ИНФОРМАЦИЯ И ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ. ИЗМЕРЕНИЕ И КОДИРОВАНИЕ ИНФОРМАЦИИ

6. Измерение информации

6.6. Статистическтй подход Шеннона

статистический подход Шеннона к измерению количества информации

$$X$$

$$x_1, x_2, \ldots, x_n \qquad p_1, p_2, \ldots, p_n$$

$$p_1 + p_2 + \ldots + p_n = 1$$

x_i	x_1	x_2	$\ldots$	x_n
p_i	p_1	p_2	$\ldots$	p_n

статистический подход Шеннона

(исходные позиции к измерению количества информации)

определено количество состояний системы (количество возможных сообщений)

для каждого сообщения определена вероятность его появления

система описана как дискретная или непрерывная случайная величина

Понятие энтропии системы

$$H_i = -\log_a p_i$$

$$x_1, x_2, \ldots, x_n \qquad p_1, p_2, \ldots, p_n$$

Понятие энтропии системы

$$H_i = -log_a \, p_i$$

$$H_i \geq 0, \text{т.к. } p_i \leq 1 \text{ и } a > 1$$

степень неопределенности состояния (сообщения) *i* системы тем больше, чем меньше его вероятность

степень неопределенности состояния *i* системы равна нулю, если его вероятность равна 1 (система может принимать только одно состояние, значит она полностью определена)

Понятие энтропии системы

$$x_1, x_2, \ldots, x_n \qquad p_1, p_2, \ldots, p_n$$

Степень неопределенности состояния x_i	$-\log_a p_1$	$-\log_a p_2$	$\ldots$	$-\log_a p_n$
Вероятность состояния p_i	p_1	p_2	$\ldots$	p_n

Понятие энтропии системы

$$H(X) = -(p_1 \cdot \log_a p_1 + p_2 \cdot \log_a p_2 + \dots + p_n \cdot \log_a p_n)$$

$$H(X) = -\sum_{i=1}^{n} p_i \times \log_a p_i$$

Понятие энтропии системы

энтропией системы (источника информации) называется сумма произведений вероятностей различных состояний системы на логарифмы их вероятностей, взятая со знаком минус

Основные свойства энтропии

Энтропия всегда положительна, т.к.

$$a > 1, 0 \leq p_i \leq 1$$

и логарифмы всегда отрицательны

Энтропия обращается в нуль, когда одно из состояний системы достоверно, а другие—невозможны

При увеличении числа состояний системы энтропия увеличивается

При заданном числе состояний энтропия обращается в максимум, когда эти состояния равновероятны

Энтропия обладает свойством аддитивности

Связь формулы Шеннона и формулы Хартли

$$p = \frac{1}{n}$$

$$H(X) = -\sum_{i=1}^{n} \frac{1}{n} \cdot \log_a \frac{1}{n} = -\frac{1}{n} \cdot n \cdot \log_a \frac{1}{n} = \log_a n$$

$$H(X) = n \cdot (-p \cdot \log_a p)$$

Связь понятий *энтропия* и *информация*

естественно количество информации измерять уменьшением энтропии той системы, для уточнения состояния которой предназначены сведения

$$X$$

$$H(X)$$

$$H'(X)$$

$$I_x = H(X) - H'(X)$$

$$I_x = H(X) - 0$$

$$I_x = H(X)$$

Связь понятий *энтропия* и *информация*

количество информации, приобретаемое при полном выяснении состояния некоторой системы равно энтропии этой системы

Единицы измерения количества *информации*

состояния системы	x1	x2
вероятности состояний	0,5	0,5

$$I_x = log_2 2 = 1$$

$$I_x = -\left(\frac{1}{2}\cdot log_2 \frac{1}{2} + \frac{1}{2}\cdot log_2 \frac{1}{2}\right) = 1$$

бит **binary unit** **binary digit**

двоичная единица или бит, есть единица измерения степени неопределенности, представляющая неопределенность, которая содержится в одном опыте, имеющем два равновероятных исхода

Единицы измерения количества *информации*

трит **дит** **нат**

$$\log_2 n = \frac{\log_{10} n}{\log_{10} 2} \approx 3{,}322 \cdot \log_{10} n$$

Определение количества *информации в сообщении*

$$I_x = H(X)$$

$$H(X) = -\sum_{i=1}^{n} p_i \times \log_a p_i$$

$$I_x = -\sum_{i=1}^{n} p_i \cdot \log_a p_i = -(p_1 \cdot \log_a p_1 + p_2 \cdot \log_a p_2 + \cdots + p_n \cdot \log_a p_n)$$

$$-\log_a p_i$$

частная или собственная информация

$$X \quad x_i \qquad I_{x_i} \qquad I_{x_i} = -log_a\, p_i$$

средняя или полная информация Ix **представляется как математическое ожидание (среднее) случайной величины,**
i**-е значение которой есть собственная информация, получаемая от сообщения** x_i

Свойства частной *информации*

$$I_{x_i} = -\log_a p_i$$

Частная информация неотрицательна

$$a > o, p_i \le 1$$

Частная информация, получаемая от сообщения x_i тем больше, чем меньше априорная вероятность *pi* получения этого сообщения

Если сообщение имеет вероятность, равную единице, то информация, содержащаяся в нем, равна нулю

Частная информация обладает свойством аддитивности

частная и полная (средняя) *информация*

$$I_x = -\sum_{i=1}^{n} p_i \cdot \log_a p_i = -(p_1 \cdot \log_a p_1 + p_2 \cdot \log_a p_2 + \dots + p_n \cdot \log_a p_n)$$

$$p = \frac{1}{n}$$

$$I_{x_i} = -\log_a p_i = -\log_a \frac{1}{n} = \log_a n$$

частная и полная (средняя) информация

Если все состояния системы равновероятны, то частная информация от каждого отдельного сообщения I_{x_i} равна средней (полной) информации о состоянии системы I_x

Если же система может принимает свои состояния с различными вероятностями, то информации от разных сообщений не одинаковы: наибольшую информацию несут сообщения о тех событиях, которые априори были наименее вероятны

ВОПРОС ЗАКОНЧЕН

семантический подход к измерению количества информации

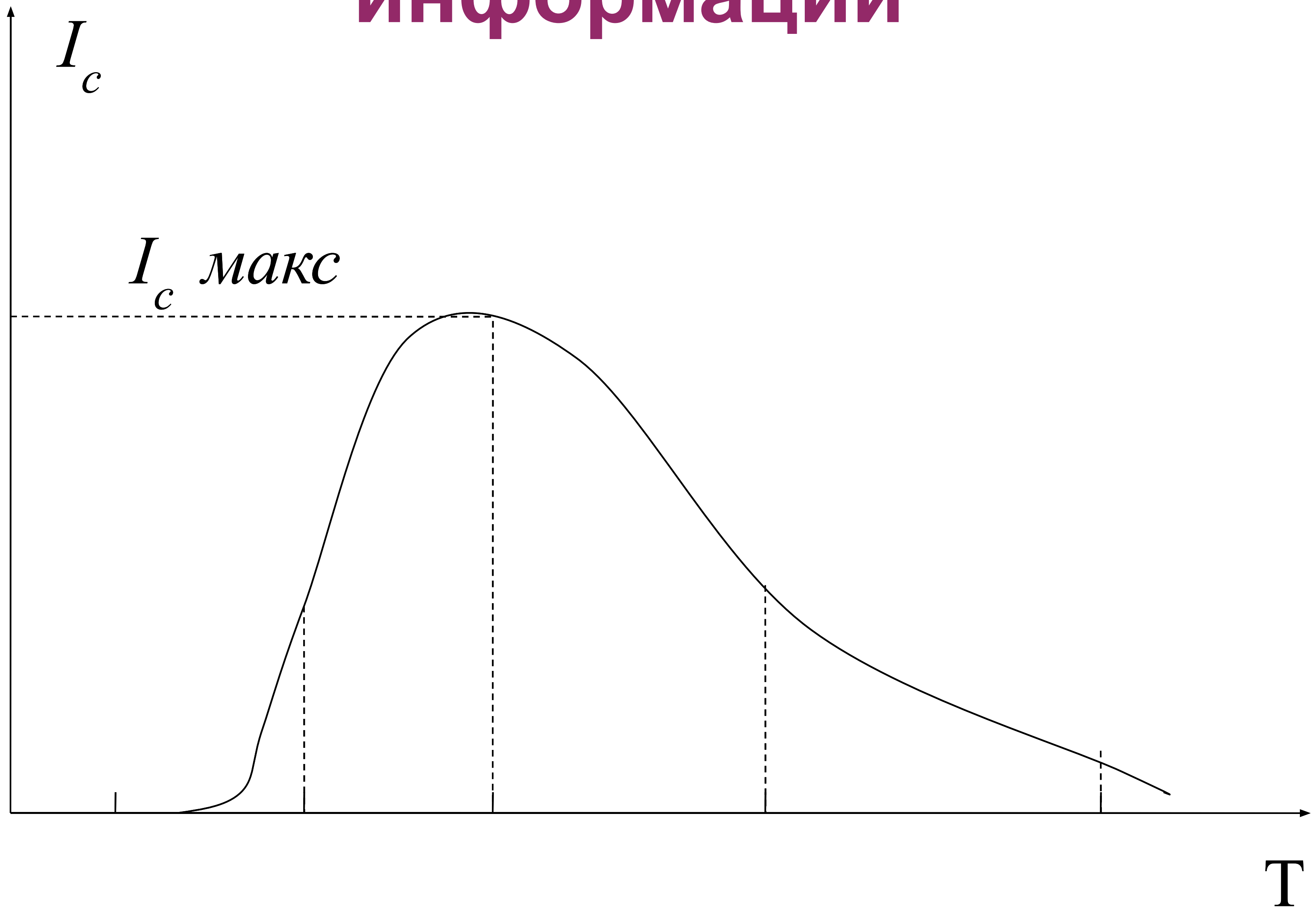
Тезаурусная мера Ю.А. Шрейдер

Тезаурус ?

Под тезаурусом (от греческого сокровище) понимается некий обобщенный справочник, определяющий уровень знаний получателя сообщений. Такой справочник включает не только описание понятий, но и связи между ними и т.д.

семантический подход к измерению количества информации

семантический подход к измерению количества информации



семантический подход к измерению количества информации

Р. Карнапом и Бар–Хиллелом разработан подход к определению количество семантической информации на основе изменения логической вероятности.

Под логической вероятностью понимается степень подтверждения той или иной гипотезы

прагматический подход к измерению количества информации

Прагматическая мера определяет ценность информации, ее полезность для достижения поставленной цели

А.А. Харкевич

$p_0 \quad p_1$

$$I_ц = \log_a p_1 - \log_a p_0 = \log_a \frac{p_1}{p_0}$$

$p_1 > p_0$

$p_1 = p_0$

$p_1 < p_0$

$I_ц > 0$

$I_ц = 0$

$I_ц < 0$