Заметки к курсу „Теория информации“
Лекции 1 и 2

14 февраля 2019 г.

Аннотация

Курс посвящён изучению подходов к определению понятия „количество информации“. Последовательность изложения материала данного курса основана на классической статье Колмогорова „Три подхода к определению понятия количества информации“ (1965).

В курсе будет рассмотрено три подхода к определению „количества информации“: комбинаторный (информация по Хартли), вероятностный (энтропия Шеннона) и алгоритмический (Колмогоровская сложность). Кроме этого мы поговорим про различные применения аппарата теории информации в различных областях компьютерных наук: в криптографии, в коммуникационной сложности, в теории кодирования, в теории конечных автоматов, в теории сложности вычислений и некоторых других.

1. Комбинаторный подход

1.1. Информация по Хартли

Пусть задано некоторое конечное множество A — *множество исходов*.

Определение 1.1 (1928)**.** Определим *количество информации в* A как $\chi(A) = \log_2 |A|$ (мы будем измерять количество информации в битах, поэтому все логарифмы будут по основанию 2, для байтов основание нужно было бы заменить на 256).

Если про некоторый $x \in A$ стало известно, что $x \in B$, то теперь для идентификации x нам достаточно $\chi(A \cap B) = \log |A \cap B|$ битов, т.е. нам сообщили $\chi(A) - \chi(A \cap B)$ битов информации.

Пример 1.1. Предположим, что мы хотим узнать некоторое неизвестное упорядочение множества $\{a_1, a_2, \ldots, a_5\}$. Нам стало известно, что $a_1 > a_2$ или $a_3 > a_4$. Сколько битов информации мы узнали? Множество A состоит из 5! перестановок, множество B — из перестановок, которые удовлетворяют новому условию. Легко проверить, что $|B| = 90$. Итого мы узнали $\log 120 - \log 90 = \log(4/3)$ битов.

Пусть $A \subset \{0,1\}^* \times \{0,1\}^*$. Обозначим через $\pi_1(A)$ и $\pi_2(A)$ проекции множества A на первую и вторую координату соответственно, а $\chi_1(A) = \log|\pi_1(A)|$ и $\chi_2(A) = \log|\pi_2(A)|$ — их сложность по Хартли.

Теорема 1.1. $\chi(A) \le \chi_1(A) + \chi_2(A)$.

Определение 1.2. Количество информации в второй компоненте A при известной первой

$$\chi_{2|1} = \log\left(\max_{a \in \pi_1(A)} |A_a|\right),$$

где $A_a = \{(a,x) \mid x \in \pi_2(A)\}$.

Теорема 1.2. $\chi(A) \le \chi_1(A) + \chi_{2|1}(A)$.

Теорема 1.3. *Для* $A \subset \{0,1\}^* \times \{0,1\}^* \times \{0,1\}^*$

$$2 \cdot \chi(A) \le \chi_{12}(A) + \chi_{13}(A) + \chi_{23}(A).$$

Следствие 1.1. *Квадрат объёма трёхмерного тела не превосходит произведение площадей его проекций на координатные плоскости.*

1.2. Применение: игра в 10 вопросов

Сколько вопросов на ДА/НЕТ нужно задать, чтобы определить загаданное число от 1 до N, если (а) можно задавать вопросы адаптивно; (б) вопросы нужно написать на бумажке заранее.

Оценка $\lceil \log N \rceil$ достигается в обоих случаях, если задавать вопросы про биты двоичного представления загаданного числа.

Докажем нижнюю оценку. Пусть $A = [N]$. Множество $Q = \{(q_1, q_2, \ldots, q_k)\}$ — множество протоколов (ответы на вопросы). Можно рассматривать A и Q как проекции некоторого множества исходов игры $S \subset A \times Q$ на разные координаты. Тогда верны следующие неравенства:

- $\chi_Q(S) = \chi(Q) \le \chi_1(Q) + \chi_2(Q) + \cdots + \chi_k(Q) \le k$,
- $\chi_A(S) = \chi(A) \le \chi(S) \le \chi_Q(S) + \chi_{A|Q}(S) \le k + 0 = k$.

Таким образом получаем, что $\log N = \chi(A) \le k$.

1.3. Цена информации

Пусть имеется некоторое неизвестное число от 1 до n (где $n \ge 2$). Разрешается задавать любые вопросы с ответами ДА/НЕТ. При ответе ДА мы заплатим 1 рубль, а при ответе НЕТ — два рубля. Сколько необходимо и достаточно заплатить для отгадывания числа?

Верхняя оценка. Давайте задавать вопросы так, чтобы отрицательные ответы приносили в два раза больше информации, чем положительные. Тогда за каждый бит информации мы заплатим некоторое константное количество рублей c. Пусть все вопросы будут вида „$x \in T$?“. Потребуем, чтобы

$$2 \cdot (\log|X| - \log|X \cap T|) = \log|X| - \log|X \cap \overline{T}|.$$

Пусть $|X \cap T| = \alpha|X|$, тогда $|X \cap \overline{T}| = (1-\alpha)|X|$, т.о. $\alpha^2 = 1 - \alpha$, $\alpha = (\sqrt{5}-1)/2$. При любом ответе мы заплатим $c = 1/(-\log\alpha) \approx 1.44$ рублей за бит, а в целом — $\log n/(-\log\alpha)$ рублей.

Нижняя оценка. Применим рассуждение про злонамеренного противника (adversary argument). Пусть противник выбирает ответ ДА/НЕТ в зависимости от того, какое из двух значений $1/(\log|X| - \log|X \cap T|)$ и $2/(\log|X| - \log|X \cap \overline{T}|)$ больше. При любых X, T одно из этих значений не меньше $c = 1/(-\log\alpha)$. Таким образом мы заставляем алгоритм платить не менее c рублей за бит, а значит любой алгоритм в худшем случае заплатит $\lceil c \log n \rceil$ рублей.

1.4. Применение: упорядочивание камней по весу

1.4.1. Верхняя и нижняя оценки для произвольного N

Сколько сравнений нужно сделать для того, чтобы упорядочить N камней по весу?

Нижняя оценка. Потребуется $\lceil \chi(S_N) \rceil = \lceil \log n! \rceil$ сравнений.

Верхняя оценка. Будем сортировать вставкой с бинарным поиском места вставки. Количество сравнений:

$$\lceil \log 2 \rceil + \lceil \log 3 \rceil + \cdots + \lceil \log n \rceil \le \log n! + n - 1 = n \log n + O(n).$$

1.4.2. Точные оценки для маленьких N

Упражнение 1.1. Сколько нужно взвешиваний, чтобы упорядочить N камней по весу? Найдите точный ответ на этот вопрос для $N = 2, 3, 4, 5$. Указание: воспользуйтесь жадной стратегией, при которой каждое взвешивание приносит максимум информации.

1.5. Применение: поиск фальшивой монетки

- 20 монет, одна фальшивая легче остальных.

  Каждое взвешивание даёт не более $\log 3$ битов. Итого $k \ge \log N / \log 3 = \log_3 N$.

- 13 монет, одна фальшивая (с неизвестным относительным весом), 3 взвешивания.

  Два варианта первого шага:

– если взвешиваем по 4, то при равенстве нельзя из 5 за два взвешивания найти фальшивую (остаётся 10 исходов),

– если взвешиваем по 5, то при неравенстве остаётся 10 возможных исходов.

- 15 монет, одна фальшивая, три взвешивания. Не требуется узнавать относительный вес монеты.

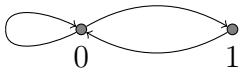
  Всего исходов $2 \cdot 14 + 1 > 27$, т.к. только в случае трёх равенств мы можем не узнать относительный вес фальшивой монеты.

- 14 монет, одна фальшивая, три взвешивания. Не требуется узнавать относительный вес монеты.

  Всего исходов $2 \cdot 13 + 1 \leq 27$, но определить тем не менее нельзя. Аппарата информации по Хартли недостаточно.

Упражнение 1.2. За три взвешивания найти одну фальшивую монету из 12, если её относительный вес неизвестен. Указание: воспользуйтесь „жадной“ стратегией, при которой каждое взвешивание приносит максимум информации.

Упражнение 1.3. Пусть L_n — множество путей длины n в графе.



Чему равен предел $\lim_{n\to\infty} \frac{\chi(L_n)}{n}$?

2. Вероятностный подход

2.1. Энтропия Ше́ннона

Энтропия Шеннона определяет количество информации $H(\alpha)$ в распределении вероятностей для некоторой случайной величины α. Пусть α принимает значения из множества $\{a_1, a_2, \ldots, a_k\}$ с вероятностями $\{p_1, p_2, \ldots, p_k\}$, $p_i \geq 0$, $\sum_i p_i = 1$.

Нам бы хотелось, чтобы это определение согласовывалось с определением Хартли, т.е. имеют место следующие „граничные условия“:

- если $p_1 = \cdots = p_k$, то $H(\alpha) = \log k$,
- если $p_1 = 1$, $p_2 = \cdots = p_k = 0$, то $H(\alpha) = 0$.

Будем искать $H(\alpha)$ в виде математического ожидания информации, которую мы получаем от каждого исхода.

$$H(\alpha) = \sum_i p_i \cdot (\text{информация в } a_i).$$

Как оценить, сколько информации в исходе a_i? Пусть U — всё пространство элементарных исходов, все исходы которого равновероятны. Тогда событию $\alpha = a_i$ соответствует множеству элементарных исходов меры p_i. Соответственно, если случилось событие $\alpha = a_i$, то размер множества согласованных с этим событием исходов уменьшается с $|U|$ до $p_i|U|$, т.е. событие $\alpha = a_i$ сообщает нам $\log|U| - \log(p_i|U|) = \log\frac{1}{p_i}$ битов информации.

Определение 2.1 (1948)**.** Энтропия Шеннона случайной величины α

$$H(\alpha) = \sum_{i=1}^{k} p_i \cdot \log\frac{1}{p_i}.$$

(По непрерывности доопределим $0 \cdot \log\frac{1}{0} = 0$.)

Лемма 2.1. *Для энтропии Шеннона выполняются следующие соотношения.*

- $H(\alpha) \geq 0$, *причём* $H(\alpha) = 0 \iff$ *распределение* α *вырождено.*
- $H(\alpha) \leq \log k$, *причём* $H(\alpha) = \log k \iff$ *величина* α *распределена равномерно.*

Для доказательства нам потребуется следующая теорема.

Теорема 2.1 (Неравенство Йенсена)**.** *Пусть функция* $f(x)$ *является вогнутой на некотором промежутке* $\mathcal{X}$ *и числа* $q_1, q_2, \ldots, q_n > 0$ *таковы, что* $q_1 + \cdots + q_n = 1$*. Тогда для любых* $x_1, x_2, \ldots, x_n$ *из промежутка* $\mathcal{X}$ *выполняется неравенство:*

$$\sum_{i=1}^{n} q_i f(x_i) \leq f\left(\sum_{i=1}^{n} q_i x_i\right).$$

Доказательство леммы 2.1. Первое свойство следует напрямую из определения: каждый член суммы $H(\alpha)$ неотрицателен и равен нулю только в случае, если $p_i = 0$ или $p_i = 1$.

Для доказательства второго неравенства перенесём всё в левую часть и применим неравенство Йенсена:

$$H(\alpha) - \log k = \sum_{i=1}^{k} p_k \cdot \log\frac{1}{p_i} - \sum_{i=1}^{k} p_i \cdot \log k = \sum_{i=1}^{k} p_k \cdot \log\frac{1}{p_i k} \leq \log\left(\sum_{i=1}^{k} p_i \frac{1}{p_i k}\right) = \log 1 = 0.$$

□

Энтропию совместного распределения пары случайных величин α и β будем обозначать $H(\alpha, \beta)$.

Лемма 2.2. *Выполняются следующие свойства:*

- $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$, *причём равенство достигается тогда и только тогда, когда случайные величины независимы;*

- $H(\alpha) \leq H(\alpha, \beta)$, *причём равенство достигается тогда и только тогда, когда* β *полностью определяется значением* α, *т.е.* $\beta = f(\alpha)$.

Доказательство. Введём обозначения для вероятностей событий совместного распределения вероятностей (α, β). Пусть пара (a_i, b_j) имеет вероятность $p_{i,j}$, событие $[\alpha = a_i]$ имеет вероятность $p_{i,*} = p_{i,1} + \cdots + p_{i,n}$, а событие $[\beta = b_j]$ — вероятность $p_{*,j} = p_{1,j} + \cdots + p_{k,j}$. В этих обозначениях неравенство $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$ переписывается как

$$\sum_{i,j} \cdot p_{i,j} \cdot \log \frac{1}{p_{i,j}} \leq \sum_i \sum_j p_{i,j} \cdot \log \frac{1}{p_{i,*}} + \sum_j \sum_i p_{i,j} \cdot \log \frac{1}{p_{*,j}}.$$

Перенесём всё в левую часть и применим неравенство Йенсена.

$$\sum_{i,j} p_{i,j} \cdot \log \frac{p_{i,*} \cdot p_{*,j}}{p_{i,j}} \leq \log \left(\sum_{i,j} p_{i,j} \cdot \frac{p_{i,*} \cdot p_{*,j}}{p_{i,j}} \right) = \log \left(\sum_{i,j} p_{i,*} \cdot p_{*,j} \right) =$$

$$= \log \left(\left(\underbrace{\sum_i p_{i,*}}_{1} \right) \cdot \left(\underbrace{\sum_j p_{*,j}}_{1} \right) \right) = 0.$$

Равенство в неравенстве Йенсена для $f(x) = \log(x)$ достигается только, если все точки равны, т.е. для любых i, j $\frac{p_{i,*}p_{*,j}}{p_{i,j}} = c$ для некоторой константы c. Несложно заметить, что $c = 1$, т.к. выполняется следующее равенство $\sum_{i,j} p_{i,*}p_{*,j} = c \sum_{i,j} p_{i,j}$ в котором обе суммы равны 1. Таким образом в случае равенства α и β независимы.

Доказательство второго свойства мы получим как следствие из свойств условной энтропии. □

Определение 2.2. Энтропия α при условии $\beta = b_j$

$$H(\alpha \mid \beta = b_j) = \sum_i \Pr[\alpha = a_i \mid \beta = b_j] \cdot \log \frac{1}{\Pr[\alpha = a_i \mid \beta = b_j]}.$$

Определение 2.3. *Условная (относительная) энтропия* α относительно β

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

Другими словами

$$H(\alpha \mid \beta) = \mathop{\mathrm{E}}_{b_j \leftarrow \beta} [H(\alpha \mid \beta = b_j)].$$

Если подставить определение 2.2, то можно получить выражение для условной энтропии через отдельные вероятности событий.

$$H(\alpha \mid \beta) = \sum_j \Pr[\beta = b_j] \cdot \sum_i \Pr[\alpha = a_i \mid \beta = b_j] \cdot \log \frac{1}{\Pr[\alpha = a_i \mid \beta = b_j]} = \sum_{i,j} p_{i,j} \cdot \log \frac{p_{*,j}}{p_{i,j}}.$$

Лемма 2.3. *Условная энтропия обладает следующими свойствами.*

- $H(\alpha \mid \beta) \geq 0$.
- $H(\alpha \mid \beta) = 0 \iff \alpha$ *однозначно определяется по* β.
- $H(\alpha, \beta) = H(\beta) + H(\alpha \mid \beta) = H(\alpha) + H(\beta \mid \alpha)$.

Доказательство. Первое свойство выполняется, т.к. условная энтропия это матожидание неотрицательной случайной величины. Второе свойство объясняется тем, что для любого j распределение $\langle \alpha \mid \beta = b_j \rangle$ имеет нулевую энтропию, т.е. распределение вырождено и каждому b_j соответствует ровно один a_i. Третье свойство следует из следующего равенства.

$$\sum_{i,j} p_{i,j} \cdot \log \frac{1}{p_{i,j}} = \sum_{i,j} p_{i,j} \cdot \log \frac{1}{p_{*,j}} + \sum_{i,j} p_{i,j} \cdot \log \frac{p_{*,j}}{p_{i,j}}.$$

(Нужна аккуратность, если есть строки, которые состоят из одних нулей, т.е $p_{*,j} = 0$ — такие строки не нужно включать в эти суммы.) □

Следствие 2.1. $H(\alpha, \beta) \geq H(\alpha)$, *причём равенство достигается тогда и только тогда, когда* $\beta = f(\alpha)$.

Доказательство. $H(\alpha, \beta) - H(\alpha) = H(\beta \mid \alpha) \geq 0$. По второму свойству условной энтропии равенство достигается тогда и только тогда, когда $\beta = f(\alpha)$. □

2.2. Взаимная информация

Определение 2.4. *Информация в* α *о величине* β определяется следующим соотношением:

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

Эту величину так же *называют взаимной информацией случайных величин* α *и* β.

Лемма 2.4. *Для взаимной информации выполняются следующие соотношения.*

1. $I(\alpha : \beta) \leq H(\alpha)$.
2. $I(\alpha : \beta) \leq H(\beta)$.
3. $I(\alpha : \alpha) = H(\alpha)$.
4. $I(\alpha : \beta) = I(\beta : \alpha)$.
5. $I(\alpha : \beta) = H(\alpha) + H(\beta) - H(\alpha, \beta)$.

Определение 2.5. Пусть α, β, γ — случайные величины. Определим *взаимную информацию в* α *о* β *при условии* γ.

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

2. $I(\alpha : \beta \mid \gamma) = \sum_\ell I(\alpha : \beta \mid \gamma = c_\ell) \cdot \Pr[\gamma = c_\ell]$.

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

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

Лемма 2.5. *Все определения условной взаимной информации эквивалентны.*

Доказательство. (3) $\iff$ (4).

$(3) = H(\alpha \mid \gamma) + H(\beta \mid \gamma) - H(\alpha, \beta \mid \gamma) = H(\alpha, \gamma) - H(\gamma) + H(\beta, \gamma) - H(\gamma) - H(\alpha, \beta, \gamma) + H(\gamma)$.

□

Утверждение 2.1 (chain rule for mutual information)**.** *Имеют место следующие соотношения:*

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

2. $I((\alpha, \beta) : \gamma \mid \delta) = I(\alpha : \gamma \mid \delta) + I(\beta : \gamma \mid \alpha, \delta)$.

2.3. Применение: опять о поиске фальшивой монетки

Теперь у нас достаточно знаний, чтобы доказать, что за три взвешивания нельзя найти одну фальшивую монету из 14, даже если не нужно определять её относительный вес.

Доказательство. Предположим, что существует способ найти фальшивую монету за три взвешивания. Тогда протокол взвешивания можно представить в виде полного троичного дерева, где каждый лист помечен номером монетки, которая оказалась фальшивой (у нас как раз ровно $3^3 = 27$ исходов).

Давайте введём следующее распределение вероятностей α. Пусть монета, номер которой находится в листе, соответствующем трём равенствам (такой лист только один), имеет номер i. В нашем распределении вероятностей монета с номером i будет фальшивой с вероятностью $1/27$. Оставшиеся монеты оказываются фальшивыми с вероятностями $2/27$, причём с вероятностью $1/27$ монета оказывается легче, чем настоящая, и с такой же вероятностью она оказывается тяжелее настоящей.

$$H(\alpha) = \log 27 = 3 \log 3.$$

Пусть случайные величины β_1, β_2, β_3 соответствуют результатам первого, второго и третьего взвешивания соответственно. Значение α однозначно определяется после трёх взвешиваний: $H(\alpha \mid \beta_1, \beta_2, \beta_3) = 0$, а следовательно

$$H(\alpha) \leq H(\beta_1, \beta_2, \beta_3) \leq H(\beta_1) + H(\beta_2) + H(\beta_3) \leq 3 \log 3.$$

Таким образом каждое взвешивание должно иметь энтропию ровно $\log 3$. Рассмотрим первое взвешивание. Пусть на чашах весов лежит по k монет. Вероятность каждого исхода взвешивания ($<$, $>$, $=$) относительно распределения α должна быть ровно $1/3$.

$$\Pr[<] = \frac{k}{27} + \frac{k}{27} = \frac{1}{3}.$$

Таким образом $2k = 9$, а значит нет такого целого k. □

Упражнение 2.1. Пусть у нас есть N камней разного веса и чашечные весы. Сколько нужно взвешиваний, чтобы найти

1. самый тяжёлый и второй по тяжести камень,
2. самый тяжёлый и самый лёгкий камни.

Список литературы

[1] Н.К. Верещагин, Е.В. Щепин. *Информация, кодирование, предсказание*, МЦНМО, 2012.

[2] Н.К. Верещагин. *Коммуникационная сложность*, Computer Science клуб, 2017. http://compsciclub.ru/courses/communicationcomplexity/2017-spring/

[3] А.Е. Ромащенко. *Введение в теорию информации,* Computer Science клуб, 2015. http://compsciclub.ru/courses/informationtheory/2015-spring/

[4] А.Е. Ромащенко. *Краткий конспект лекций курса "Введение в теорию информации",* 2014. http://www.mccme.ru/~anromash/courses/lecture-notes-it-2014.pdf

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

[6] А. Шень. *Алгоритмическая теория информации,* Computer Science клуб, 2008. http://compsciclub.ru/courses/algo-information-theory/2008-autumn/

[7] D. Gavinsky, O. Meir, O. Weinstein, A. Wigderson. *Toward better formula lower bounds: an information complexity approach to the KRW composition conjecture.* STOC 2014.

[8] T.Kaced, A.E. Romashchenko, N.K.Vereshchagin, *A Conditional Information Inequality and Its Combinatorial Applications*. IEEE Trans. Information Theory, 2018.

[9] E. Nisan, N. Kushilevitz. *Communication complexity*, 1997.

[10] A. Rao. *Notes for CSE533: Information Theory in Computer Science*, 2010. https://homes.cs.washington.edu/~anuprao/pubs/CSE533Autumn2010/