Информационные технологии

Часть 3

Информация и энтропия

Меры количества информации

Мера	Учитывает	Требует наличия
Синтаксическая	**Способ представления информации**	**Только самой информации**
Семантическая	**Смысловое содержание информации**	**Информации и ее пользователя**
Прагматическая	**Потребительские свойства информации**	**Информации, пользователя и цели использования**

Синтаксическая мера информации

Минимальным количеством информации является **1 бит**.

Эту единицу количественной меры информации впервые предложил **Р. Хартли** в 1928 году

Количество информации по формуле Хартли:

$$i = \log_2 N$$

где **i** – количество информации (бит); **N** – количество элементарных сообщений.

Сам термин произошел от слияния слов **bi**nary digi**t** - двоичная цифра

Из формулы Хартли вытекает основное определение бита

Бит равен информации, которая передается при приеме одного из двух равновероятных сообщений.

Синтаксическая мера информации

Другие свойства бита

Бит - двоичная единица информации.

Математически **бит** отображается состоянием **1** или **0** одного разряда двоичной системы счисления.

При получении информации в **1 бит** неопределенность уменьшается в **2** раза

1 бит – это количество информации, которое можно получить при ответе на вопрос типа «да - нет»

Группа из 8 битов информации называется **байт**
Производные единицы информации:
килобайт (кбайт, кб), **мега**байт (Мбайт, Мб) **гига**байт (Гбайт, Гб) и т.д.
1 кб =1024 байта - 2^{10} (1024) байтов.
1 Мб = 1024 кбайта = 2^{20}(1024 х 1024) байтов.
1 Гб = 1024 Мбайта - 2^{30} (1024 х 1024 х 1024)байтов.

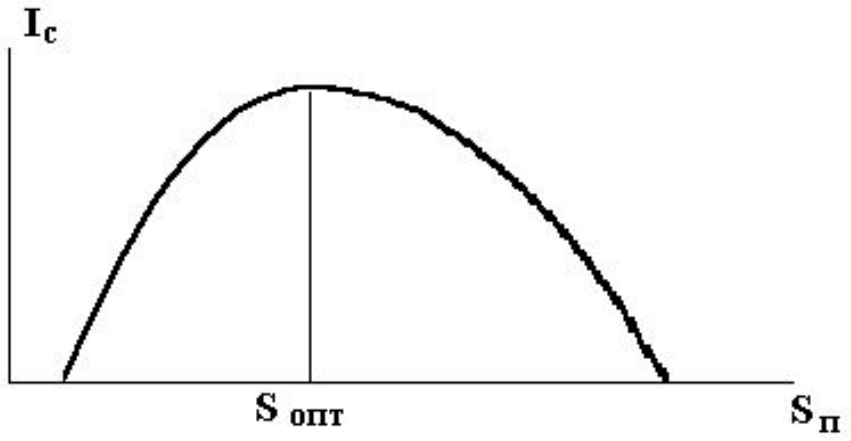
Семантическая мера информации

Если

$S_п$ – тезаурус пользователя;

I_c – количество семантической информации, то:

1. При $S_п \approx 0$ пользователь не понимает поступающую информацию и $I_c \approx 0$
2. При $S_п \to \infty$ пользователю уже известна поступающая информация и $I_c \approx 0$



Максимальное значение I_c приобретает при таком согласовании с тезаурусом пользователя $S_п$, когда поступающая информация с одной стороны понятна пользователю, а с другой, несет ранее неизвестные ему сведения

Прагматическая мера информации

Прагматическое количество информации определяется приростом экономического эффекта функционирования системы, где эта информация используется

$$I_{П}β(α) = П(α/β) - П(α)$$

Где
$I_{П}β(α)$ - ценность информационного сообщения **β** для системы управления **α**;
П(α) - ожидаемый экономический эффект функционирования системы управления **α**;
П(α/β) - ожидаемый экономический эффект функционирования системы управления **α**, при условии, что для управления будет использована информация, содержащаяся в сообщении **β**.

В такой постановке единицей измерения ценности информации являются денежные единицы.

Четыре основных класса случайных процессов

По признакам, связанным с пространством состояний и с параметром времени выделяют следующие классы случайных процессов

Дискретный случайный процесс с дискретным временем

Случайный процесс с дискретным временем

Дискретный случайный процесс с непрерывным временем

Случайный процесс общего типа

Неопределенность и количество информации

Получение какой-либо информации всегда связано с изменением **степени неосведомленности** получателя

Пусть мера неосведомленности равна $H(\alpha)$.
После получения некоторого сообщения β неосведомленность уменьшилась и стала $H(\alpha/\beta)$.
Тогда количество информации $I(\beta)$ в сообщении β, определится так:

$$I(\beta) = H(\alpha) - H(\alpha/\beta)$$

Количество полученной информации измеряется уменьшением неопределенности состояния системы

Основные положения теории К. Шеннона

1. Источник информации, порождающий сообщения, должен обладать тем или иным алфавитом, который должен быть заранее известен приемнику информации

2. Синтаксическое количество информации зависит, по определению, только от статистических свойств появления элементов алфавита в сообщении

3. В качестве меры неопределенности следует использовать величину равную логарифму вероятности, взятому с противоположным знаком

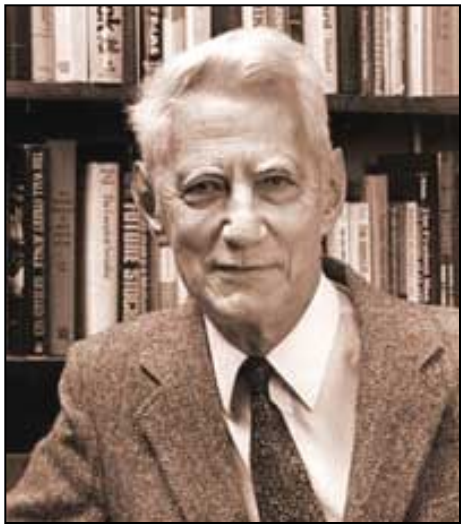
Этот показатель получил название

информационная энтропия

Клод Элвуд Шеннон

Claude Elwood Shannon

(30.04.1916 - 24.02.2001)



В 1936 году выпускник Мичиганского университета Клод Шеннон, которому было тогда 21 год, имея два диплома бакалавра - по электротехнике и по математике, сумел ликвидировать разрыв между алгебраической теорией логики и ее практическим приложением. Свои идеи относительно связи между двоичным исчислением, булевой алгеброй и электрическими схемами Шеннон развил в докторской диссертации, опубликованной в 1938 году

В 1948 году опубликовал фундаментальную работу A Mathematical Theory of Communication, в которой сформулированы основы теории информации.

Большую ценность представляет другая работа – Communication Theory of Secrecy Systems (1949), в которой сформулированы математические основы криптографии.

С 1956 – член Национальной академии наук США и Американской академии искусств и наук.

Клод Элвуд Шеннон – один из создателей математической теории информации, в значительной мере предопределил своими результатами развитие общей теории дискретных автоматов, которые являются важными составляющими кибернетики

Энтропия дискретного сигнала

Энтропия системы, имеющей **N** возможных состояний

$$H(x) = -\sum_{i=1}^{N} p(i) \log[p(i)]$$

Формула К. Шеннона

Где **p(1), p(2),....p(i),.....p(N)** – вероятности появления элементов алфавита $x_1, x_2, \ldots x_i, \ldots x_N$

Для случая, когда вероятность появления для всех элементов алфавита одинакова **p(1) = p(2) ...= p(N) = 1/N**

$$H(x) = -\sum_{i=1}^{N} \left(\frac{1}{N}\right) \log\left(\frac{1}{N}\right) = \log N$$

Энтропия и количество информации

Из уравнения $I(\beta) = H(\alpha) - H(\alpha/\beta)$ при $H(\alpha/\beta) = 0$ получаем:

$$I(\beta) = H(\alpha)$$

***Количество информации*, приносимое в среднем одним элементом (знаком) сообщения *равно энтропии* источника**

При использовании двоичного алфавита $N = 2$ и $p(1) = p(2) = 0{,}5$ это количество информации равно одному биту

$$I = -\sum_{1}^{2}\left(\frac{1}{2}\right)\log_2\left(\frac{1}{2}\right) = 1$$

Если сообщение содержит n разрядов и его алфавит состоит из m различных символов то при одинаковой вероятности их появления

$$I = \log N = n\log m$$

Так как $N = m^n$

Энтропия зависимой последовательности

Если появление в последовательности элемента x_i зависит от того, какой элемент x_j был предшествующий, а условная вероятность их совместного появления **p(i|j),** то сначала вычисляется условная энтропия

$$H(x_i \mid x_j) = -\sum_{i=1}^{N} p(i \mid j) \log[p(i \mid j)]$$

Для получения безусловной энтропии источника необходимо провести усреднение по вероятностям появления элементов

$$H(x) = \sum_{j=1}^{N} \left[\sum_{i=1}^{N} p(i \mid j) \log[p(i \mid j)] \right] p(j)$$

Основные свойства энтропии

1. Энтропия является величиной вещественной и неотрицательной
2. Энтропия - величина ограниченная
3. Энтропия равна **0**,если вероятность одного из состояний источника информации равна **1**
4. Энтропия максимальна при равной вероятности всех состояний источника информации
5. Энтропия объединенных статистически независимых источников информации равна сумме их энтропий
6. Энтропия характеризует среднюю неопределенность выбора одного состояния из ансамбля состояний

Энтропия непрерывных сигналов

Энтропия случайной величины **x**,
описываемой плотностью распределения вероятности $\mathbf{p_x(x)}$

$$H(x) = -\int_{-\infty}^{\infty} p_x(x) \log p_x(x) dx$$

Если возможное число реализаций случайной аналоговой величины сделать конечным и равным **N,** то Каждая из реализаций $C_1, C_2, ...C_i, ...C_N$ будет иметь определенную вероятность появления. Тогда энтропия и среднее количество информации в каждой реализации определятся известным равенством

$$H = -\sum_{i=1}^{N} p(C_i) \log[p(C_i)]$$