

Теория информации

Кабанов Александр Николаевич
к.ф.-м.н., доцент кафедры кибернетики

1. Энтропия и информация

Вероятностная схема

- Пусть $A = \{a_1, a_2, \dots, a_n, \dots\}$ – полная группа попарно несовместных событий (исходов), $p_i = p(a_i)$ – вероятность события a_i .

- Тогда $A = \begin{pmatrix} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}$ – вероятностная схема, $\sum_i p_i = 1$

Дискретная вероятностная схема

- Если множество A не более чем счётно, то вероятностная схема A называется дискретной.
- Иначе вероятностная схема A называется непрерывной.
- В этом случае вероятностное распределение на исходах задаётся с помощью плотности распределения вероятности $p(x)$.
- Если множество A конечно, то и схема называется конечной.
- В дальнейшем будем рассматривать конечные вероятностные схемы.

Количество информации по Хартли

1. Пусть сообщение T_1 , записанное в алфавите A_1 , $|A_1| = n_1$, имеет длину l_1 , а сообщение T_2 , записанное в алфавите A_2 , $|A_2| = n_2$, имеет длину l_2 . Тогда сообщения T_1 и T_2 несут одинаковое количество информации, если $n_1^{l_1} = n_2^{l_2}$.
2. Количество информации, содержащееся в сообщении, пропорционально его длине.
 - Количество информации в сообщении длины l , записанном в алфавите A мощностью n :

$$I = \log_2 n^l$$

Количество информации по Шеннону

1. Пустое сообщение не содержит информации.
2. Количество информации, содержащееся в сообщении, пропорционально его длине.
 - Пусть символы алфавита $A = \{a_1, a_2, \dots, a_n\}$ появляются в сообщениях с вероятностями p_1, p_2, \dots, p_n .
 - Количество информации в сообщении длины l , записанном в алфавите A :

$$I = l \cdot \left(- \sum_{i=1}^n p_i \log_2 p_i \right)$$

Энтропия

- Энтропия вероятностной схемы A:

$$H(A) = -\sum_{i=1}^n p_i \log_2 p_i = \sum_{i=1}^n p_i \log_2 \frac{1}{p_i}$$

- Т.о. энтропия – количество информации, приходящееся на один символ сообщения.
- Энтропия – количественная мера неопределенности в реализации вероятностной схемы.

Единицы измерения

- Основание логарифма определяет единицу измерения количества информации.
- Если основание логарифма 2, то информацию измеряют в двоичных единицах – битах.
- Если основание логарифма 10, то информацию измеряют в десятичных единицах – дитах.
- Если основание логарифма e , то информацию измеряют в натуральных единицах – натах.

Аксиомы Хинчина

- Энтропия конечной вероятностной схемы с точностью до постоянного множителя однозначно определяется системой аксиом:

1. $H(p_1, \dots, p_n)$ – ненулевая непрерывная функция, $0 \leq p_i \leq 1$, $\sum_{i=1}^n p_i = 1$.
2. $H(p_1, \dots, p_n)$ симметрична по переменным.
3. $H(p_1, \dots, p_n, 0) = H(p_1, \dots, p_n)$.
4. $H(q_{11}, \dots, q_{1m}, q_{21}, \dots, q_{2m}, \dots, q_{n1}, \dots, q_{nm}) = H(p_1, \dots, p_n) \sum_{i=1}^n p_i H\left(\frac{q_{i1}}{p_i}, \dots, \frac{q_{im}}{p_i}\right)$,
где $p_i = q_{i1} + \dots + q_{im}$.
5. $H(p_1, \dots, p_n) \leq H\left(\frac{1}{n}, \dots, \frac{1}{n}\right)$.

Аксиомы Фадеева

- Система аксиом Фадеева эквивалентна системе аксиом Хинчина:

1. $H(p, 1-p)$ – непрерывная и положительная хотя бы в одной точке функция, $0 \leq p_i \leq 1$,

2. $H(p_1, \dots, p_n)$ симметрична по переменным $p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right)$,

3. При $n \geq 2$ $H(p_1, \dots, p_{n-1}, q_1, q_2) = H(p_1, \dots, p_n) +$

где $p_n = q_1 + \dots + q_2$.

Минимальная энтропия

- Докажем, что $H(1,0) = 0$ из аксиом Хинчина.
- По аксиоме X3: $H(\frac{1}{2}, \frac{1}{2}, 0, 0) = H(\frac{1}{2}, \frac{1}{2})$.
- По аксиоме X2: $H(\frac{1}{2}, \frac{1}{2}, 0, 0) = H(\frac{1}{2}, 0, \frac{1}{2}, 0)$.
- По аксиоме X4: $H(\frac{1}{2}, 0, \frac{1}{2}, 0) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2}H(1,0) + \frac{1}{2}H(1,0) = H(\frac{1}{2}, \frac{1}{2}) + H(1,0)$.
- Следовательно, $H(\frac{1}{2}, \frac{1}{2}) = H(\frac{1}{2}, \frac{1}{2}) + H(1,0)$.
- Значит, $H(1,0) = 0$.

Минимальная энтропия

- Докажем, что $H(1,0) = 0$ из аксиом Фадеева.
- По аксиоме Ф2: $H(\frac{1}{2}, \frac{1}{2}, 0) = H(0, \frac{1}{2}, \frac{1}{2})$.
- По аксиоме Ф3: $H(0, \frac{1}{2}, \frac{1}{2}) = H(0, 1) + 1 \cdot H(\frac{1}{2}, \frac{1}{2}) = H(1, 0) + H(\frac{1}{2}, \frac{1}{2})$.
- По аксиоме Ф3: $H(\frac{1}{2}, \frac{1}{2}, 0) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2}H(1, 0)$.
- Следовательно, $H(1, 0) + H(\frac{1}{2}, \frac{1}{2}) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2}H(1, 0)$.
- Значит, $H(1, 0) = \frac{1}{2}H(1, 0)$.
- Значит, $\frac{1}{2}H(1, 0) = 0$.
- Значит, $H(1, 0) = 0$.

Объединённая вероятностная схема

- Рассмотрим вероятностные схемы $A = \begin{pmatrix} a_1 & \dots & a_n \\ p_1 & \dots & p_n \end{pmatrix}$, $\sum_{i=1}^n p_i$, и

$$B = \begin{pmatrix} b_1 & \dots & b_m \\ q_1 & \dots & q_m \end{pmatrix}, \sum_{j=1}^m q_j.$$

- Вероятностная схема

$$C = AB = \begin{pmatrix} a_1 b_1 & \dots & a_1 b_m & a_2 b_1 & \dots & a_2 b_m & \dots & a_n b_m \\ p_{11} & \dots & p_{1m} & p_{21} & \dots & p_{2m} & \dots & p_{nm} \end{pmatrix}$$

называется объединённой вероятностной схемой.

Объединённая вероятностная схема

- Для вероятностей p_{ij} выполняются следующие соотношения:

$$\sum_{i=1}^n \sum_{j=1}^m p_{ij} = 1, \quad \sum_{i=1}^n p_{ij} = q_j, \quad \sum_{j=1}^m p_{ij} = p_i.$$

- Энтропией объединённой вероятностной схемы АВ называется величина

$$H(AB) = - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log_2 p_{ij}$$

Объединённая вероятностная схема

- Преобразуем выражение $H(AB)$

$$\begin{aligned} H(AB) &= -\sum_{i=1}^n \sum_{j=1}^m p_{ij} \log_2 p_{ij} = -\sum_{i=1}^n \sum_{j=1}^m p_{ij} \log_2 p(a_i b_j) = \\ &= -\sum_{i=1}^n \sum_{j=1}^m p_{ij} \log_2 p(a_i) p(b_j | a_i) = \\ &= -\sum_{i=1}^n \sum_{j=1}^m p_{ij} \left(\log_2 p(a_i) + \log_2 p(b_j | a_i) \right) = \end{aligned}$$

Объединённая вероятностная схема

$$\begin{aligned} & - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log_2 p(a_i) - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log_2 p(b_j | a_i) = \\ & = - \sum_{i=1}^n p_i \log_2 p_i - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log_2 p(b_j | a_i) = \\ & = H(A) - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log_2 p(b_j | a_i) \end{aligned}$$

Условная энтропия

$$H(B | A) = - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log_2 p(b_j | a_i)$$

- Эта величина называется условной энтропией вероятностной схемы B относительно вероятностной схемы A.
- Т.о. $H(AB) = H(A) + H(B | A)$.
- Аналогично, $H(AB) = H(B) + H(A | B)$.

$$H(A | B) = - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log_2 p(a_i | b_j)$$

Условная энтропия

$$H(B | a_i) = - \sum_{j=1}^m p(b_j | a_i) \log_2 p(b_j | a_i)$$

- Эта величина называется условной энтропией вероятностной схемы B относительно исхода a_i .
- Имеет место следующее соотношение:

$$\begin{aligned} H(B | A) &= - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log_2 p(b_j | a_i) = \\ &= - \sum_{i=1}^n \sum_{j=1}^m p(a_i) p(b_j | a_i) \log_2 p(b_j | a_i) = \end{aligned}$$

Условная энтропия

$$= \sum_{i=1}^n p(a_i) \left(- \sum_{j=1}^m p(b_j | a_i) \log_2 p(b_j | a_i) \right) =$$
$$= \sum_{i=1}^n p_i H(B | a_i)$$

- Поэтому $H(B | A)$ называют средней условной энтропией.

Условная энтропия

- Пусть A и B – независимые вероятностные схемы. Тогда:

$$\begin{aligned} H(B | A) &= - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log_2 p(b_j | a_i) = \\ &= - \sum_{i=1}^n \sum_{j=1}^m p_i q_j \log_2 q_j = - \sum_{i=1}^n p_i \sum_{j=1}^m q_j \log_2 q_j = \\ &= - \sum_{j=1}^m q_j \log_2 q_j = H(B) \end{aligned}$$

Энтропия объединённой и составляющих схем

- **Теорема:** Для любых двух конечных вероятностных схем справедливо неравенство:

$$H(AB) \leq H(A) + H(B).$$

Равенство имеет место тогда и только тогда, когда схемы A и B независимы.

- **Следствие 1:** $H(B|A) \leq H(B)$.
- **Следствие 2:** $H(A_1 \dots A_k) \leq H(A_1) + \dots + H(A_k)$.
- **Следствие 3:** $H(A|BC) \leq H(A|B)$.

Взаимная информация

- Рассмотрим меру изменения количества информации, содержащейся в исходе a_i из A , при условии, что реализовался исход b_j из B .

$$I(a_i, b_j) = \log_2 \frac{p(a_i | b_j)}{p(a_i)}$$

- Это величина называется взаимной информацией исходов a_i и b_j .

Взаимная информация

$$\begin{aligned} I(a_i, b_j) &= \log_2 \frac{p(a_i|b_j)}{p(a_i)} = \log_2 \frac{p(a_i|b_j)p(b_j)}{p(a_i)p(b_j)} = \\ &= \log_2 \frac{p(a_i b_j)}{p(a_i)p(b_j)} = \log_2 \frac{p(a_i)p(b_j|a_i)}{p(a_i)p(b_j)} = \\ &= \log_2 \frac{p(b_j|a_i)}{p(b_j)} = I(b_j, a_i) \end{aligned}$$

Взаимная информация

- Составим закон распределения случайной величины:

$$\begin{pmatrix} I(a_1, b_1) & \dots & I(a_i, b_j) & \dots & I(a_n, b_m) \\ p_{11} & \dots & p_{ij} & \dots & p_{nm} \end{pmatrix}$$

- Вычислим математическое ожидание:

$$I(A, B) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} I(a_i, b_j)$$

- Эта величина называется средней взаимной информацией вероятностных схем A и B.

Взаимная информация

$$I(A, B) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} I(a_i, b_j) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} I(b_j, a_i) = I(B, A)$$

- Т.о. $I(A, B) = I(B, A)$.

Собственная информация

$$I(a_i) = \log_2 \frac{1}{p_i}$$

- Эта величина называется собственной информацией, содержащейся в исходе a_i .
- Собственную информацию, содержащуюся в исходе a_i , интерпретируют как неопределённость события a_i или как количество информации, необходимое для разрешения этой неопределённости.

Собственная информация

- Составим закон распределения случайной величины:

$$\begin{pmatrix} I(a_1) & \dots & I(a_i) & \dots & I(a_n) \\ p_1 & \dots & p_i & \dots & p_n \end{pmatrix}$$

- Вычислим математическое ожидание:

$$I(A) = \sum_{i=1}^n p_i I(a_i)$$

- Эта величина называется средней собственной информацией вероятностной схемы A .

Собственная информация

$$I(A) = \sum_{i=1}^n p_i I(a_i) = \sum_{i=1}^n p_i \log_2 \frac{1}{p_i} = - \sum_{i=1}^n p_i \log_2 p_i = H(A)$$

- Т.о. $I(A) = H(A)$.

Условная собственная информация

$$I(a_i|b_j) = \log_2 \frac{1}{p(a_i|b_j)}$$

- Эта величина называется условной собственной информацией, содержащейся в исходе a_i при условии реализации исхода b_j .

Условная собственная информация

- Составим закон распределения случайной величины:

$$\begin{pmatrix} I(a_1|b_1) & \dots & I(a_i|b_j) & \dots & I(a_n|b_m) \\ p_{11} & \dots & p_{ij} & \dots & p_{nm} \end{pmatrix}$$

- Вычислим математическое ожидание:

$$I(A|B) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} I(a_i|b_j)$$

- Эта величина называется средней условной собственной информацией вероятностной схемы A относительно вероятностной схемы B.

Условная собственная информация

$$\begin{aligned} I(A|B) &= \sum_{i=1}^n p_{ij} I(a_i | b_j) = \\ &= \sum_{i=1}^n p_{ij} \log_2 \frac{1}{p(a_i | b_j)} = - \sum_{i=1}^n p_{ij} \log_2 p(a_i | b_j) = H(A|B) \end{aligned}$$

- Таким образом, средняя условная собственная информация равна условной энтропии.

Связь между видами информации

$$\begin{aligned} I(a_i, b_j) &= \log_2 \frac{p(a_i|b_j)}{p(a_i)} = \log_2 p(a_i|b_j) - \log_2 p(a_i) = \\ &= -\log_2 \frac{1}{p(a_i|b_j)} + \log_2 \frac{1}{p(a_i)} = I(a_i) - I(a_i|b_j) \end{aligned}$$

Аналогично, $I(a_i, b_j) = I(b_j) - I(b_j|a_i)$.

Связь между видами информации

- Рассмотрим собственную информацию, содержащуюся в совместном исходе a_i, b_j .

$$\begin{aligned} I(a_i, b_j) &= \log_2 \frac{1}{p(a_i, b_j)} = \log_2 \frac{1}{p(a_i)p(b_j|a_i)} = \\ &= \log_2 \frac{1}{p(a_i)} + \log_2 \frac{1}{p(b_j|a_i)} = I(a_i) + I(b_j|a_i) = \\ &= I(a_i) + I(b_j) - I(a_i, b_j) \end{aligned}$$

Связь между видами информации

- Таким образом, $I(a_i, b_j) = I(a_i) + I(b_j) - I(a_i, b_j)$
- Усредняя это выражение, получаем:
- $H(A, B) = H(A) + H(B) - I(A, B)$
- Отсюда: $I(A, B) = H(A) + H(B) - H(A, B)$
- $I(A, B) = H(A) + H(B) - H(A) - H(B|A) = H(B) - H(B|A)$
- Так как $H(B|A) \leq H(B)$, следовательно $I(A, B) \geq 0$.

Свойство средней взаимной информации

- **Теорема:** Пусть A , B и C – вероятностные схемы, $\phi: A \rightarrow C$ – сюръективное отображение. Тогда $I(A, B) \geq I(C, B)$.
- Таким образом, средняя взаимная информация не увеличивается при преобразовании схем.
- Равенство имеет место в том случае, если ϕ – биекция.

Непрерывные вероятностные схемы

- Пусть A – непрерывная вероятностная схема. Тогда вероятностное распределение задается функцией плотности распределения вероятностей.
- Тогда энтропия схемы A :

$$H(A) = \int_{-\infty}^{\infty} p(x) \log_2 p(x) dx$$

Непрерывные вероятностные схемы

- Пусть V – непрерывная вероятностная схема с плотностью распределения $q(y)$.
- Для объединенной вероятностной схемы AB существует совместная плотность распределения вероятностей $p(x,y)$.
- Энтропия объединенной непрерывной вероятностной схемы:

$$H(AB) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log_2 p(x, y) dx dy$$

Непрерывные вероятностные схемы

- Частные распределения:

$$p(x) = \int_{-\infty}^{\infty} p(x, y) dy, \quad q(y) = \int_{-\infty}^{\infty} p(x, y) dx$$

- Условные распределения

$$p(x|y) = \frac{p(xy)}{q(y)}, \quad q(y|x) = \frac{p(xy)}{p(x)}$$

Ёмкость

- Максимальная энтропия системы равна

$$H_{\max}(A) = -\sum_{i=1}^n \frac{1}{n} \log_2 \frac{1}{n} = n \cdot \frac{1}{n} \log_2 n = \log_2 n$$

- Таким образом при равновероятных выборах формула энтропии преобразуется в формулу Хартли.
- $H_{\max}(A)$ называется ёмкостью системы A.

Избыточность

- Абсолютной избыточностью называется величина

$$D = H_{\max}(A) - H(A)$$

- Относительной избыточностью называется величина

$$D_{\text{отн}} = \frac{H_{\max}(A) - H(A)}{H_{\max}(A)}$$

- Относительная избыточность показывает, насколько рационально применяются символы данного алфавита.

Дискретный источник сообщений

- Под дискретным источником сообщений будем понимать устройство, порождающее последовательности, составленные из букв конечного алфавита $A = \{a_1, a_2, \dots, a_n\}$. При этом буквы последовательностей порождаются в дискретные моменты времени.
- Любой непрерывный источник информации можно заменять с заданной степенью точности некоторым дискретным источником.

Дискретный источник сообщений

- Пусть $c_{t_1 t_2 \dots t_m}(a_{i_1} a_{i_2} \dots a_{i_m})$ – событие, заключающееся в том, что в момент времени t_j источник порождает символ a_{ij} .
- Для математического описания источника кроме алфавита нужно также знать распределение вероятностей $p(c)$ для событий указанного выше вида.

Стационарный источник сообщений

- Источник сообщений называется стационарным, если $p(c_{t_1 t_2 \dots t_m} (a_{i_1} a_{i_2} \dots a_{i_m})) = p(c_{t_1+j t_2+j \dots t_m+j} (a_{i_1} a_{i_2} \dots a_{i_m}))$ для любого j .
- Другими словами, источник стационарный, если вероятность того, что источник порождает некоторую последовательность, составленную из букв алфавита A , в моменты времени $1, 2, \dots, m$, равна вероятности того, что в точности такая же последовательность порождается в моменты времени $1+k, 2+k, \dots, m+k$ для любого k .
- Таким образом, стационарность означает неизменность во времени всех конечномерных распределений.

Энтропия источника

- Для стационарного источника множество всех событий $C_{t_1 t_2 \dots t_m}(a_{i_1} a_{i_2} \dots a_{i_m})$ можно рассматривать как вероятностную схему, событиями которой являются всевозможные наборы символов длины m .

- Для такой вероятностной схемы можно вычислить энтропию:

$$H(C_m) = - \sum p(a_{i_1} a_{i_2} \dots a_{i_m}) \log_2 p(a_{i_1} a_{i_2} \dots a_{i_m}).$$

- Здесь сумма берется по всевозможным наборам символов алфавита A длины m .

Энтропия источника

- В среднем на одну порождаемую источником букву приходится количество информации, равное $H(C_m)/m$.
- **Теорема:** Для любого стационарного источника сообщений существует предел

$$\lim_{m \rightarrow \infty} \frac{H(C_m)}{m}$$

- Эта величина называется энтропией источника.

$$H_\infty = \lim_{m \rightarrow \infty} \frac{H(C_m)}{m}$$

Энтропия источника

- Пусть источник порождает последовательности по схеме независимых испытаний.
- Тогда $H(C_m) = mH(C_1) = mH(A)$.
- Следовательно, $H_\infty = H(A)$.
- Такие источники, для которых буквы, порожденные в предыдущие моменты времени, не влияют на буквы, порождаемые в последующие моменты времени, называются источниками без памяти.

Первая теорема Шеннона

- **Первая теорема Шеннона:** Рассмотрим источник без памяти, имеющий энтропию H_∞ . Для любых чисел $\varepsilon > 0$ и $\eta > 0$ существует число k_0 такое, что при $k > k_0$ все реализации источника длины k могут быть разбиты на 2 класса: $C_k = C'_k \cup C''_k$. Причем для любой последовательности c'_k из класса C'_k имеет место условие:

$$\left| \frac{1}{k} \log_2 \frac{1}{p(c'_k)} - H_\infty \right| < \eta,$$

а суммарная вероятность последовательностей из класса C''_k меньше ε .

Первая теорема Шеннона

- **Следствие 1:** Вероятность $p(c'_k)$ можно оценить следующим образом:

$$2^{-k(H_\infty + \eta)} < p(c'_k) < 2^{-k(H_\infty - \eta)}.$$

- **Следствие 2:** Суммарная вероятность последовательностей из класса C'_k не менее $1 - \varepsilon$.

Вторая теорема Шеннона

- **Вторая теорема Шеннона:** Упорядочим все последовательности длины k , полученные на источнике без памяти, по убыванию их вероятностей. Выберем некоторое произвольное число $0 < \alpha < 1$. Будем отбирать наиболее вероятные последовательности, пока их суммарная вероятность не окажется такой, что добавление к этой сумме вероятности реализации следующей последовательности из упорядоченного списка сделает эту сумму больше α . Множество отобранных таким образом высоковероятностных последовательностей обозначим $M_k(\alpha)$. Имеет место следующее условие:
$$\lim_{k \rightarrow \infty} \frac{\log_2 |M_k(\alpha)|}{k} = H_\infty$$

Марковский источник

- Дискретный стационарный источник называется марковским источником порядка m , если для любого $k > m$ и любой последовательности $a_{i_1} a_{i_2} \dots a_{i_k}$ выполняется условие:

$$p(a_{i_k} | a_{i_1} a_{i_2} \dots a_{i_{k-1}}) = p(a_{i_k} | a_{i_{k-m}} a_{i_{k-m+1}} \dots a_{i_{k-1}})$$