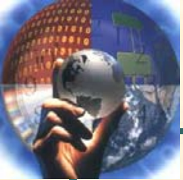
Введение в информатику

Измерение количества информации

1. Меры информации
2. Мера Хартли
3. Мера Шеннона
4. Приставки к единицам измерения информации



Понятие энтропии

Информация соответствует наличию порядка, организации в какой-либо области и отсутствию неопределенности. Противоположное понятие — **энтропия** (беспорядок).

Энтропия — мера неупорядоченности материальных систем.

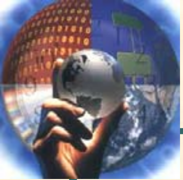
Увеличение информации ⇒ **Строгий порядок**

Полный хаос (неопределенность) ⇐ **Рост энтропии**

Количество информации о системе α в сообщении β

$$I_{\beta}(\alpha) = H(\alpha) - H_{\beta}(\alpha)$$

H(α) — мера неопределенности состояния системы (неосведомленности о системе) до получения сообщения β.



Меры информации

Количество информации — числовая величина, характеризующая информацию по разнообразию, сложности, упорядоченности, определенности и т.п.

За основную единицу измерения количества информации принят **1 бит**.

Для оценки состояния системы, которая может принимать одно из n возможных состояний, используется мера информации (события).

Мера — непрерывная действительная неотрицательная функция, определенная на множестве событий.

Мера Хартли используется для равновероятных событий.

Мера Шеннона используется для не равновероятных событий.

Мера Хартли является частным случаем меры Шеннона.

Мера количества информации по Хартли

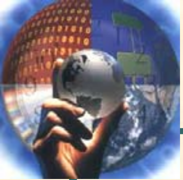
Мощность алфавита (h) — количество символов в алфавите.

$$I = L * \log_x h$$

L — длина сообщения

x — основание системы меры

Ральф Хартли (1880--1970)



Аналитическое определение бита

Пусть есть исходный алфавит {0,1}.

Надо определить количество информации, содержащейся в одной цифре.

Подставим данные в формулу Хартли

$$I = L*\log_x h,$$

где L — длина сообщения, x — основание системы меры.

Получим

$$I = 1*\log_2 2 = 1 \text{ (бит)}$$

Таким образом, на одну двоичную цифру приходится 1 бит информации.

Применение меры Хартли на практике

Пример 1. Ведущий загадывает число от 1 до 64. Какое количество вопросов типа «да-нет» понадобится, чтобы гарантировано угадать число?

Первый вопрос: «Загаданное число меньше 32?». Ответ: «Да».

Второй вопрос: «Загаданное число меньше 16?». Ответ: «Нет».

Нужно задать как можно меньше вопросов. После очередного ответа диапазон делится пополам.

Шестой вопрос (в худшем случае) точно приведет к верному ответу.

В соответствии с мерой Хартли в загадке содержится $\log_2 64 = 6$ бит информации.

Применение меры Хартли на практике-2

Пример 2. Ведущий держит за спиной ферзя и собирается поставить его на произвольную клетку доски. Насколько непредсказуемо его решение?

Всего на шахматной доске 64 клетки (8 х 8).

Цвет ферзя может быть белым или черным.

Всего возможно 8 х 8 х 2 = 128 равновероятных событий.

Количество информации по Хартли равно $\log_2 128 = 7$ бит.

Вероятность события

Вероятность — это количественная характеристика одного из исходов некоторого опыта, известная до его проведения.

Измеряется в пределах от 0 до 1.

$$0 \le p(A) \le 1,$$

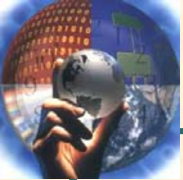
Классическое определение: существует только n равновозможных исходов эксперимента, из них m исходов приведут к событию А.

$$p(A) = m/n$$

Статистическое определение: в результате проведённых n экспериментов событие А возникло m раз.

$$p(A) = \lim_{n \to \infty} \frac{m}{n}$$

Сумма вероятностей всех возможных несовместных событий равна 1.



Мера количества информации по Шеннону

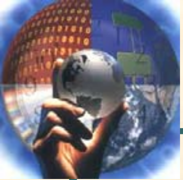
$$i(S) = -\sum_{i=1}^{N} p_i \cdot \log_2 p_i ,$$

где N – число состояний системы,

p_i – вероятность того, что система S находится в состоянии i (сумма всех p_i равна 1).

Клод Шеннон
(1916--2001)

Сообщение о наступлении достоверно наступающего события несет в себе нулевую информацию.



Аналитическое определение бита

Мера Хартли подходит только для систем с равновероятными состояниями.

Если состояния системы S не равновероятны, то используют меру Шеннона.

Если у опыта 2 равновероятных исхода, то по формуле Шеннона получим

$$i(S) = -\sum_{i=1}^{N} p_i \cdot \log_2 p_i \,,$$

$$\mathbf{i(S) = -(0{,}5 * \log_2 0{,}5 + 0{,}5 * \log_2 0{,}5) = 1\ (бит)}$$

Формула Хартли является частным случаем формулы Шеннона!

Применение меры Шеннона на практике

Пусть по результатам некоторого опыта получено n сообщений с вероятностью p_i. Тогда количество информации в i-том сообщении определяется по формуле: $\mathbf{I = -\log_2 p_i}$.

Пример. Определить количество информации в сообщении о результатах сдачи экзамена студентом, если известны вероятности получения оценок: «отлично» — p(5) = 0,1; «хорошо» — p(4) = 0,2; «удовлетворительно» — p(3) = 0,3; «неудовлетворительно» — p(2) = 0,4.

Количество информации в каждом сообщении

$I(5) = -\log_2 0{,}1 = \mathbf{3{,}32}$

$I(4) = -\log_2 0{,}2 = 2{,}32$

$I(3) = -\log_2 0{,}3 = 1{,}74$

$I(2) = -\log_2 0{,}4 = 1{,}32$

Сообщение о наступлении события с меньшей вероятностью несет в себе больше информации, чем сообщение о наступлении события с большей вероятностью.

Приставки к единицам измерения информации

Приставки единиц СИ	Новые двоичные префиксы	Δ,%
килобайт (kB) = 10^3 байт	кибибайт (KiB, КиБ) = 2^{10} байт	2
мегабайт (MB) = 10^6 байт	мебибайт (MiB, МиБ) = 2^{20} байт	5
гигабайт (GB) = 10^9 байт	гибибайт (GiB, ГиБ) = 2^{30} байт	7
терабайт (TB) = 10^{12} байт	тебибайт (TiB, ТиБ) = 2^{40} байт	10

Краткое обозначение битов и байтов

b = bit = бит, B = Б = байт

1024 B = 1024 Б = 8192 b = 8192 бит = 8 Кибит = 1 КиБ = 1 KiB

Приставки к единицам измерения информации-2

Полное произношение названий приставок

3 КиБ = «три кибибайта» = «три килобинарных (kilobinary) байта».

7 Гибит = «семь гибибитов» = «семь гигабинарных (gigabinary) битов».

Практика использования приставок

Объем памяти (HDD, RAM, Cache): 512 KiB = 524 288 bytes.

Скорость передачи данных: 512 kbps = 512 000 bps = 512 000 бит/с.