

И-Л2-13.09.2025

#лекция

Меры информации

☰ Определение

Энтропия (информационная) - величина неопределенности неполноты знаний о чем-либо. Чем больше мы знаем о чем-либо, тем энтропия *ниже*, и, наоборот, чем меньше мы знаем о чем-либо, тем энтропия *выше*.

Энтропия характеризует *непредсказуемость* появления какого-либо символа алфавита.

Ральф Хартли - впервые внес понятие единицы измерения информации.

☰ Определение

В современных вычислительных системах одновременно обрабатываются 8 бит информации, называемые **байтом**. Байт может принимать одно из 256 (2^8) значений.

Производные от байта:

- КилоБайт - 10^3 байт
- КибиБайт - 1024 байт

Производные от байта десятичная система измерения называется "килобайт", "гигабайт"...

Производные от байта двоичной системы измерения называются

"кибайт", "гибайт"...

Исторически сложилась такая ситуация, что с наименованием "байт" некорректно использовали и продолжают использовать приставки СИ.

Определение

Формула Хартли - логарифмическая мера информации, которая определяет количество информации, содержащиеся в сообщении.

$$I = K \log_2 N$$

I - количество информации в сообщении, бит.

K - количество символов в сообщении (длина сообщения)

N - мощность алфавита

Формула Хартли применима только к расчетам информационной двоичной энтропии при отсутствии информационных потерь (и только в случае равной вероятности появления каждого последующего символа в алфавите).

Основание логарифма определяет единицу *количества информации*. Если основание было 1 - "над"; 10 - "дит".

При отсутствии информационных потерь энтропия численно равна количеству информации на символ передаваемого сообщения.

Определение

Формула Шеннона

$$H = - \sum_i P_i \log_2 P_i$$

H - функция энтропии

P_i - вероятность i го события

$H(P_1 \dots P_n)$ - определена и непрерывна для всех $P_1 \dots P_n$, где

$P \in [0; 1]$ для всех $i = 1 \dots n$ и $P_1 + \dots + P_n$

Знак минус (-) в формуле Шеннона не означает, что количество информации в сообщении отрицательная величина. Объясняется это тем, что вероятность лежит в промежутке **[0; 1]**. Так как логарифм числа меньше 1 отрицателен.

Шеннон предположил, что прирост информации равен утраченной неопределенности. Следовательно, энтропия является разницей между информацией содержащейся в сообщении и той частью информации, которая точно известна или хорошо предсказуема. Примером этого является избыточность языка: имеются явные статистические закономерности.