Конспект по теории информации
IV семестр, 2021 год
Современное программирование, факультет математики и компьютерных наук, СПбГУ
(лекции Дмитрия Соколова)

Тамарин Вячеслав

April 8, 2021

Contents

Исходный код на https://github.com/tamarinvs19/theory_university

Index

Контакты: sokolov.dmt@gmail.com,

Видимо будет письменный экзамен.

Есть прошлогодний конспект, там есть существенные ошибки, плюс курс немного отличается.

Chapter 1

Введение

1.1 Информация по Хартли

Пусть у нас есть конечное множество объектов A. Выдернем какой-то элемент.

Мы хотим придумать описание этого элемента, которое будет отличать его от всех остальных.

Самый простой вариант — число битов требуемое для записи объекта.

Свойства, которые мы хотим получить от меры $\chi(A)$:

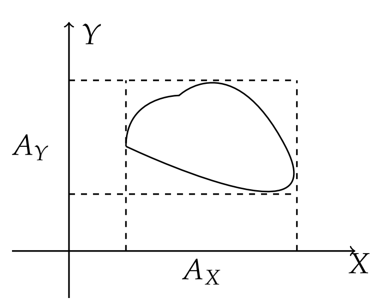
1. χ дает нам оценку на длину описаний
2. $\chi(A \cap B) \leqslant \chi(A) + \chi(B)$
3. Если наше множество $A \coloneqq B \times C$, то можно описать для B и для C, поэтому можно ограничить:
$$\chi(A) \leqslant \chi(B) + \chi(C).$$

Определение 1: Информация по Хартли
$\chi(A) \coloneqq \log|A|$

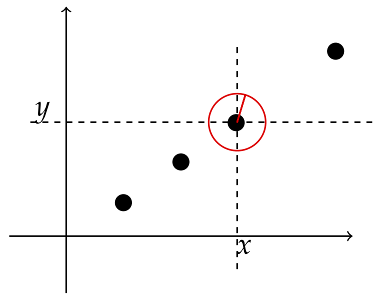
Замечание. Очевидно, второе свойство выполнено для такого определения. В третьем даже равенство.

Описание — например, битовая строка. Если логарифм нецелый, округляем вверх.

Пусть $A \subset X \times Y$. Обозначим проекции A_X и A_Y. Здесь



1. $\chi(A) \geqslant 0$
2. $\chi(A_X) \leqslant \chi(A)$
3. $\chi(A) \leqslant \chi(A_X) + \chi(A_Y)$



Рассмотрим такой пример: здесь, зная первую координату, можно сразу понять вторую.

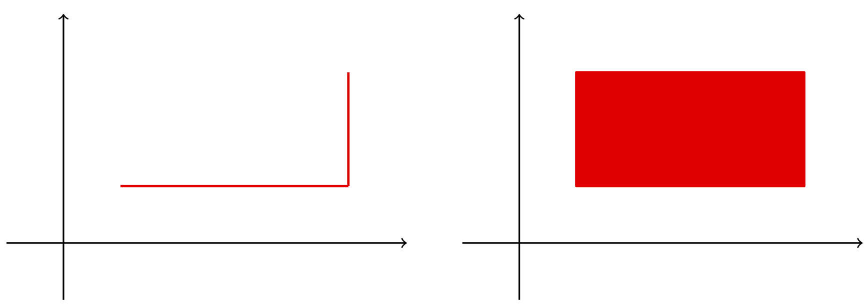
Попробуем усилить третье свойство:

3'. $\chi(A) \leqslant \chi(A_X) + \chi_{Y|X}(A)$, где $\chi_{Y|X}(A)$ — описание Y при условии X.

Как будем определять $\chi_{Y|X}(A)$? Можно взять $\max_{x \in X} \log(A(x))$.

Теперь для диагонального множество $\chi_{Y|X}$ просто обнуляется и неравенство переходит в равенство.

Но если взять такие множества. Во-первых, на первой картинке передав x столбца придется передавать



и y тоже. Во-вторых, мы не сможем отличить эти множества.

Упражнение. Пусть $A \subset X \times Y \times Z$. Доказать

$$2\chi(A) \leqslant \chi(A_{XY}) + \chi(A_{XZ}) + \chi(A_{YZ}).$$

1.1.1 Применение информации

Обозначим $[n] \coloneqq \{1, \ldots, n\}$. Первый игрок выбирает одно число, а второй должен угадывать.

Если два варианта игры:

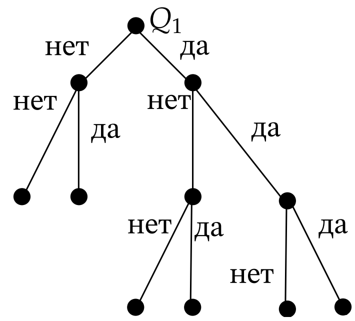
- Адаптивная — ответ сразу
- Сначала пишем все запросы, а потом получаем все ответы.

Очевидно, что нам потребуется не менее логарифма запросов: нарисуем дерево, где вершины – запросы, по двум ребрам можно перейти в зависимости от ответа. Листья должны содержать $[n]$, поэтому глубина дерева не менее логарифма.

Теперь подумаем с точки зрения теории информации. Пусть $B \coloneqq Q_1 \times \ldots \times Q_h$, h — число запросов, Q_i — ответ на запрос по некоторому протоколу. Хотим минимизировать h.

Рассмотрим $([n], B)$ — все возможные пары —. Нас интересует множество $A \subseteq ([n], B)$ — соответствует некоторым корректным запросам, здесь записаны ответы нашего протокола.

$$A = \{(m, b) \mid b = (q_1, \ldots, q_h),\ m \text{ — согласовано с ответом}\}.$$



1. $\chi_{[n]|B}(A) = 0$. Ответы на запросы должны однозначно определять число m. Это свойство говорит о корректности протокола, то есть нам ничего не нужно, чтобы, зная ответы, получить m.
2. $\log n \leqslant \chi(A)$, так как хотя бы столько мы запихнули. С другой стороны, $\chi(A) \leqslant \chi_B(A) + \chi_{[n]|B}(A) = 0$, а $\chi_B(A) \leqslant \chi(B) \leqslant \sum_{i=1}^{h} \chi(Q_i) = h$. Итого

$$\log n \leqslant h.$$

Другая формулировка

Пусть теперь за ответ «да» мы платим 1, а за «нет» 2. И мы хотим минимизировать не число запросов, а стоимость в худшем случае.

$$Q_i \overset{?}{\in} T_i.$$

Пусть A_i — множество возможных x (ответов) перед шагом i. В начале это все $[n]$, в конце – одно число.

$$A_i = \{a \in [n] \mid a \text{ согласовано с } Q_1, \dots Q_{i-1}\}.$$

Стратегия минимальной цены бита информации: берем такое T_i, что

$$2\big(\chi(A_i) - \chi(\underbrace{T_i}_{A_{i+1}})\big) = \chi(A_i) - \chi(\underbrace{A \setminus T_i}_{A_{i+1}}).$$

Докажем, что эта стратегия оптимальна. То есть для любой другой стратегии найдется число, с которым мы заплатим больше.

Если заплатили 1, то перешли в $A_i \to T_i$. Если заплатили 2, то $A_i \to A_i \setminus T_i$. Заметим, что каждый раз мы заплатили за каждый бит одинаково.

Докажем оптимальность. Пусть второй игрок меняет число, чтобы мы заплатили как можно больше, причем он знает нашу стратегию.

Если в нашем неравенстве знак $\geqslant$, он будет направлять на по «нет», а при $\leqslant$ «да», за счет чего каждый бит он будет отдавать по цене большей, чем, если бы мы действовали в точности по стратегии.

Следовательно, любая другая стратегия будет требовать большего вклада.

Можем решить уравнение на T_i, должно получиться:

$$\Phi(|T_i|) = |A_i|, \quad \Phi \text{ — золотое сечение.}$$

Упражнение (Задача про взвашивания монеток). Есть n монеток и рычажные весы. Хотим найти фальшивую (она одна).

1. Пусть $n = 30$ и весы показывают, что больше, что меньше. Теперь запрос приносит $\log 3$ информации, так как три ответа.

$$\log 30 \leqslant \sum_{i=1}^{h} \chi(a_i) \leqslant h \log 3.$$

2. $n = 15$, но мы не знаем относительный вес фальшивой монеты. В прошлом неравенстве можно заменить 30 на 29. Если в какой-то момент у нас было неравенство, можем в конце узнать не только номер, но и относительный вес, поэтому у нас 29 исходов.

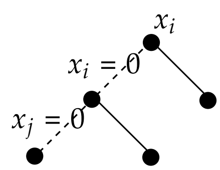
3. Вопрос: можно ли при $n = 14$? Нет.

1.1.2 Жизненные применения

Мы хотим решать задачу выполнимости.

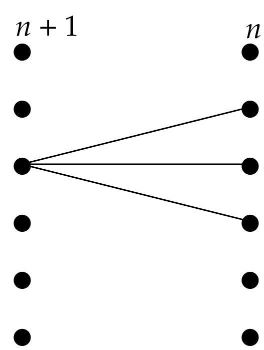
Вход: $\Phi = \wedge C_i$ — формула в КНФ.

Подставим $x_i = 0$. Если один из слозов нарушился, вернемся на шаг назад и подставим $x_j = 1$, а иначе подставляем дальше.



Это достаточно эффективный алгоритм, причем мы не ограничиваем выбор последовательности подстановок, порядок 0 и 1.

Рассадка голубей



Вопросы — сажаем ли мы голубя в клетку i?

Пусть один игрок загадал расстановку голубей, а второй хочет найти дизъюнкт, для которого нарушается эта расстановка.

1.2 Новая мера информации

На прошлой лекции поняли, что не всегда можем отличить некоторые множества.

Попробуем исправить данную ситуацию. Хотим понять состояния в Y, зная информацию об X. В среднем нам нужно сильно меньше информации, чем в крайнем случае.

Введем новую меру информации $\mu(\alpha)$, где α — распределение (множество и вероятности каждого элемента). Причем хотим, чтобы основные свойства были согласованы: [1]

1. $\mu(U_n) = \log n$;

[1] $\mu(x, y) = \mu((x, y))$

2. $\mu(\alpha) \geqslant 0$;
3. $\mu(\alpha, \beta) = \mu(\alpha) + \mu(\beta)$, если α и β независимы.

Если действовать как настоящие математики, можно переписать эти свойства в более общие:

1. $\mu(U_M) \geqslant \mu(U_{M'})$, если $|M| \geqslant |M'|$;
2. $\mu(\alpha, \beta) = \mu(\alpha) + \mu(\beta)$, если α и β независимы;
3. $\mu(B_p)$ непрерывно по $p \in [0, 1]$, где B_p — распределение для монетки, вероятность орла – p.
4. $\mu(B_p, \alpha) = \mu(B_p) + Pr[B_p = 0] \cdot \mu(\alpha \mid B_p = 0) + Pr[B_p = 1] \cdot \mu(\alpha \mid B_p = 1)$.

Определение 2: Энтропия

Этим аксиомам удовлетворяет примерно одна функция $\mu(\alpha) \coloneqq k \cdot H(\alpha)$, где $H(\alpha)$ — **энтропия**.[a]

$$H(\alpha) = \sum_{i=1}^{|\operatorname{supp}(\alpha)|} p_i \log \frac{1}{p_i}.$$

Энтропия обозначает среднее по распределению α необходимое количество информации для записи элемента.

[a] supp α — все возможные события, то есть имеющие ненулевую вероятность

Замечание. Энтропия равномерного распределения равна $\log n$, если $p_i = n$.

Замечание. Далее $H(p)$ обозначает энтропию для распределения монетки.

Теорема 1.2.1. $H(\alpha) \leqslant \log|\operatorname{supp}(\alpha)|$

□ Применим неравенство Йенсена

$$\sum_{i=1}^{|\operatorname{supp}(\alpha)|} p_i \log \frac{1}{p_i} \leqslant \log\left(\sum_i p_i \frac{1}{p_i}\right) = |\operatorname{supp}(\alpha)|$$

■

Теорема 1.2.2. $H(\alpha, \beta) \leqslant H(\alpha) + H(\beta)$

□

$$H(\alpha, \beta) = \sum_{i,j} p_{i,j} \log \frac{1}{p_{i,j}}$$

$$H(\alpha) + H(\beta) = \sum_i p_i \log \frac{1}{p_i} + \sum_j p_j \log \frac{1}{p_j}$$

Заметим, что $p_i = \sum_j p_{i,j}$ и $p_j = \sum_i p_{i,j}$.

$$H(\alpha, \beta) - H(\alpha) - H(\beta) = \sum_{i,j} p_{i,j} \log \frac{1}{p_{i,j}} - \sum_i p_i \log \frac{1}{p_i} + \sum_j p_j \log \frac{1}{p_j} = \sum_{i,j} p_{i,j} \log \frac{p_i p_j}{p_{i,j}}.$$

Если α и β независимы, то все логарифмы обнуляются. Иначе по неравенству Йенсена

$$\sum_{i,j} p_{i,j} \log \frac{p_i p_j}{p_{i,j}} \leqslant \log\left(\sum_{i,j} p_i p_j\right) = 0.$$

■

Определение 3: Условная энтропия

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

$$H(\alpha \mid \beta) = \mathbb{E}_{b=\beta} H(\alpha \mid p = b) = \sum_b H(\alpha \mid \beta = b) Pr[beta = b].$$

Свойства.

1. $\forall f\colon\ H(\alpha \mid \beta) \geqslant H(f(\alpha) \mid \beta)$
2. $H(\alpha, \beta) = H(\alpha) + H(\beta \mid \alpha)$
3. $H(\alpha) \geqslant H(\alpha \mid \beta)$
   $\square$
   $$H(\alpha \mid \beta) - H(\alpha) = \sum p_{i,j} \frac{1}{\log Pr[\alpha = i \mid \beta = j]} - \sum p_{i,j} \log \frac{1}{p_i} \leqslant \sum p_{i,j} \log \frac{p_i}{Pr[\alpha = i \mid \beta = j]}.$$
   По неравенству Йенсена полученное выражение меньше нуля. $\blacksquare$
4. $H(\alpha \mid \beta) \geqslant H(\alpha \mid \beta, \gamma)$

Попробуем решить задачу с монетками. Мы взвешиваем 14 монеток и хотим найти фальшивую за три взвешивания, причем неизвестен относительный вес. В нашем графе есть только один исход со всеми равенствами. Докажем, что нет такой стратегии.

Пусть нам дали текущее состояние и стратегия Сделаем так, чтобы каждый лист был равновероятен. Вернем с вероятностью $\frac{1}{27}$, что i фальшивая, и с $\frac{1}{27}$ – больше ($l > i$ фальшивая), и также с $l < i - \frac{1}{27}$.

При равномерном распределении энтропия $3 \log 3$.

Если стратегия верная, то

$$\begin{aligned} \log 27 &\leqslant H(\alpha, q_1, q_2, q_3) \leqslant \\ &\leqslant H(q_1) + H(q_2 \mid q_1) + H(q_3 \mid q_1, q_2) + H(\alpha \mid q_1, q_2, q_3) \leqslant \qquad \text{(Cain rule)} \\ &\leqslant H(q_1) + H(q_2) + H(q_3) + 0 \end{aligned}$$

Так как $H(q_i) \leqslant \log 3$, для все i выполнено равенство.

Чтобы было так, мы должны в каждый ход равновероятно получать все три ответа. Пусть мы взвешиваем кучки из k монет[2]. Вероятность равенства должна быть $\frac{2k}{27} \frac{1}{3}$, то есть $k \notin \mathbb{N}$. Противоречие.

1.3 Биномиальное распределение

$$\sum_{i=0}^{k} \binom{n}{i} \leqslant 2^{nH(\frac{k}{n})}.$$

Обозначим сумму за C.

[2]Очевидно, что взвешивать кучки разного размера, информацию извлечь не получиться даже по Хартли

Будем выбирать множество размера не больше k, а затем проверять, попало ли i наше множество. Пусть X — индикатор того, что i выбрали.

$$\begin{aligned}
\log C = H(X) &\leqslant H(X_1, \dots, X_n) \leqslant \\
&\leqslant \sum H(X_i \mid X_{<i}) \leqslant && \text{(Chain rule)} \\
&\leqslant \sum H(X_i) = nH(X_1) \leqslant && \text{(считаем, что } k \leqslant \tfrac{n}{2}\text{)} \\
&\leqslant nH\left(\frac{k}{n}\right)
\end{aligned}$$