, причём равенство достигается если и только если распределение равномерное.

Для пары совместно определённых случайных величин α, β мы имеем энтропии $H(\alpha)$, $H(\beta)$, $H(\alpha, \beta)$. Кроме того, при каждом фиксированном значении b величины β мы имеем условное распределение вероятностей на значениях α . Обозначим его энтропию $H(\alpha|\beta = b)$. Определим относительную энтропию:

$$H(\alpha|\beta) = \sum_b H(\alpha|\beta = b) \cdot \text{Prob}[\beta = b]$$

Основные свойства:

- $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$, причём равенство достигается тогда и только тогда, когда α и β независимы (в обычном смысле теории вероятностей).
- $H(\alpha, \beta) = H(\alpha|\beta) + H(\beta)$
- $H(\alpha|\beta) \geq 0$, равенство достигается если и только если α есть детерминированная функция β .
- $H(\alpha|\beta) \leq H(\alpha)$, причём равенство достигается только при независимости α и β

Определим *информацию*, в α о величине β как разницу между простой и относительной энтропиями:

$$I(\alpha : \beta) = H(\beta) - H(\beta|\alpha)$$

Основные свойства:

- $I(\alpha : \beta) = I(\beta : \alpha) = H(\alpha) + H(\beta) - H(\alpha, \beta)$
- $I(\alpha : \beta) \leq H(\alpha)$, $I(\alpha : \beta) \leq H(\beta)$
- $I(\alpha : \beta) \geq 0$, равенство достигается если и только если величины α и β независимы.

- $I(\alpha : \alpha) = H(\alpha)$
- $I(\alpha : \beta) = H(\alpha)$ если и только если α есть детерминированная функция β

Аналогично определим относительную взаимную информацию $I(\alpha : \beta|\gamma)$.

Упражнение 10 Докажите, что $I(\alpha : \beta|\gamma) \geq 0$ и

$$H(\alpha, \gamma) + H(\beta, \gamma) \geq H(\alpha, \beta, \gamma) + H(\gamma).$$

Упражнение 11 Докажите, что $2H(\alpha, \beta, \gamma) \leq H(\alpha, \beta) + H(\alpha, \gamma) + H(\beta, \gamma|\alpha)$.

Упражнение 12 (а) Пусть α, β, γ образуют цепь Маркова (относительные распределения вероятностей γ при условии $\alpha = a$ и $\beta = b$ такое же, как и при условии $\beta = b$). Докажите, что $I(\alpha : \gamma) \leq I(\alpha : \beta)$

(б) Пусть четвёрка случайных величин $\alpha, \beta, \gamma, \delta$ образуют цепь Маркова. Докажите, что $I(\alpha : \delta) \leq I(\beta : \gamma)$.

3 Лекция 3, 15 марта: энтропии пар и троек случайных функций

Утверждение 1

$$H(\alpha, \beta, \gamma) + H(\gamma) \leq H(\alpha, \gamma) + H(\beta, \gamma)$$

Доказательство: применяем неравенство Йенсена.

Пусть дано совместное распределение тройки случайных величин α, β, γ . При каждом фиксированном значении $\gamma = c$ мы имеем распределение условных вероятностей на значениях α и β . Для каждого такого относительного распределения определена взаимная информация пары величин, которую мы будем обозначать $I(\alpha : \beta|\gamma = c)$. Далее, определим *информацию, в α о величине β относительно величины γ* как среднее взаимных информационных α и β при всех возможных значениях γ :

$$I(\alpha : \beta|\gamma) = \sum_c H(\alpha : \beta|\gamma = c) \cdot Prob[\gamma = c]$$

Основные свойства:

- $I(\alpha : \beta|\gamma) \geq 0$
- $I(\alpha : \beta|\gamma) = H(\alpha, \gamma) + H(\beta, \gamma) - H(\alpha, \beta, \gamma) - H(\gamma)$

Утверждение 2 Для любого вещественного $r \geq 0$ найдётся случайная величина α такая, что $H(\alpha) = r$.

Утверждение 3 (а) Для любых вещественных $r_1, r_2, r_{12} \geq 0$ найдутся случайные величины α_1, α_2 такие, что $H(\alpha_1|\alpha_2) = r_1$, $H(\alpha_2|\alpha_1) = r_2$, $I(\alpha_1 : \alpha_2) = r_{12}$.

(б) Если r_1, r_2, r_{12} удовлетворяют неравенствам

$$\begin{aligned} 0 &\leq r_1 &\leq r_{12}, \\ 0 &\leq r_2 &\leq r_{12}, \\ r_1 + r_2 &\geq r_{12}, \end{aligned}$$

то найдутся случайные величины α_1, α_2 такие, что $H(\alpha_1) = r_1$, $H(\alpha_2) = r_2$, $H(\alpha_1, \alpha_2) = r_{12}$.

Для тройки случайных величин $\alpha_1, \alpha_2, \alpha_3$ все энтропии можно задать набором из $2^3 - 1 = 7$ параметров. Достаточно указать величины

$$H(\alpha_1), H(\alpha_2), H(\alpha_3), H(\alpha_1, \alpha_2), H(\alpha_1, \alpha_3), H(\alpha_2, \alpha_3), H(\alpha_1, \alpha_2, \alpha_3)$$

Часто бывает удобна другая система координат:

$$H(\alpha_1|\alpha_2, \alpha_3), H(\alpha_2|\alpha_1, \alpha_3), H(\alpha_3|\alpha_1, \alpha_2), I(\alpha_1 : \alpha_2|\alpha_3), I(\alpha_1 : \alpha_3|\alpha_2), I(\alpha_2 : \alpha_3|\alpha_1), I(\alpha_1 : \alpha_2 : \alpha_3),$$

где взаимная информация тройки определяется следующим образом:

$$I(\alpha_1 : \alpha_2 : \alpha_3) = I(\alpha_1 : \alpha_2) - I(\alpha_1 : \alpha_2|\alpha_3)$$

Заметим, что величина $I(\alpha_1 : \alpha_2 : \alpha_3)$ симметрична по всем трём параметрам. Обратим также внимание, что для некоторых распределений вероятностей $I(\alpha_1 : \alpha_2 : \alpha_3) < 0$.

Для энтропий тройки случайных величин выполняются следующие “базисные” ограничения:

$$\begin{aligned} H(\alpha_i|\alpha_j, \alpha_k) &\geq 0, \\ I(\alpha_i : \alpha_j) &\geq 0, \\ I(\alpha_i : \alpha_j|\alpha_k) &\geq 0, \end{aligned}$$

Упражнение 13 Докажите, что всякое неравенство для энтропий тройки случайных величин есть комбинация “базисных” неравенств с положительными коэффициентами.

Другими словами, если набор из семи чисел удовлетворяет указанным базисным неравенствам, то для сколь угодно малого $\varepsilon > 0$ можно найти распределение вероятностей, соответствующие энтропии которого ε -близки к заданным семи числам. (В отличие от двумерного случая, мы не требуем, чтобы полученные энтропии в точности равнялись заданному набору чисел.)

Пример комбинации базисных неравенств: $H(\gamma) \leq H(\gamma|\alpha) + H(\gamma|\beta) + I(\alpha : \beta)$. Это неравенство показывает, что если γ является детерминированной функцией α и детерминированной функцией β , то энтропия γ не превосходит $I(\alpha : \beta)$.

Базисные неравенства задают выпуклый конус в 7-мерном вещественном пространстве. Для того, чтобы набор из семи чисел задавал энтропии какой-то тройки случайных величин, этот набор должен лежать внутри указанного конуса. Это условие является необходимым, но не является достаточным.

Упражнение 14 Если $H(\alpha_1) = H(\alpha_2) = H(\alpha_3) = n$, $I(\alpha_1 : \alpha_2) = I(\alpha_2 : \alpha_3) = I(\alpha_1 : \alpha_3) = 0$, $H(\alpha_1, \alpha_2, \alpha_3) = 2n$, то величина n есть двоичный логарифм некоторого целого числа N .

Определение 1 Набор слов $c_1, \dots, c_n \in \{0, 1\}^*$ называется однозначно декодируемым кодом, если никакое слово $x \in \{0, 1\}^*$ нельзя двумя разными способами представить в виде конкатенации слов c_i .

Примерами однозначно декодируемого кода являются (бес)префиксные коды.

Теорема 1 (неравенство Крафта) Если слова $c_1, \dots, c_n \in \{0, 1\}^*$ образуют однозначно декодируемый код, то $\sum 2^{-|c_i|} \leq 1$.

Теорема 2 Если для набора слов $c_1, \dots, c_n \in \{0, 1\}^*$ выполнено неравенство $\sum 2^{-|c_i|} \leq 1$, то существует префиксный код $d_1, \dots, d_n \in \{0, 1\}^*$ с такими же длинами кодовых слов (т.е. $|c_i| = |d_i|$ для $i = 1, \dots, n$).

Таким образом, всякий однозначно декодируемый код можно переделать в код префиксный, не меняя длин кодовых слов.

Упражнение 15 Сформулируйте и докажите неравенство Крафта для алфавита из $k > 2$ букв.

4 Лекция 4, 22 марта: энтропия Шеннона и экономные коды

Теорема 3 (а) Для любого распределения вероятностей p_1, \dots, p_n и любого однозначно декодируемого кода c_1, \dots, c_n

$$\sum p_i |c_i| \geq \sum p_i \log \frac{1}{p_i}$$

(б) Для любого распределения вероятностей p_1, \dots, p_n найдётся такой префиксный код c_1, \dots, c_n , что

$$\sum p_i |c_i| < \sum p_i \log \frac{1}{p_i} + 1$$

Конструкции кодов: код Шеннона-Фано, код Хаффмана (и его оптимальность), арифметическое кодирование.

5 Лекция 5, 5 апреля

Равномерный блочный код для канала без шума.

Теорема 4 (Шеннона о блочном кодировании источника) Пусть случайная величина α имеет энтропию h .

(1) Для всякого $L > h$ существуют функции кодирования и декодирования

$$C_k : A^k \rightarrow \{0, 1\}^{Lk}$$

и

$$D_k : \{0, 1\}^{Lk} \rightarrow A^{Lk}$$

такие, что вероятность ошибки

$$\varepsilon_k = \text{Prob}[D_k(C_k(a_{i_1} \dots a_{i_k})) \neq (a_{i_1} \dots a_{i_k})]$$

(где буквы a_{i_s} для каждой позиции $s = 1 \dots k$ выбираются независимо друг от друга, по распределению α) стремится к нулю при $k \rightarrow \infty$.

(2) Для всякого $L < h$ и для любой последовательности функций кодирования и декодирования

$$C_k : A^k \rightarrow \{0, 1\}^{Lk}$$

и

$$D_k : \{0, 1\}^{Lk} \rightarrow A^{Lk}$$

вероятность ошибки

$$\varepsilon_k = \text{Prob}[D_k(C_k(a_{i_1} \dots a_{i_k})) \neq (a_{i_1} \dots a_{i_k})]$$

(где буквы a_{i_s} для каждой позиции $s = 1 \dots k$ выбираются независимо друг от друга, по распределению α) стремится к единице при $k \rightarrow \infty$.

Комбинаторная модель шума. Расстояние кода и число исправляемых ошибок. Скорость кода: отношение логарифма числа кодовых слов к длине кодового слова.

Оценки Хэмминга и Гилберта для соотношения скорости кода и кодового расстояния.

6 Лекция 6, 12 апреля

Коды Хэмминга (совершенные коды, исправляющие одну ошибку).

Упражнение 16 Вычислите минимальное число вопросов, позволяющее угадать целое число от 1 до 100, если один из ответов может быть ошибочным.

Код Рида–Соломона. Расстояние кода Рида–Соломона и оценка Синглтона. Полиномиальный алгоритм декодирования для кода Рида–Соломона.

Упражнение 17 Пусть двоичный код с длиной кодового слова n имеет расстояние $d \geq cn$ для некоторой константы $c > 1/2$. Покажите, что число кодовых слов ограничено некоторой величиной $F(c)$ (которая не зависит от n).

Упражнение 18 Пусть двоичный код с длиной кодового слова n имеет расстояние $d > n/2$. Покажите, что число кодовых слов не превосходит $n + 1$.

Упражнение 19 Пусть двоичный код с длиной кодового слова n имеет расстояние $d \geq n/2$. Покажите, что число кодовых слов не превосходит $2n$.

Вывод из этих трёх упражнений: асимптотически хороший код может исправлять лишь менее 25% ошибок.

7 Лекция 7, 19 апреля

Построение асимптотически хорошего кода, исправляющего 1% ошибок, имеющего скорость не менее $1/4$ и допускающего кодирование и декодирование за полиномиальное время (конкатенация кода Рида–Соломона и неконструктивного кода с границы Варшамова–Гилберта).

Упражнение 20 Докажите, что код Адамара имеет кодовое расстояние, равное 50% длины кодового слова.

Упражнение 21 Найдите расстояние и объём кода Рида–Маллера (Reed–Muller) порядка s и степени r .

Вероятностная модель канала с шумом. Дискретный канал без памяти (канал задаётся входным алфавитом A , выходным алфавитом B и набором условных вероятностей p_{ij} для $i = 1, \dots, |A|$ и $j = 1, \dots, |B|$). Определение пропускной способности дискретного канала без памяти.

Пример: вычисление пропускной способности для двоичного симметричного канала.

Упражнение 22 Рассмотрим двоичный канал без памяти, который передаёт единицу без искажения, а нули с вероятностью p меняет на единицы. Вычислите пропускную способность этого канала.

Теорема 5 (Шеннона о кодировании для дискретного канала с шумом)

Пусть пропускная способность канала без памяти (с входным алфавитом A и выходным алфавитом B) равна R .

(1) Для всякого $L < R$ существуют функции кодирования и декодирования

$$C_k : \{0, 1\}^k \rightarrow A^{k/L}$$

и

$$D_k : B^{k/L} \rightarrow \{0, 1\}^k$$

такие, что вероятность ошибки ε_k при кодировании блоков из k битов

$$\{0, 1\}^k \xrightarrow{\text{кодирование } C_k} A^{k/L} \xrightarrow{\text{искажение в канале}} B^{k/L} \xrightarrow{\text{декодирование } D_k} \{0, 1\}^k$$

стремится к нулю при $k \rightarrow \infty$.

(2) Если $L > R$, то для любых функции кодирования и декодирования

$$C_k : \{0, 1\}^k \rightarrow A^{k/L}$$

и

$$D_k : B^{k/L} \rightarrow \{0, 1\}^k$$

вероятность ошибки ε_k не стремится к нулю при $k \rightarrow \infty$.

(2') При тех же предположениях, что и в (2), вероятность ошибки ε_k стремится к единице.

Доказательство (1) для двоичного симметричного канала (метод случайного кодирования). Доказательство (2) для произвольного канала.

8 Лекция 8, 26 апреля: коммуникационная сложность

Определение детерминированного коммуникационного протокола для вычисления функции $f : X \times Y \rightarrow Z$ (Алиса хранит первый аргумент функции, принадлежащий X , а Боб – второй аргумент, принадлежащий Y). Коммуникационная сложность функции $CC(f)$. Тривиальная верхняя оценка сложности $\log |X| + \log |Z|$ и тривиальная нижняя оценка $\log |Y|$.

Примеры протоколов, вычисляющих предикат равенства слов длины n , медиану двух множеств из $\{1, \dots, n\}$.

Упражнение 23 Придумайте протокол сложности $O(\log n)$, вычисляющий медиану двух множеств из $\{1, \dots, n\}$.

Прямоугольное множество для заданной функции $f : X \times Y \rightarrow Z$. Каждому листу протокола соответствует прямоугольное множество. Трудное множество.

Нижняя оценка для коммуникационной сложности предиката EQ , для предиката дизъюнктивности пары множеств из $\{1, \dots, n\}$.

Упражнение 24 Докажите, что коммуникационная сложность предиката GT (на парах n -битных чисел) равна $n + 1$.

Вероятностные коммуникационные протоколы, вероятностная коммуникационная сложность CC_ε .

Вероятностный протокол сложности $O(\log n)$ для предиката EQ_n .

Теорема 6 $CC_\varepsilon(f) = \Omega(\log CC(f))$ для любого предиката f .

Следствие: Вероятностная коммуникационная сложность предиката равенства равна $\Omega(\log n)$.

Вероятностные коммуникационные протоколы с нулевой вероятностью ошибки, коммуникационная сложность $CC_0(f)$. Оценка $CC_0(EQ_n) \geq n + 1$.

Теорема 7 Пусть $f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$ некоторый предикат, и дана одноленточная машина Тьюринга M , которая допускает за время $T(n)$ все слова вида

$$\{x0^n y \mid |x| = |y| = n, f(x, y) = 1\}$$

и отвергает за время $T(n)$ все слова вида

$$\{x0^n y \mid |x| = |y| = n, f(x, y) = 0\}.$$

Тогда $CC_0(f) = O\left(\frac{T(n)}{n}\right) + O(\log n)$

Следствие: Язык палиндромов распознаётся одноленточной машиной Тьюринга за время $\Omega(n^2)$.

Теорема 8 Пусть $f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$ некоторый предикат, и дана многоленточная машина Тьюринга M , которая допускает за время $T(n)$ и используя память $S(n)$ все слова вида

$$\{x0^n y \mid |x| = |y| = n, f(x, y) = 1\}$$

и отвергает за время $T(n)$ и используя память $S(n)$ все слова вида все слова вида

$$\{x0^n y \mid |x| = |y| = n, f(x, y) = 0\}.$$

Тогда $CC(f) = O\left(\frac{S(n)T(n)}{n}\right)$

Следствие: Если многоленточная машина распознаёт язык палиндромовна зоне $S(n)$ за время $T(n)$, то $T(n) \cdot S(n) = \Omega(n^2)$.

9 Лекция 9, 3 мая: колмогоровская сложность

Определение 2 Пусть $U : \{0,1\}^* \rightarrow \{0,1\}^*$ есть (частичная) вычислимая функция. Определим $K_U(x) = \min\{|p| : U(p) = x\}$. (Минимум пустого множества считается равным бесконечности.)

Теорема 9 Существует такой способ описания (частичная вычислимая функция) U , что для любой другой V и для всех слов x $K_U(x) \leq K_V(x) + O(1)$

Фиксируем какой-либо оптимальный способ описания U и обозначаем соответствующую сложность $K(x)$. Называем эту величину *колмогоровской сложностью* слова x .

Простейшие свойства колмогоровской сложности:

- $K(x) \leq |x| + O(1)$;
- $K(f(x)) \leq K(x) + O(1)$ для всякой вычислимой f ;
- для всякого n существует слово длины n и сложности не менее n ;
- существует такая константа C , что для всякого n не менее, чем для 99% слов длины n выполнено $n - c \leq K(x) \leq n + c$.

Определение относительной колмогоровской сложности.

Теорема 10 (Колмогорова–Левина)

$$K(x, y) = K(x) + K(y|x) + O(\log K(x, y))$$

Замечание: логарифмический член в теореме Колмогорова–Левина устранить нельзя.

Теорема 11 (Колмогорова–Левина) Пусть слово x длины n состоит из pn единиц и $(1-p)n$ нулей. Тогда

$$K(x) \leq \left(p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} \right) n + O(\log n)$$

Следствие: Существует последовательность префиксных блоковых кодов $C_k : A^k \rightarrow \{0,1\}^*$ таких, что для любого распределения вероятностей ξ на A , для последовательности независимых случайных величин ξ_i (каждая из которых распределена как ξ)

$$\lim_{k \rightarrow \infty} \frac{E|C_k(\xi_1 \dots \xi_k)|}{k} = H(\xi)$$

Теорема 12 Не существует алгоритма, который по заданному n находил бы слово с колмогоровской сложностью не менее n .

Следствие: Колмогоровская сложность $K(x)$ не является вычислимой функцией.

Пример применения колмогоровской сложности: для любых слов выполнено неравенство

$$2K(x, y, z) \leq K(x, y) + K(x, z) + K(y, z) + O(\log K(x, y, z))$$

Следствие: для любого компактного множества $A \subset \mathbb{R}^3$ выполнено следующее неравенство

$$\text{vol}(A)^2 \leq S(A_{12}) \cdot S(A_{13}) \cdot S(A_{23})$$

(где $S(A_{ij})$ обозначает площадь проекции A на координатную плоскость $Ox_i x_j$).

Упражнение 25 Докажите, что $K(x, y, z) + K(z) \leq K(x, z) + K(y, z) + O(\log K(x, y, z))$

10 Лекция 10–11, 10 мая: колмогоровская сложность и алгоритмическая случайность

Теорема 13 (Пример 1 применения колмогоровской сложности) Для того чтобы скопировать слово длины n одноленточной машине Тьюринга (с одной головкой) требуется время $\Omega(n^2)$.

Теорема 14 (Пример 2 применения колмогоровской сложности) Для любого k существует язык L_k , распознаваемый конечным автоматом с k читающими головками, но не распознаваемый автоматом с меньшим числом головок (все головки движется вдоль входного слова слева направо).

Случайность по Мартин-Лёфу: Определение эффективно-нулевого множества. Примеры эффективно нулевых множеств: индивидуальная вычислимая последовательность; последовательности, у которых во всех позициях с нечётными номерами стоят нули; последовательности, для которых нарушен закон больших чисел.

Теорема 15 Существует максимальное эффективно-нулевое множество.

Теорема 16 Для всякой бесконечной двоичной последовательности $\omega_0 \omega_1 \omega_2 \dots$ найдутся сколь угодно большие номера n такие, что

$$K(\omega_0 \dots \omega_{n-1}) \leq n - \log n + O(1)$$

Определение префиксной колмогоровской сложности. Простейшие свойства префиксной сложности.

Лемма 1 Если $P(x)$ перечислимая снизу полумера, то $KP(x) \leq \log \frac{1}{P(x)} + O(1)$.

Теорема 17 Бесконечная двоичная последовательность $\omega_0\omega_1\omega_2\dots$ случайна по Мартин-Лёфу тогда и только тогда, когда для всех n

$$KP(\omega_0\dots\omega_{n-1}) \geq n - O(1)$$

Теорема 18 (Закон больших чисел в форме Харди–Литтлвуда) Для почти всех последовательностей $\omega_0\omega_1\omega_2\dots$

$$[\text{доля единиц среди первых } n \text{ битов последовательности}] = \frac{1}{2} + O\left(\sqrt{\frac{\ln n}{n}}\right)$$

Упражнение 26 Пусть вычислимая функция $U(\cdot)$ есть оптимальный способ описания (для простой колмогоровской сложности). Верно ли, что $U(U(\cdot))$ также является оптимальным способом описания?

Упражнение 27 Существует ли такой оптимальный способ описания, что колмогоровская сложность всех слов чётна? является некоторой степенью двойки?

Упражнение 28 Докажите, что $K(x, K(x)) = K(x) + O(1)$

Список литературы

- [1] А.М. Яглом, И.М. Яглом. Вероятность и информация, 1973.
- [2] Р. Галлагер. Теория информации и надёжная связь, 1974.
- [3] Т.М. Cover, J.A. Thomas. Elements of information Theory, 2006.
- [4] И. Чисар, Я. Кернер. Теория информации, 1985.
- [5] R.W. Yeung. A First Course in Information Theory, 2002.
- [6] В.М. Сидельников. Теория кодирования, 2008.
- [7] Ф.Дж.А. Мак-Вильямс, Н.Дж.А. Слоэн. Теория кодов, исправляющих ошибки, 1979.
- [8] E. Nisan, N. Kushilevitz. Communication complexity, 1997.
- [9] M. Li, and P. Vitanyi. An Introduction to Kolmogorov Complexity and Its Applications. 2008.
- [10] В. А. Успенский, Н. К. Верещагин, А. Шень. Колмогоровская сложность (черновик книги).