# Энтропия Шеннона. Энтропийные профили. (ДЗ)

17 марта 2017 г.

1. Докажите эквивалентность определений:

   Пусть $\alpha, \beta, \gamma$ — случайные величины. Определим *взаимную информацию в $\alpha$ о $\beta$ при условии $\gamma$*.

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

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

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

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

2. Докажите следующее обобщение предыдущего неравенства. Пусть $T_1, ..., T_k$ — произвольные кортежи, составленные из переменных $\alpha_1, \dots \alpha_n$, причем каждая переменная входит ровно в $r$ кортежей. Тогда $rH(\alpha_1, \dots, \alpha_n) \leq H(T_1) + H(T_2) + \dots + H(T_k)$

3. Если величины $\alpha, \beta$ различаются с вероятностью $\epsilon$, и число различных значений величины $\alpha$ равно $a$, то выполнуется неравенство Фано:

$$H(\alpha \mid \beta) \leq \epsilon \log a + h(\epsilon, 1 - \epsilon),$$

   где $h(\epsilon, 1-\epsilon)$ — энтропия случайной величины с двумя значениями, имеющими вероятности $\epsilon, 1 - \epsilon$.

4. Пусть $\alpha \to \beta \to \gamma$ образуют Марковскую цепь, т.е. распределение $\langle \gamma \mid \beta \rangle = \langle \gamma \mid \alpha, \beta \rangle$. Докажите, что $I(\alpha : \gamma) \leq I(\alpha : \beta)$ и $I(\alpha : \gamma) \leq I(\beta : \gamma)$.

5. Пусть $\alpha \to \beta \to \gamma \to \delta$ образуют Марковскую цепь. Докажите, что $I(\alpha : \delta) \leq I(\beta : \gamma)$.

6. Если $a, x, y$ такие, что

$$\{ H\ (a \mid y, x) = 0, I(x : y \mid a) = 0.$$

то $H(a \mid x) + H(a \mid y) \leq H(a)$.

7. Первый игрок задумал число x от 1 до n. На своем ходу второй игрок задает ему вопросы вида: “верно ли, что $x \leq k$” выбирая конкретное k по своему усмотрению. После каждого такого хода первый игрок обязан ему ответить на этот вопрос утвердительно, если в самом деле задуманное число меньше или равно k, и может дать любой ответ (ДА/НЕТ) иначе. В конце игры второй игрок должен найти x при условии, что общее количество ошибочных ответов, данных первым игроком, не превысило сотую часть от общего количества заданных ему вопросов. Докажите, что второй игрок может справиться с этой задачей, задав не более $O(logn)$ вопросов. [Указание. Нужно запустить алгоритм бинарного поиска, модифицированный следующим образом. Каждая вершина дерева поиска соответствует некоторой гипотезе $x \in [a, b]$. При этом мы имеем гарантию, что $x > a$. На каждом шаге алгоритма мы сначала спрашиваем, верно ли, что $x < b$. Если ответ опровергают текущую гипотезу, то мы возвращаемся в дереве поиска к отцу текущей вершины. А иначе делим отрезок пополам и переходим к одному из сыновей текущей вершины, как в обычном би- нарном поиске (если мы уже находимся в некотором листе, то деление отрезка пополам пропускаем). Рассмотрим следующую “потенциаль- ную” функцию: (расстояние в дереве поиска от текущей вершины до цели) 2 (количество неправильных ответов, полученных к текущему моменту). Нетрудно видеть, что если мы находимся не в целевом листе (дерева поиска), то значение потенциальной функции убывает не менее чем на 1 в ходе выполнения одного шага алгоритма. В начале значение этой функции равно $\log n$, а в конце оно не меньше - 2(количество неправильных ответов) $>$ -0.02 (количество заданных вопросов). На каждом шаге мы задаем не более двух вопросов. Поэтому если мы ни разу не попали в целевой лист, то уменьшение целевой функции не меньше половины количества заданных вопросов. Это доказывает, что задав $\log n/0.48$ вопросов, мы хотя бы однажды попадем в целевой лист. Осталось заметить, что однажды попав в него, мы никогда из него не выйдем. Действительно, при $x = a = b$ на вопрос x $\leq b$ противник обязан отвечать положительно.]

8. (a) Докажите, что код Шеннона–Фано является префиксным.

(б) Докажите, что если центральный отрезок относить туда, куда попала его большая часть, то кодирование Шеннона–Фано не является сбалансированным (то есть не существует константы d, для которой выполнено $l(c_i) < -\log p_i + d$ для любых $k$ и любых исходных вероятностей $p_1, ..., p_k$).

(в) Докажите, что если центральный отрезок всегда относить к правой половине, то кодирование Шеннона–Фано также не является сбалансированным

9. Докажите, что кодирование Хаффмана не является сбалансированным.

Минимальное количество задач для зачета по домашнему заданию — 5. Большее число решенных задач увеличивает вероятность получения зачета. Ранее оформленные задачи переоформлять не надо.