ФИВТ МФТИ. Краткий конспект лекций курса
Введение в теорию информации [1]**.**
(А.Е. Ромащенко, весна 2014).

1 Определение информации по Хартли

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

$$\mathcal{X}(A) = \log|A|,\ \mathcal{X}_1(A) = \log|\pi_1 A|,\ \mathcal{X}_2(A) = \log|\pi_2 A|$$

(здесь $\pi_i A$ обозначает проекцию множества A на i-ую координату). Заметим, что $\mathcal{X}(A) \le \mathcal{X}_1(A) + \mathcal{X}_2(A)$, 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 смысл независимы).

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

$$\mathcal{X}_{2|1}(A) = \log\left(\max_{a \in \pi_1(A)} |\{b\ :\ (a,b) \in A\}|\right)$$

Аналогично можно определить "количество информации" и для бо́льшего количества координат. Нетрудно заметить, что

$$\mathcal{X}(A) \le \mathcal{X}_1(A) + \mathcal{X}_{2|1}(A)$$

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

Домашнее задание 1.1. *Докажите, что для любого 3-мерного множества A выполняется неравенство $2\mathcal{X}(A) \le \mathcal{X}_{12}(A) + \mathcal{X}_{23}(A) + \mathcal{X}_{13}(A)$.*

1.1 Информация по Хартли в детских задачах.

Пример 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 (адаптивная стратегия), а прислать список сразу со всеми вопросами (неадаптивная стратегия)?

[1] Последние исправления в текст внесены 23.01.2016.

Те же вопросы для отгадывания задуманной пары различных чисел из интервала $1..N$.

В решении этой Задачи 1 доказательство нижней оценки на минимальное число вопросов можно изложить тремя альтернативными способами: (а) подсчет числа листьев в дереве–стратегии фиксированной высоты, (б) рассуждение со "злонамеренным противником" (adversary argument), и (в) "информационная" оценка на языке информации Хартли.

Домашнее задание 1.2. *Имеется n камней попарно разного веса и чашечные весы, которые позволяют сравнить веса любых двух камней.*

(а) Сколько нужно операций взвешивания, чтобы найти самый тяжелый камень?

(б) Сколько нужно операций взвешивания, чтобы найти самый тяжелый камень и второй по весу камень?

Пример 2. 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).

Домашнее задание 1.3. *Имеется* 12 *монет, одна из них фальшивая. Все настоящие монеты имеют одинаковый вес, а фальшивая легче или тяжелее. Есть чашечные весы без гирь, с помощью которых можно сравнить веса двух любых групп монет. Сколько нужно произвести взвешиваний, чтобы найти фальшивую монету и определить, легче она или тяжелее?*

Домашнее задание 1.4. *Решите аналогичную задачу для* 13 *монет.*

Пример 4. 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)?

Домашнее задание 1.5. *Решите аналогичную задачу для* 14 *монет.*

Домашнее задание 1.6. *Сколько нужно взвешиваний, чтобы упорядочить N камне по весу? Найдите точный ответ на этот вопрос для $N = 2, 3, 4, 5$.*

2 Энтропия Шеннона: определение и простейшие свойства энтропии дискретной случайной величины

Определение 1. *Шенноновской энтропией случайной функции α, которая принимает k значений с вероятностям $p_1, \ldots, p_k$ ($\sum_{i=1}^{k} p_i = 1$), называется число $H(\alpha) = -\sum p_i \log p_i$.*

Мы будем применять это определение в том числе и для распределений, в которых некоторые p_i равны нулю. Как обычно, мы по непрерывности доопределяем функцию $x \log x$ в нуле и полагаем $0 \cdot \log \frac{1}{0} = 0$.

Простейшие свойства энтропии Шеннона:

- $H(\alpha) \geq 0$, причём равенство достигается, если и только если величина α является вырожденной (вероятность p_i одного из значений равна единице, а вероятности всех остальных значений равны нулю).
- для случайной величины с k значениями $H(\alpha) \leq \log k$, причём равенство достигается, если и только если распределение равномерное, т.е., $p_1 = \ldots = p_k = \frac{1}{k}$ (доказательство: воспользуемся вогнутостью логарфма и неравенством Йенсена).

Для пары совместно определённых случайных величин α, β мы имеем энтропии $H(\alpha)$, $H(\beta)$, $H(\alpha, \beta)$.

- $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$, причём равенство достигается тогда и только тогда, когда α и β независимы (в обычном смысле теории вероятностей).

Также для пары совместно определённых случайных величин α, β при каждом фиксированном значении b величины β мы имеем некоторое условное распределение вероятностей на значениях α. Обозначим энтропию этого условного распределения $H(\alpha \mid \beta = b)$. *Условная* (или *относительная*) энтропия Шеннона величины α относительно величины β определяется как усреднение энтропий $H(\alpha \mid \beta = b)$ по всем значениям β:

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

Простейшие свойства относительной энтропии:

- $H(\alpha, \beta) = H(\alpha \mid \beta) + H(\beta)$
- $H(\alpha \mid \beta) \geq 0$, причём равенство достигается тогда и только тогда, когда α есть детерминированная функция β (т.е., по значению β с веротяностью один можно однозначно восстановить значение α).

- $H(\alpha \mid \beta) \leq H(\alpha)$, причём равенство достигается только при независимости α и β

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

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

Основные свойства взаимной информации:

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

Аналогично определим *относительную взаимную информацию* (взаимную информацию между α и β при известном значении γ). Мы рассмотрим три варианта этого определения.

Первый вариант определения:

$$I(\alpha : \beta \mid \gamma) := \sum_c I(\alpha : \beta \mid \gamma = c) \cdot \mathrm{Prob}[\gamma = c]$$

(здесь $I(\alpha : \beta \mid \gamma = c)$ обозначает взаимную информацию между α и β в условном распределении информации на парах (α, β) при зафиксированном значения $\gamma = c$.

Второй вариант определения:

$$I(\alpha : \beta \mid \gamma) := H(\beta \mid \gamma) - H(\beta \mid \alpha, \gamma).$$

Третий вариант определения:

$$I(\alpha : \beta \mid \gamma) := H(\alpha \mid \gamma) + H(\beta \mid \gamma) - H(\alpha, \beta \mid \gamma).$$

Домашнее задание 2.1. *Докажите, что три определения относительной взаимной информации эквивалентны.*

Домашнее задание 2.2. *Докажите, что*
(а) если α, β, γ образуют цепь Маркова (относительные распределения вероятностей γ при условии $\alpha = a$ и $\beta = b$ такое же, как и при условии $\beta = b$), то $I(\alpha : \gamma) \leq I(\alpha : \beta)$ и $I(\alpha : \gamma) \leq I(\beta : \gamma)$;
(б) если четвёрка случайных величин $\alpha, \beta, \gamma, \delta$ образуют цепь Маркова, то $I(\alpha : \delta) \leq I(\beta : \gamma)$.

2.1 Использования энтропии в «детских» задачах

Пример 1: Энтропийная эвристика (жадный «энтропийный» алгоритм) в задаче о поиске одной фальшивой монеты и её относительного веса из 13 (см. Домашнее задание 1.4).

Пример 2: Энтропийное доказательство нижней оценки для числа взвешиваний при поиске одной фальшивой монеты из 14 без обязательно определения относительного веса (см. Домашнее задание 1.5).

Пример 3: Энтропийная эвристика (жадный «энтропийный» алгоритм) в задаче о сортировки 5 камней по весу.

3 Относительная взаимная информация: определения и простейшие свойства

Напомним, что для распределение тройки случайных величин α, β, γ мы тремя эквивалентными способами определили относительную взаимную информацию $I(\alpha : \beta \mid \gamma)$. Из этих определений немедленно вытекают следующие свойства:

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

Замечание. Существуют такие распределения вероятностей (α, β, γ), что $I(\alpha : \beta) = 0$, но $I(\alpha : \beta \mid \gamma) > 0$. И наоборот, для некоторых распределений $I(\alpha : \beta \mid \gamma) = 0$, но $I(\alpha : \beta) > 0$. (*Приведите примеры таких распределений!*)

3.1 Энтропийные профили распределений

Далее мы рассмотрим вопрос: какими могут быть *энтропийные профили* — значения энтропий набора совместно распределенных случайных величин. Начнем с самого простого вопроса: какой может быть энтропия одной случайной величины?

Утверждение 1. *Для любого вещественного $h \geq 0$ найдётся случайная величина α такая, что $H(\alpha) = h$.*

Для совместо распределнной пары (α_1, α_2) у нас имеется уже довольно много информационных величин: энтропии $H(\alpha_1), H(\alpha_2), H(\alpha_1, \alpha_2)$, относительные энтропии $H(\alpha_1 \mid \alpha_2)$ и $H(\alpha_2 \mid \alpha_1)$, а также взаимная информация $I(\alpha_1 : \alpha_2)$. Однако не все эти величны независимы. Например, зная значения $H(\alpha_1), H(\alpha_2), H(\alpha_1, \alpha_2)$, можно вычислить значения обеих относительных энтропий и взаимной информации. И наоборот, зная $H(\alpha_1 \mid \alpha_2)$, $H(\alpha_2 \mid \alpha_1)$ и $I(\alpha_1 : \alpha_2)$, мы можем восстановить $H(\alpha_1), H(\alpha_2), H(\alpha_1, \alpha_2)$.

Все энтропии принимают неотрицательные занчения. Кроме того, мы значем, что энтропия пары $H(\alpha_1, \alpha_2)$ не меньше каждой из энтропий $H(\alpha_1)$, $H(\alpha_2)$ и не больше суммы этих двух энтропий. Следующее утверждение показывает, что никаких других ограничений на значения энтропий пары случайных величин нет.

Утверждение 2. *(а) Если h_1, h_2, h_{12} удовлетворяют неравенствам*

$$\begin{array}{rcl} 0 & \leq & h_1, h_2, \\ h_1 & \leq & h_{12}, \\ h_2 & \leq & h_{12}, \\ h_{12} & \leq & h_1 + h_2, \end{array}$$

то найдутся случайные величины α_1, α_2 такие, что

$$H(\alpha_1) = h_1,\ H(\alpha_2) = h_2,\ H(\alpha_1, \alpha_2) = h_{12}.$$

(б) Для любых трёх неотрицательных вещественных чисел $h_{1|2}, h_{2|1}, h_{1:2}$ найдутся случайные величины α_1, α_2 такие, что

$$H(\alpha_1 \mid \alpha_2) = h_{1|2},\ H(\alpha_2 \mid \alpha_1) = h_{2|1},\ I(\alpha_1 : \alpha_2) = h_{1:2}.$$

Замечание 1: Пункты (а) и (б) этого утверждения эквивалентны. По существу, это одно и то же утверждение, записанное в двух разных системах координат.
Замечание 2: Данное утверждение говорит, что множество троек чисел, представляющих энтропии всевозможных распределений (α_1, α_2), является замкнутым и выпуклым конусом в $\mathbb{R}^3$. Этот конус ограничивается тремя плоскостями (является пересечением трёх полупространств в $\mathbb{R}^3$). Пункты (а) и (б) утверждения описывают этот конус в двух разных системах координат. Во второй системе координат (пункт (б)) данный конус выглядит совсем просто — это первый координатный октант, т.е., множество точек, все три координаты которых неотрицательны.

Для тройки совместно распределенных случайных величин $\alpha_1, \alpha_2, \alpha_3$ все энтропии можно задать набором из $2^3 - 1 = 7$ параметров. В самом деле, чтобы однозначно определить все энтропии (условные и безусловные) и взаимные информации (условные и безусловные), включающие $\alpha_1, \alpha_2, \alpha_3$, достаточно указать значения

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

Иногда удобнее пользоваться другой система координат и описывать энтропии тройки $(\alpha_1, \alpha_2, \alpha_3)$ другим набором из 7 параметров:

$$\begin{array}{l} H(\alpha_1 \mid \alpha_2, \alpha_3), H(\alpha_2 \mid \alpha_1, \alpha_3), H(\alpha_3 \mid \alpha_1, \alpha_2), \\ I(\alpha_1 : \alpha_2 \mid \alpha_3), I(\alpha_1 : \alpha_3 \mid \alpha_2), I(\alpha_2 : \alpha_3 \mid \alpha_1), \\ I(\alpha_1 : \alpha_2 : \alpha_3), \end{array}$$

где *взаимная информация тройки* $I(\alpha_1 : \alpha_2 : \alpha_3)$ определяется следующим образом:

$$I(\alpha_1 : \alpha_2 : \alpha_3) = I(\alpha_1 : \alpha_2) - I(\alpha_1 : \alpha_2 \mid \alpha_3).$$

Заметим, что величина $I(\alpha_1 : \alpha_2 : \alpha_3)$ симметрична по своим трём аргументам. Взаимная информация тройки не имеет наглядного смысла. Ниже мы увидим, что для некоторых распределений вероятностей величина $I(\alpha_1 : \alpha_2 : \alpha_3)$ отрицательна.

Для энтропий троек случайных величин выполняются следующие *базисные* ограничения:

$$H(\alpha_i \mid \alpha_j, \alpha_k) \geq 0,$$
$$I(\alpha_i : \alpha_j \mid \alpha_k) \geq 0,$$
$$I(\alpha_i : \alpha_j) \geq 0$$

для всех наборов попарно различных i, j, k. Отметим, что последнее неравенство можно переписать как

$$I(\alpha_1 : \alpha_2 : \alpha_3) + I(\alpha_i : \alpha_j \mid \alpha_k) \geq 0.$$

Эти 9 базисных неравенств задают выпуклый замкнутый конус в 7-мерном вещественном пространстве. Чтобы набор из семи чисел представлял энтропии какой-то тройки случайных величин, *необходимо*, чтобы данная точка лежала в описанном конусе. Однако данное условие не является *достаточным.*

Домашнее задание 3.1. *Известно, что в некотором совместном распределении (a, b, c) значение случайной величины c однозначно определяется по значению a и однозначно определяется по значению b. Докажите, что $H(c) \leq I(a : b)$.*

4 Информационные неравенства и энтропии для распределений троек случайных величин

Информационные диаграммы для троек случайных величин. Доказательство информационных неравенств с помощью диаграмм. Перевод доказательств с языка диаграмм на язык формальных неравенств.

Домашнее задание 4.1. *Постройте пример распределения (α, β, γ), для которого*

$$H(\alpha \mid \beta, \gamma) = H(\beta \mid \alpha, \gamma) = H(\gamma \mid \alpha, \beta) = 0,$$
$$I(\alpha : \beta \mid \gamma) = I(\alpha : \gamma \mid \beta) = I(\beta : \gamma \mid \alpha) = 1,$$
$$I(\alpha : \beta : \gamma) = -1.$$

Домашнее задание 4.2. *Для некоторого совместного распределения (a, b, c) выполнены равенства*

$$I(a : b|c) = I(a : c|b) = I(b : c|a) = 0.$$

Докажите, что можно определить (на том же вероятностном пространстве) случайную величину w такую, что

$$H(w|a) = H(w|b) = H(w|c) = 0$$

и случайные величины a, b, c независимы относительно w.

Домашнее задание 4.3. *!!! Если $H(\alpha_1) = H(\alpha_2) = H(\alpha_3) = h, H(\alpha_1, \alpha_2) = H(\alpha_2, \alpha_3) = H(\alpha_1, \alpha_3) = 2h, H(\alpha_1, \alpha_2, \alpha_3) = 2h$, то величина h есть двоичный логарифм некоторого целого числа N.*

5 Оптимальное кодирование дискретного распределения.

Определение 2. *Набор слов $c_1, \ldots, c_n \in \{0,1\}^*$ называется однозначно декодируемым кодом, если никакое слово $x \in \{0,1\}^*$ нельзя двумя разными способами представить в виде конкатенации слов c_i.*

Стандартным примером однозначно декодируемого кода является класс (бес)префиксных кодов.

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

Теорема 2. *Если для набора чисел $l_1, \ldots, l_n \in N$ выполнено неравенство $\sum 2^{-|c_i|} \leq 1$, то существует префиксный код $c_1, \ldots, c_n \in \{0,1\}^*$ с такими длинами кодовых слов (т.е. $|c_i| = l_i$ для $i = 1, \ldots, n$).*

Из этих двух теорем следует, что всякий однозначно декодируемый код можно переделать в код префиксный, не меняя длин кодовых слов.

Теорема 3. *(а) Для любого распределения вероятностей $p_1, \ldots, p_n$ и любого однозначно декодируемого кода $c_1, \ldots, c_n$*

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

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

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

Доказательство: пользуемся вогнутостью логарифма и неравенством Крафта.

Обсуждение классических конструкций кодов: код Шеннона–Фано, код Хаффмана (и доказательство его оптимальности), арифметическое кодирование (с оценкой средней длины кодового слова).

Домашнее задание 5.1. *Докажите неравенство Крафта для однозначно декодируемых двоичных кодов.*

Домашнее задание 5.2. *Докажите оптимальность кода Хаффмана.*

Домашнее задание 5.3. *Приведите пример распределения вероятностей, для которого код Шеннона–Фано не является оптимальным.*

Домашнее задание 5.4. *Существует ли распределения вероятностей на 4 элементах, для которого код Шеннона–Фано не является оптимальным?*

Домашнее задание 5.5. *Приведите пример распределения вероятностей, для которого арифметический код не является оптимальным.*

Домашнее задание 5.6. *Энтропия некоторого распределения α равна $H(\alpha)$. Может ли средняя длина арифметического код для этого распределения быть больше $H(\alpha)+1$?*

6 Лекция 5, 19 марта: блоковое кодирование для канала без шума.

Теорема 4 (Шеннона о блоковом кодировании источника)**.** *Пусть случайная величина α распределена на конечном множестве $\{a_1,\dots,a_k\}$. Рассмотрим последовательность независимых одинаково распределенных копий этой случайной величины и кодирование блоков из n таких случайных величин.*

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

$$C_n : A^n \to \{0,1\}^{\lfloor L\cdot n\rfloor}$$

и

$$D_n : \{0,1\}^{\lfloor L\cdot n\rfloor} \to A^n$$

такие, что вероятность ошибки

$$\varepsilon_n = \mathrm{Prob}_{a_{i_1}\dots a_{i_n}}[D_n(C_n(a_{i_1}\dots a_{i_n})) \neq (a_{i_1}\dots a_{i_n})]$$

стремится к нулю при $n\to\infty$ (буквы a_{i_s} для каждой позиции $s = 1\dots n$ выбираются по распределению α, независимо для всех i).

(2) [*слабое обращение*] *Для всякого $L < h$ и для любой последовательности функции кодирования и декодирования*

$$C_n : A^n \to \{0,1\}^{\lceil L\cdot n\rceil}$$

и

$$D_n : \{0,1\}^{\lceil L\cdot n\rceil} \to A^n$$

вероятность ошибки

$$\varepsilon_n = \mathrm{Prob}_{a_{i_1}\dots a_{i_n}}[D_n(C_n(a_{i_1}\dots a_{i_n})) \neq (a_{i_1}\dots a_{i_n})]$$

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

(2') [*сильное обращение*] *Для всякого $L < h$ и для любой последовательности функции кодирования и декодирования*

$$C_n : A^n \to \{0,1\}^{\lceil L\cdot n\rceil}$$

и

$$D_n : \{0,1\}^{\lceil L\cdot n\rceil} \to A^n$$

вероятность ошибки

$$\varepsilon_n = \mathrm{Prob}_{a_{i_1}\dots a_{i_n}}[D_n(C_n(a_{i_1}\dots a_{i_n})) \neq (a_{i_1}\dots a_{i_n})]$$

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

Домашнее задание 6.1. *Докажите, что для каждого фиксированного алфавита $A = \{a_1, \dots, a_k\}$ существует последовательность блоковых кодов (C_n, D_n)*

$$C_n : A^n \to \{0,1\}^* \text{ и } D_n : \{0,1\}^* \to A^n$$

таких, что для любого распределения α на алфавите A

$$\lim_{n\to\infty} \mathrm{Prob}_{a_{i_1}\dots a_{i_n}}[D_n(C_n(a_{i_1}\dots a_{i_n})) \neq (a_{i_1}\dots a_{i_n})] = 0$$

(вероятность ошибки стремится к нулю) и

$$\lim_{n\to\infty} \frac{|E(C_n(a_{i_1}\dots a_{i_n}))|}{n} = H(\alpha)$$

(средняя длина кодового слова составляет $H(\alpha) + o(1)$ битов не один символ), где буквы a_{i_s} для каждой позиции $s = 1\dots n$ выбираются независимо друг от друга по распределению α.

Замечание: Верно и немного более сильное утверждение: множество всех кодовых слов (образ функции C_n) можно сделать префиксным, а вероятность ошибки декодирования можно сделать равной нулю.

7 Кодирование Вульфа–Слепяна.

Теорема 5 (О кодировании и декодировании с помощником)**.** *Пусть пара случайных величин (α, β) совместно распределена на некотором конечном множестве $A \times B$. Рассмотрим последовательность из n независимых одинаково распределенных копий этой пары случайных величин (α_i, β_i).*

(1) [*кодирование и декодирование с помощником*] *Пусть значения* β_i *известны отправителю и получателю, а значения* α_i *только отправителю.*

Для пары функций

$$C_n : (A \times B)^n \to \{0,1\}^{Ln}$$

и

$$D_n : \{0,1\}^k \times B^{Ln} \to A^n$$

вероятность ошибки определяется как

$$\varepsilon_n = \mathrm{Prob}[D_n(C_n(a_{i_1} \dots a_{i_n}, b_{i_1} \dots b_{i_n}), a_{i_1} \dots a_{i_n}) \neq (a_{i_1} \dots a_{i_n})]$$

(кодирующее отображение C_n *в качестве аргументов получает* α_i *и* β_i *для* $i = 1, \dots, n$*, а декодирующее отображение в качестве аргументов получает* D_n *к* β_i *для* $i = 1, \dots, n$ *и кодовое слово — значение функции* C_n*.).*

При $n \to \infty$ *вероятность ошибки стремится к нулю, если* $L > H(\alpha \,|\, \beta)$*, и стремится к единице, если* $L < H(\alpha \,|\, \beta)$*.*

(2) [*декодирование с помощником*] *Аналогичное утверждение (вероятность ошибки стремится к нулю, если* $L > H(\alpha \,|\, \beta)$*, и стремится к единице, если* $L < H(\alpha \,|\, \beta)$*) выполнено, если доступ к* β_i *имеет только декодирующее отображение* D_n*, но не кодирующее отображение* C_n*.*

8 Теория информации в криптографии

Следующая классическая теорема Шеннона показывает, насколько длинным должен быть секретный ключ, чтобы схема шифрования была абсолютно надежна.

Теорема 6 (Шеннона о надежном шифровании с секретным ключом)**.** *Опишем схему шифрования с секретным ключом.* Отправитель *хочет переслать* Получателю *сообщение* M *(выбранное по некоторому заранее известному распределению вероятностей). Отправитель и Получатель заранее договорились о значении секретного ключа* K*. Отправитель производит «шифрование» с использованием секретного ключа и вычисляет значение* $C = C(M, K)$*. Затем он пересылает по открытому каналу связи полученное зашифрованное сообщение* C *. Получатель производит расшифровку и восстанавливает по зашифрованному сообщению* C *и значению ключа* K *исходное сообщение* M *. При этом требуется, чтобы передаваемое по открытому каналу зашифрованное сообщение* C *не имело никакой взаимной информации с сообщением* M *(свойство секретности). Для любой схему такого вида энтропия секретного ключа* K *не может быть меньше энтропии передаваемого соощения* M*.*

Условие теоремы можно переписать в виде следующих информационных неравенств:

$$\begin{cases} H(C|M,K) &= 0, \\ H(M|C,K) &= 0, \\ I(C:M) &= 0. \end{cases}$$

Из этих трех условий и базисных неравенств для энтропии нетрудно вывести $H(K) \geq H(M)$.

Замечание 1: Утверждение теоремы остается верным и без первого условия $H(C|M,K) = 0$ (достаточно потребовать $H(M|C,K) = 0$ и $I(C:M) = 0$). Содержательно это означает, что размер (энтропия) ключа должна быть большой не только дл] детерминированных схем шифрования, но и для схем шифрования с вероятностной процедурой кодирования $(M,K) \mapsto C$.

Замечание 2: В известной схеме Вернама (Gilbert Vernam) шифрование состоит в побитовом XOR n-битового сообщения M с n-битовым же ключом K, т.е., $C_i = M_i \oplus K_i$, $i = 1,\ldots,n$. (Декодирование выполняется аналогично: $M_i = C_i \oplus K_i$, $i = 1,\ldots,n$.) Если исходное сообщение M равномерно распределено на всех n-битовых строках и биты ключа K также выбираются случайно и равномерно (независимо от M), то $H(M) = H(K) = n$. В данном случае энтропии сообшения и ключа совпадают.

9 Схемы разделения секрета.

Определение совершенное схемы разделения секрета.

Утверждение 3. *Предположим, что не один участник схемы разделения секрета не может восстановить секрет в одиночку. Тогда любую совершенную схему разделения секрета для такой схемы можно модифицировать так, чтобы распределение на множестве секретовстало равномерным, а энтропии долей секрета каждого участника схемы не изменились.*

Участник номер i схемы разделения секрета называется существенным, если некоторый набор участников $\{j_1,\ldots,j_r\}$ не имеет никакой информации о секрете, но участники $\{i,j_1,...,j_r\}$ вместе могут однозначно восстановить секрет (добавление i-го участника превращает некотроую группу из неавторизованной в авторизованную).

Домашнее задание 9.1. *Докажите, что для каждого существенного участника совершенной схемы разделения секрета выполнено неравенство $H(S_i) \geq H(S_0)$ (энтропия доли такого участника не может быть меньше, чем энтропия самого секрета).*

Определение идеальной схемы разделения секрета. Идеальное совершенное разделение секрета для пороговой структуры доступа (полиномиальная схема Шамира).

Пример структуры доступа, не имеющей соврешенной идеальной схемы разделения секрета: структура из 4 участников, минимальными авторизованными группами в которой являются пары $\{\{1,2\},\{2,3\},\{3,4\}\}$.

Утверждение 4. *Докажите, что для описанной выше структуры доступа существует схема разделения секрета (S_0,S_1,S_2,S_3,S_4), в которой*

$$\max_{i>0} \frac{H(S_i)}{H(S_0)} = 3/2.$$

Домашнее задание 9.2. *Рассмотрим структуру доступа для 4 участников, минимальными авторизованными группами в которой являются пары* $\{\{1,2\},\{2,3\},\{3,4\},\{2,4\}\}$. *Докажите, что для данной схемы не существует совершенной идеальной схемы разделения секрета.*

10 Простейшие конструкции теории кодирования (напоминание из курса теории кодирования)

Комбинаторная модель канала с шумом. Определение кода длины n над алфавитом Σ, исправляющего e ошибок. Расстояние кода d (минимальное расстояние между двумя несовпадающими кодовыми словами); связь кодового расстояния и числа исправляемых ошибок.

Утверждение 5 (граница Хэмминга (volume bound))**.** . *Для любого кода с параметрами* (n,k,d) *выполнено неравенство*

$$2^n \geq 2^k \cdot \left(C_n^0 + \ldots + C_n^e\right),$$

где $e = \lfloor \frac{d-1}{2} \rfloor$.

Утверждение 6 (граница Гилберта (Gilbert))**.** *Если*

$$2^n \geq 2^k \cdot \left(C_n^0 + \ldots + C_n^{d-1}\right),$$

то существует код с параметрами (n,k,d).

Асимптотическое поведение границы Хэмминга и границы Гилберта для кодов с расстоянием $d = \alpha n$ при $n \to \infty$. Определение линейных кодов над алфавитом $\Sigma = \mathbb{F}_q$ (алфавит Σ состоит из элементов конечного поля размера q).

Утверждение 7. *Расстояние линейного кода* C *равно*

$$\min_{0 \neq x \in C} \operatorname{dist}(x,0),$$

что совпадает с

$$\min_{0 \neq x \in C} \omega(x),$$

где $\omega(x)$ *(вес слова* x*) есть число ненулевых элементов в* x.

Порождающая и проверочная матрицы линейного кода. *Замечание:* в линейном коде с расстоянием d любые $d-1$ столбцов проверочной матрицы линейно независимы.

Утверждение 8 (Граница Варшамова–Гилберта)**.** *Если*

$$2^n > 2^{k-1}\left(C_n^0 + \ldots + C_n^{d-1}\right),$$

то существует линейный код над полем из двух элементов с парамерами $[n,k,\geq d]$.

11 Вероятностная модель канала с шумом.

Дискретный канал без памяти (канал задаётся входным алфавитом A, выходным алфавитом B и набором условных вероятностей p_{ij} для $i = 1, \ldots, |A|$ и $j = 1 \ldots, |B|$). Формальное определение пропускной способности дискретного канала без памяти.
Примеры каналов:

(1) двоичный симметричный канал, в котором пересылаемый бит меняется на противоположный с веротяностью ε;

(2) двоичный несимметричный канал, в котором пересылаемый ноль всегда передается без ошибок, а пересылаемая единица с вероятностью ε превращается в ноль;

(3) канал с алфавитом $\{0, 1\}$ на входе и $\{0, 1, *\}$ на выходе: ноль и единица с вероятностью $(1 - \varepsilon)$ пересылаются без ошибок, и с вероятностью ε превращаются в $*$.

На лекции мы вычислили пропускную способность для двоичного симметричного канала.

Домашнее задание 11.1. *Вычислите пропускную способность каналов из двух других приведенных выше примеров.*

Теорема 7 (Шеннона о кодировании для дискретного канала с шумом)**.** *Пусть пропускная способность канала без памяти (с входным алфавитом A и выходным алфавитом B) равна R.*

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

$$C_k : \{0, 1\}^k \to A^{\lfloor k/L \rfloor}$$

и

$$D_k : B^{\lfloor k/L \rfloor} \to \{0, 1\}^k$$

такие, что вероятность ошибки σ_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 $k \to \infty$.

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

$$C_k : \{0, 1\}^k \to A^{\lfloor k/L \rfloor}$$

и

$$D_k : B^{\lfloor k/L \rfloor} \to \{0, 1\}^k$$

вероятность ошибки σ_k не стремится к нулю при $k \to \infty$.

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

12 Колмогоровская сложность

Определение 3. *Пусть* $U : \{0,1\}^* \to \{0,1\}^*$ *есть* (*частичная*) *вычислимая функция. Определим* $KS_U(x) = \min\{|p| \ : \ U(p) = x\}$. (*Минимум пустого множества считается равным бесконечности.*)

Теорема 8. *Существует такой способ описания* (*частичная вычислимая функция*) U*, что для любой другой* V *и для всех слов* x

$$KS_U(x) \leq KS_V(x) + O(1)$$

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

Простейшие свойства колмогоровской сложности:

- $KS(x) \leq |x| + O(1)$;
- $KS(xx) \leq |x| + O(1)$;
- $KS(f(x)) \leq KS(x) + O(1)$ 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 $n - c \leq KS(x) \leq n + c$.

Теорема 9. *Не существует алгоритма, который по заданному* n *находил бы слово с колмогоровской сложностью не менее* n.

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

Следствие: Оптимальный способ описания не может быть всюду определенной функцией.

Следствие: Почти все утверждения вида $KS(x) > c$ истинны и при этом недоказуемы.

Утверждение 9. *(а)* $KS(x,y) \leq KS(x) + KS(y) + O(\log(|x| + |y|))$.

(б) Для любого $C > 0$ *найдутся такие слова* x *и* y*, что*

$$KS(x,y) > KS(x) + KS(y) + C.$$

Определение *относительной* (или *условной*) колмогоровской сложности. Простейшие свойства относительной колмогоровской сложности:

- $KS(x|y) \leq |x| + O(1)$;
- $KS(x|\Lambda) = K(x) + O(1)$;
- $KS(x|y) \leq K(x) + O(1)$ 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 x длины n выполнено $n - c \leq KS(x|y) \leq n + c$.

Теорема 10 (Колмогорова–Левина)**.**

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

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

Домашнее задание 12.1. *Существует ли такой оптимальный способ описания U, для которого $KS(x)$ для каждого x оказывается (а)* чётным числом*? (б) Степенью двойки?*

Домашнее задание 12.2. *Пусть слово x длины n состоит из pn единиц и $(1-p)n$ нулей. Тогда*

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

Домашнее задание 12.3. *Докажите, что*
(а) $K(x, y, z) + K(z) \leq K(x, z) + K(y, z) + O(\log K(x, y, z))$,
(б) $K(x, y, z) + K(z) \leq K(x, z) + K(y, z) + O(\log K(x, y, z))$,
(в) $K(z) \leq K(z|x) + K(z|y) + I(x : y)$.

Домашнее задание 12.4. $KS(x, KS(x)) = KS(x) + O(1)$.

13 Применение колмогоровской сложности

Теорема 11. *Для того чтобы распознать языка палиндромов, одноленточной машине Тьюринга (с одной головкой) требуется время $\Omega(n^2)$.*

Теорема 12. *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).*

14 Случайность по Мартин-Лёфу.

Утверждение 10. *Для всякой бесконечной двоичной последовательности $\omega_0\omega_1\omega_2\ldots$ найдутся сколь угодно большие номера n такие, что*

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

Определение префиксной колмогоровской сложности $KP_U(x)$ для вычислимой функции U с префиксной областью определения. Теорема о существовании оптимального префиксного способа описания. Простейшие свойства префиксной сложности.

Утверждение 11.

$$\sum_{x\in\{0,1\}^*} 2^{-KP(x)} \leq 1.$$

Домашнее задание 14.1. *Докажите, что*

$$KP(x) \leq |x| + 2\log|x| + O(1).$$

Домашнее задание 14.2. *Докажите, что*

$$KP(x,y) \leq KP(x) + KP(y) + O(1).$$

Случайность по Мартин-Лёфу: Бесконечная двоичная последовательность $\omega_0\omega_1\omega_2\ldots$ называется случайной, если существует такая константа C, что для всех n

$$KP(\omega_0\ldots\omega_{n-1}) \geq n - C.$$

Утверждение 12. *Почти все (по равномерной бернуллиевской мере) бесконечные двоичные последовательности $\omega_0\ldots\omega_n\ldots$ случайны по Мартин-Лёфу.*

Домашнее задание 14.3. *(а) Докажите, что случайная по Мартин-Лёфу последовательность не может быть вычислимой.*

(б) Докажите, что последовательность вида $\omega_0 0\omega_1 0\omega_2 0\ldots\omega_n 0\ldots$ не может быть сучайной по Мартин-Лёфу.

(в) Докажите, что последовательность вида $\omega_0\omega_0\omega_1\omega_1\ldots\omega_n\omega_n\ldots$ не может быть сучайной по Мартин-Лёфу.

Теорема 13 (Закон больших чисел в форме Харди–Литлвуда)**.** *Для почти всех двоичных последовательностей $\omega_0\omega_1\omega_2\ldots$*

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

15 Принцип кратчайшего описания.

Принцип кратчайшего описания (MDL, minimum description length principle) в машинном обучении и анализе данных. Примеры использования метода MDL.

16 Коммуникационная сложность

Определение детерминированного коммуникационного протокола для двух участников. Определение детерминированной коммуникационная сложности $\mathrm{CC}(f)$ функции $f : A \times B \to C$ для конечных множеств A, B, C. Доказательство оценок $\mathrm{CC}(f) \leq \lceil \log |A| \rceil + \lceil \log |B| \rceil$ и $\mathrm{CC}(f) \leq \lceil \log |A| \rceil + \lceil \log |C| \rceil$.

Теорема 14. *Детерминированная коммуникационная сложность предиката равенства* $\mathrm{EQ}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, *определяемого как*

$$\mathrm{EQ}_n(a,b) = \begin{cases} 1, & \text{если } a \neq b, \\ 0, & \text{иначе,} \end{cases}$$

равна $n+1$.

Схема доказательства теоремы: Верхняя оценка очевидна. Для получения нижней оценки замечаем, что каждому листу в коммуникационном протоколе соответствует «комбинаторный прямоугольник».

Вероятностные коммуникационный протоколы раздельными источниками случайных битов Алисы и Бобаопределение вероятностной коммуникационной сложности $\mathrm{RCC}_\varepsilon(f)$. Доказательство оценки

$$\mathrm{RCC}_\varepsilon(\mathrm{EQ}_n) = O(\log \frac{n}{\varepsilon}).$$

Теорема 15. *Для любого* $\varepsilon < 1/2$ *и для любого предиката* $f(x,y)$

$$\mathrm{RCC}_\varepsilon(f) = \Omega(\log \mathrm{CC}(f)).$$

Следствие: $\mathrm{RCC}^{\mathrm{pub}}_\varepsilon(\mathrm{EQ}_n) = \Theta(\log n)$ для каждого $\varepsilon < 1/2$.

Домашнее задание 16.1. *Докажите, что для предиката*

$$\mathrm{GT}_n(a,b) = \begin{cases} 1, & \text{если } a > b, \\ 0, & \text{иначе} \end{cases}$$

$\mathrm{CC}(\mathrm{GT}_n) = n+1$.

Домашнее задание 16.2. *Докажите, что для предиката дизъюнктности* $\mathrm{DISJ}_n : 2^{\{1,\ldots,n\}} \times 2^{\{1,\ldots,n\}} \to \{0,1\}$ (*предикат дизъюнктности пары множеств* $A, B \subset \{1,\ldots,n\}$), *определяемого как*

$$\mathrm{DISJ}_n(A,B) = \begin{cases} 1, & \text{если } A \cap B = \emptyset, \\ 0, & \text{иначе,} \end{cases}$$

детерминированная коммуникационная сложность равна $n+1$.

Домашнее задание 16.3. *Рассмотрим функцию*

$$\mathrm{MAX}_n : \{1,\ldots,n\} \times \{1,\ldots,n\} \to \{1,\ldots,n\}$$

(*максимум из двух целых чисел из интервала от* 1 *до* 2^n). *Докажите, что*

(а) $\mathrm{CC}(\mathrm{MAX}_n) \leq 2n$,
(б) $\mathrm{CC}(\mathrm{MAX}_n) \leq \frac{3}{2}n + O(1)$,
(в) $\mathrm{CC}(\mathrm{MAX}_n) \leq n + O(\sqrt{n})$.

Домашнее задание 16.4. *Докажите, что для предиката* GT_n *из домашнего задания 16.1 и любого* $\varepsilon > 0$

(а) $\mathrm{RCC}_\varepsilon(\mathrm{GT}_n) = O(\log^2 n)$,

*(б)** $\mathrm{RCC}_\varepsilon(\mathrm{GT}_n) = O(\log n)$.

Список литературы

[1] T.M. Cover, J.A. Thomas. Elements of information Theory, 2006.

[2] R.W. Yeung. A First Course in Information Theory, 2002.

[3] Н.К. Верещагин, Е.В. Щепин. Информация, кодирование, предсказание, 2012.

[4] И. Чисар, Я. Кернер. Теория информации, 1985.

[5] Г.А. Кабатянский. Математика разделения секрета, *Математическое просвещение*, т. 2, No 3, стр. 115–126, 1998.

[6] M. Li, and P. Vitanyi. An Introduction to Kolmogorov Complexity and Its Applications, 2008.

[7] В. А. Успенский, Н. К. Верещагин, А. Шень. Введение в колмогоровскую сложность, 2012.

[8] Румянцев, Ромащенко, Шень. Заметки о теории кодирования, 2011. `http://www.mccme.ru/~anromash/courses/coding-theory.ps`

[9] А.М. Яглом, И.М. Яглом. Вероятность и информация, 1973.

[10] Ф.Дж.А. Мак-Вильямс, Н.Дж.А. Слоэн. Теория кодов, исправляющих ошибки, 1979.

[11] E. Nisan, N. Kushilevitz. Communication complexity, 1997.

Дополнительная литература.

[12] Р. Галлагер. Теория информации и надежная связь, 1974.

[13] В.М. Сидельников. Теория кодирования, 2008.