



ИНФОРМА́ЦИИ ТЕО́РИЯ

Авторы: Ю. В. Прохоров

ИНФОРМА́ЦИИ ТЕО́РИЯ, раздел математики, исследующий процессы хранения, преобразования и передачи *информации*. И. т. — часть *кибернетики*. В основе И. т. лежит определённый способ измерения количества информации, содержащейся в к.-л. данных (сообщениях).

И. т. исходит из представления о том, что сообщения, предназначенные для сохранения в запоминающем устройстве или для передачи по каналу связи, не известны заранее с полной определённой. Заранее известен лишь источник сообщений, т. е. множество x_1, x_2, \dots, x_n , из которого могут быть выбраны эти сообщения, вероятности p_1, p_2, \dots, p_n появления этих сообщений. В И. т. неопределённость, с которой сталкиваются в подобной обстановке, допускает количественное выражение, и именно это количество, а не конкретная природа самих сообщений, определяет возможность их хранения и передачи. Рассматриваются всевозможные способы записи сообщений цепочками символов 0 и 1, называемые двоичными кодами, удовлетворяющие условиям: а) разл. сообщениям соответствуют разл. цепочки; б) по записи любой последовательности сообщений в кодированной форме эта последовательность должна однозначно восстанавливаться. В качестве меры неопределённости источника сообщений принимают среднее значение длины кодовой цепочки, соответствующее самому экономному способу кодирования; единицей измерения служит один двоичный знак.

Пример. Пусть некоторые сообщения x_1, x_2, x_3 появляются с вероятностями, равными соответственно $\frac{1}{2}$, $\frac{3}{8}$, $\frac{1}{8}$. К.-л. короткий код, напр. $x_1=0, x_2=1, x_3=01$, не пригоден, т. к. нарушается условие б), поскольку цепочка 01 может означать как x_1x_2 , так и x_3 . Код $x_1=0, x_2=10, x_3=11$ удовлетворяет условиям а) и б). Ему соответствует среднее значение длины кодовой цепочки, равное $1 \cdot \frac{1}{2} + 2 \cdot \frac{3}{8} + 2 \cdot \frac{1}{8} = 1,5$. Оказывается, что никакой другой код не может дать меньшего значения, т. е. указанный код — самый экономный, мера неопределённости данного источника сообщений равна 1,5 (двоичных знаков).

Не существует простой формулы, выражающей точный минимум H' среднего числа двоичных знаков, необходимых для кодирования сообщений x_1, x_2, \dots, x_n , через вероятности p_1, p_2, \dots, p_n этих сообщений. Однако этот минимум не меньше величины $H = -\sum_{i=1}^n p_i \log_2(1/p_i)$ и может превосходить её не более чем на единицу. Величина H называемая *энтропией* источника сообщений, обладает простыми свойствами, а для всех выводов И. т., которые носят асимптотич. характер, т. е. соответствуют случаю $H' \rightarrow \infty$, различие между H и H' несущественно. Поэтому именно энтропия принимается в качестве меры неопределённости данного источника. В приведенном выше примере энтропия равна $H = \frac{1}{2} \log_2 2 + \frac{3}{8} \log_2 \frac{8}{3} + \frac{1}{8} \log_2 8 = 1,40564$.

Энтропия бесконечной совокупности сообщений x_1, x_2, \dots , появляющихся с вероятностями p_1, p_2, \dots , оказывается, как правило, бесконечной, поэтому в применении к источникам с бесконечным числом сообщений

поступают иначе. Именно, задаются определённым уровнем точности и вводят понятие ϵ -энтропии как энтропии сообщения, записываемого с точностью до ϵ , если сообщение представляет собой непрерывную величину или функцию, напр., времени.

Так же, как и понятие энтропии, понятие количества информации, содержащейся в одном случайном объекте (случайной величине, случайном векторе, случайной функции и т. д.) относительно другого, вводится сначала для объектов с конечным числом n возможных значений. Затем общий случай изучается при помощи предельного перехода при $n \rightarrow \infty$. В отличие от энтропии количество информации, напр. в одной непрерывно распределённой случайной величине относительно другой непрерывно распределённой величины, часто оказывается конечным.

Понятие [канала связи](#) в И. т. носит весьма общий характер. Канал связи задаётся указанием множества допустимых сообщений на входе канала, множеством сообщений на выходе и набором условных вероятностей получения того или иного сообщения на выходе при данном входном сообщении. Эти условные вероятности описывают влияние помех, искажающих передаваемые сообщения. Присоединяя к каналу связи источник сообщений, можно рассчитать количество информации относительно сообщения на входе, содержащееся в сообщении на выходе. Верхняя грань таких количеств информации, взятая по всем допустимым источникам, называется пропускной способностью (ёмкостью) канала. Пропускная способность канала – его осн. информац. характеристика. Несмотря на влияние (быть может, сильное) помех в канале, при определённом соотношении между энтропией поступающих сообщений и пропускной способностью канала при надлежащем кодировании возможна почти безошибочная передача сообщений.

В И. т. изучаются оптимальные в смысле скорости и надёжности способы передачи информации и устанавливаются теоретич. пределы достижимого качества. И. т. носит существенно вероятностный характер и значительная часть её математич. методов заимствуется из теории вероятностей.

Основы И. т. были заложены в 1948–1949 К. [Шенноном](#). Её теоретич. разделы разрабатывались А. Н. [Колмогоровым](#) и А. Я. [Хинчиным](#), а разделы, связанные с применениями, – В. А. [Котельниковым](#).

Литература

Лит. см. при ст. [Информация](#).