

*

Измерение информации: вероятностный подход

Урок 9

Система основных понятий

Измерение информации – содержательный подход

Измеряется количество информации в сообщении о результате некоторого события

Равновероятные результаты:
никакой результат не имеет преимущества перед другими

Неопределенность знания – число возможных результатов (вариантов сообщения) - N

Количество информации в сообщении об одном результате события – i битов

Главная формула информатики: $2^i = N$

Частный случай: два равновероятных результата события

$N=2$

$i=1$ бит

1 бит – количество информации в сообщении об одном из двух равновероятных результатов некоторого события

Вопросы для повторения:

1. Что такое неопределенность знания о результате какого-либо события? Приведите примеры, когда неопределенность знания можно выразить количественно.
2. Как определяется единица измерения количества информации?
3. В каких случаях и по какой формуле можно вычислить количество информации, содержащейся в сообщении, используя содержательный подход?

Измерение информации

Содержательный подход

ИЗМЕРЕНИЕ

Алфавитный подход



ИНФОРМАЦИЯ

N

Число равновероятных возможных событий

i

Количество информации в сообщении о том, что произошло одно из N равновероятных событий

N

Число символов в алфавите (его размер) – **МОЩНОСТЬ АЛФАВИТА**

i

ИНФОРМАЦИОННЫЙ ВЕС СИМВОЛА
количество информации в одном символе

$$2^i = N$$

K

Число символов в символьном сообщении

I

Количество информации в символьном сообщении

$$I = K \times i$$

$$N = 256$$

$$i = 8 \text{ бит} = 1 \text{ байт}$$

$$N = 2$$

$$i = 1 \text{ бит}$$

1 байт

1 Кб

1 Мб

1 Гб

1024

1024

1024

Вероятностный подход к измерению информации

- в 1928 году американский инженер Ричард Хартли подметил закономерность и предложил меру для измерения количества информации:

$$I = \log_2 N$$

где N - количество *равновероятных* событий;

I - количество бит в сообщении о том, что любое из N событий произошло.

- Не все ситуации имеют одинаковые вероятности реализации. Существует много таких ситуаций, у которых вероятности реализации различаются. Например, если бросают несимметричный предмет
- Еще один бытовой пример – «правило бутерброда»

Вероятностный подход к измерению информации

В 1948 г. американский инженер и математик К. Шеннон предложил формулу для вычисления количества информации для событий с различными вероятностями.

p – вероятность события,

$p = K/N$, где K – количество благоприятных исходов

N – общее число исходов

$$I = \log_2 \left(\frac{1}{p} \right)$$

Формулу Хартли теперь можно рассматривать как частный случай формулы Шеннона.

При равновероятных событиях получаемое количество информации максимально.

Пример 1:

- В корзине лежит 8 черных шаров и 24 белых. Сколько информации несет сообщение о том, что достали **черный шар**?

$$K_{\text{ч}} = 8; K_{\text{б}} = 24;$$

$$N = K_{\text{ч}} + K_{\text{б}} = 8 + 24 = 32$$

$$P_{\text{черн}} = \frac{K_{\text{ч}}}{N} = \frac{8}{32} = \frac{1}{4};$$

$$I = \log_2 \frac{1}{1/4} = \log_2 4 = 2(\text{бита})$$

Пример 1:

- Для случая равновероятных событий
(если все шары разного цвета):

$$I = \log_2 N = \log_2 32 = 5(\text{бит})$$

Формула Шеннона

$$I = - \sum_{i=1}^N p_i \cdot \log_2 p_i$$

I - количество информации

N - количество возможных событий

p_i - вероятности отдельных событий

$$I = p_1 \cdot \log_2 \frac{1}{p_1} + p_2 \cdot \log_2 \frac{1}{p_2} + p_3 \cdot \log_2 \frac{1}{p_3} + \dots$$

Пример 2:

- В озере обитает 12500 окуней, 25000 пескарей, а карасей и щук по 6250. Сколько информации мы получим, когда поймаем какую-нибудь рыбу?

$$N = K_1 + K_2 + K_3 + K_4 = 12500 + 25000 + 2 * 6250 = 50000$$

$$p_1 = \frac{K_1}{N} = \frac{12500}{50000} = \frac{1}{4}; p_2 = \frac{K_2}{N} = \frac{25000}{50000} = \frac{1}{2};$$

$$p_3 = \frac{K_3}{N} = \frac{6250}{50000} = \frac{1}{8}; p_4 = \frac{K_4}{N} = \frac{6250}{50000} = \frac{1}{8}$$

$$I_1 = \log_2 \frac{1}{p_1} = \log_2 \frac{1}{1/4} = \log_2 4 = 2 \text{ бита}$$

$$I_2 = \log_2 \frac{1}{p_2} = \log_2 \frac{1}{1/2} = \log_2 2 = 1 \text{ бит}$$

$$I_3 = \log_2 \frac{1}{p_3} = \log_2 \frac{1}{1/8} = \log_2 8 = 3 \text{ бита}$$

$$I_4 = \log_2 \frac{1}{p_4} = \log_2 \frac{1}{1/8} = \log_2 8 = 3 \text{ бита}$$

$$I = p_1 \cdot \log_2 \frac{1}{p_1} + p_2 \cdot \log_2 \frac{1}{p_2} + p_3 \cdot \log_2 \frac{1}{p_3} + p_4 \cdot \log_2 \frac{1}{p_4}$$

$$I = \frac{1}{4} \cdot \log_2 \frac{1}{1/4} + \frac{1}{2} \cdot \log_2 \frac{1}{1/2} + \frac{1}{8} \cdot \log_2 \frac{1}{1/8} + \frac{1}{8} \cdot \log_2 \frac{1}{1/8} =$$

$$= \frac{1}{4} \cdot \log_2 4 + \frac{1}{2} \cdot \log_2 2 + \frac{1}{8} \cdot \log_2 8 + \frac{1}{8} \cdot \log_2 8 =$$

$$= \frac{1}{4} \cdot 2 + \frac{1}{2} \cdot 1 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = 1 \frac{3}{4} = 1,75(\text{бита})$$

Пример 3:

- Пусть имеется текст, содержащий 1000 букв. Буквы встречаются в тексте:
 - o – 90 раз
 - p – 40 раз
 - ф – 2 раза
 - a – 200 раз

Какое количество информации несет буква в строке?

Пример 3:

