# Теория информации

## Лекция 2

## Сводка определений

Определим *ансамбль* $X$ как тройку $(x, A_x, P_x)$, в которой $x$ - значение случайной величины из фиксированного множества элементов $A_x=\{a_1, a_2, \ldots, a_I\}$, которым соответствует совокупность вероятностей $P_x=\{p_1, p_2, \ldots, p_I\}$, так что $P(x=a_i)=p_i$, $p_i \geq 0$ и $\sum_{a_i \in A_x} P(x=a_i) = 1$. Множество $A_x$ еще называют *алфавитом*. Естественный пример это множество букв алфавита какого-нибудь языка вместе с пробелом и знаками препинания. Тогда в качестве $P_x$ можно взять вероятности, с которыми встречаются эти знаки в каком-нибудь тексте.

Пусть $T$ некоторое подмножество $A_x$ (например гласные в русском алфавите), тогда *вероятность подмножества* $T$ *в* $A_x$ определяется как

$$P(T) = P(x \in T) = \sum_{a_i \in T} P(x=a_i) \quad . \tag{2.1}$$

*Совместный ансамбль* $XY$ это ансамбль состоящий из упорядоченных пар $x, y$, таких что $x \in A_x=\{a_1, a_2, \ldots, a_I\}$ и $y \in A_y=\{b_1, b_2, \ldots, b_J\}$. Обозначим $P(x, y)$ совместную вероятность $x$ и $y$. В таком ансамбле не обязательно случайные величины независимы.

*Безусловная (маргинальная) вероятность* может быть получена из совместной вероятности суммированием:

$$P(x=a_i) = \sum_{y \in A_y} P(x=a_i, y) \tag{2.2}$$

и наоборот

$$P(y) = \sum_{x \in A_x} P(x, y) \quad . \tag{2.3}$$

*Условная вероятность* может быть выражена из совместной и безусловной как

$$P(x=a_i | y=b_j) = \frac{P(x=a_i, y=b_j)}{P(y=b_j)} \quad , \tag{2.4}$$

если конечно $P(y=b_j)\neq 0$ , в противном случае условная вероятность не определена.
*Правило умножения вероятностей*:

$$P(x,y) = P(x|y)P(y) = P(y|x)P(x) \quad . \tag{2.5}$$

*Правило суммирования вероятностей*:

$$P(x) = \sum_y P(x,y) = \sum_y P(x|y)P(y) \quad . \tag{2.6}$$

Из правого равенства в (2.5) можно получить *формулу Байеса*:

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)} = \frac{P(x|y)P(y)}{\sum_{y'} P(x|y')P(y')} \quad . \tag{2.7}$$

И наконец, величины $X$ и $Y$ статистически независимы, если

$$P(x,y) = P(x)P(y) \quad . \tag{2.8}$$

**Количество информации и энтропия**

Определение Шеннона для *количества информации*, связанной со случайной переменной $x$ имеет вид:

$$h(x) = \log_2 \frac{1}{P(x)} = -\log_2 P(x) \quad . \tag{2.9}$$

Эта величина измеряется в битах.

*Энтропия ансамбля* $X$ определяется как среднее содержание информации в реализациях:

$$H(X) = \sum_{x\in A_x} P(x)\log_2 \frac{1}{P(x)} = -\sum_{x\in A_x} P(x)\log_2 P(x) \quad , \tag{2.10}$$

причем действует соглашение, что при $P(x)=0$ , $0\cdot\log_2\frac{1}{0}=0$ (поскольку соответствующий предел сверху этому и равен). Другое название энтропии $X$ это неопределенность $X$ . Отметим некоторые *свойства энтропии*:

- $H(X)\geq 0$ , причем равенство достигается когда $p_i=1$ для одного какого-нибудь $i$ ;

- энтропия максимальна если распределение $\boldsymbol{p}$ равномерно и $H(X) \le \log_2(|A_x|)$ , причем равенство достигается тогда и только тогда, когда $p_i = 1/|A_x| \quad \forall \; i$ (здесь $|A_x|$ это число элементов в $A_x$ ).

*Избыточность* $X$ характеризуется величиной относительной разности между энтропией ансамбля $H(X)$ и ее максимально возможной величиной $\log_2|A_x|$ , т.е. величиной

$$1 - \frac{H(X)}{\log_2|A_x|} \quad . \tag{2.11}$$

*Совместная энтропия* двух ансамблей $X$ и $Y$ :

$$H(X,Y) = \sum_{xy \in A_x A_y} P(x,y) \log_2 \frac{1}{P(x,y)} \quad . \tag{2.12}$$

Энтропия (2.12) аддитивна для независимых величин:

$$H(X,Y) = H(X) + H(y) \quad \Leftrightarrow \quad P(x,y) = P(x)P(y) \quad . \tag{2.13}$$

**Разложимость энтропии**

Функция энтропии обладает рекурсивным свойством, которое может быть полезным при вычислении этой функции. Рассмотрим для примера случайную переменную $x \in \{0,1,2\}$ , которая определяется с помощью двух бросков монеты (не согнутой с одинаковыми вероятностями выпадения орла и решки). Первый бросок определяет $x=0$ или нет, если нет, то выполняется второй бросок, чтобы определить $x=1$ или $x=2$ . Очевидно, что распределение вероятностей следующее:

$$P(x=0) = \frac{1}{2}; \quad P(x=1) = \frac{1}{4}; \quad P(x=2) = \frac{1}{4} \quad .$$

Легко посчитать энтропию такого ансамбля (исходя из определения энтропии):

$$H(X) = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4 = 1.5 \quad .$$

Однако можно посчитать энтропию иначе. В качестве аргумента энтропии будем рассматривать вектор вероятностей. Рассмотрим только первый бросок, в котором определяется 0 или нет $x$. Вектор вероятностей для этого события

$\boldsymbol{p} = \{1/2, 1/2\}$ и энтропия в этом случае $H(1/2, 1/2) = 1/2 \log_2 2 + 1/2 \log_2 2 = 1$. Если первый бросок дал, что $x$ не 0, то придется выполнить второй бросок с вектором вероятностей также $\boldsymbol{p} = \{1/2, 1/2\}$. Однако этот бросок понадобится только в половине случаев, поэтому энтропия может быть записана в виде

$$H(X) = H(1/2, 1/2) + 1/2\, H(1/2, 1/2) .$$

Если обобщить это свойство на произвольный вектор вероятностей (распределение) $\boldsymbol{p} = \{p_1, p_2, \ldots, p_I\}$, то выглядеть в виде формулы оно будет следующим образом:

$$H(\boldsymbol{p}) = H(p_1, 1-p_1) + (1-p_1)\, H\left(\frac{p_2}{1-p_1}, \frac{p_3}{1-p_1}, \ldots, \frac{p_I}{1-p_1}\right) . \tag{2.14}$$

Дальнейшее обобщение для произвольного $m$:

$$\begin{aligned} H(\boldsymbol{p}) = {} & H\big((p_1 + p_2 + \ldots + p_m), (p_{m+1} + p_{m+2} \ldots + p_I)\big) + \\ & + (p_1 + \ldots + p_m)\, H\left(\frac{p_1}{p_1 + \ldots + p_m}, \ldots, \frac{p_m}{p_1 + \ldots + p_m}\right) + \\ & + (p_{m+1} + \ldots + p_I)\, H\left(\frac{p_{m+1}}{p_{m+1} + \ldots + p_I}, \ldots, \frac{p_I}{p_{m+1} + \ldots + p_I}\right) \end{aligned} . \tag{2.15}$$

Еще одной полезной характеристикой в теории информации является *относительная энтропия* или *расстояние Кульбака-Лейблера* между двумя распределениями вероятностей $P(x)$ и $Q(x)$ определенными на одном и том же алфавите $A_x$:

$$D_{KL}(P \| Q) = \sum_{x \in A_x} P(x) \log_2 \frac{P(x)}{Q(x)} . \tag{2.16}$$

Относительная энтропия удовлетворяет *неравенству Гиббса*:

$$D_{KL}(P \| Q) \geq 0 , \tag{2.17}$$

в котором равенство выполняется только если распределения $P$ и $Q$ совпадают. И хотя эта величина называется расстоянием, она не является в общем случае метрикой в пространстве распределений, поскольку $D_{KL}(P \| Q) \neq D_{KL}(Q \| P)$.