

Информационная энтропия

Материал из Википедии — свободной энциклопедии

Информационная энтро́пия — мера хаотичности [информации](#), неопределённость появления какого-либо символа [первичного алфавита](#). При отсутствии информационных потерь численно равна количеству информации на символ передаваемого сообщения.

Например, в последовательности букв, составляющих какое-либо предложение на русском языке, разные буквы появляются с разной частотой, поэтому неопределённость появления для некоторых букв меньше, чем для других. Если же учесть, что некоторые сочетания букв (в этом случае говорят об энтропии *n*-ого порядка, см. [ниже](#)) встречаются очень редко, то неопределённость ещё более уменьшается.

Для иллюстрации понятия информационной энтропии можно также прибегнуть к примеру из области [термодинамической энтропии](#), получившему название [демона Максвелла](#). Концепции информации и энтропии имеют глубокие связи друг с другом, но, несмотря на это, разработка теорий в статистической механике и теории информации заняла много лет, чтобы сделать их соответствующими друг другу.

Содержание

- [1 Формальные определения](#)
 - [1.1 Альтернативное определение](#)
- [2 Свойства](#)
 - [2.1 Математические свойства](#)
 - [2.2 Эффективность](#)
- [3 Вариации и обобщения](#)
 - [3.1 Условная энтропия](#)
 - [3.2 Взаимная энтропия](#)
- [4 История](#)
- [5 Примечания](#)
- [6 См. также](#)
- [7 Ссылки](#)
- [8 Литература](#)

Формальные определения

Определение с помощью [собственной информации](#)^[1]

Также можно определить энтропию случайной величины, введя предварительно понятия распределения случайной величины *X*, имеющей конечное число значений:

$$P_X(x_i) = p_i, \quad p_i \geq 0, \quad i = 1, 2, \dots, n$$
$$\sum_{i=1} p_i = 1$$

и собственной информации:

$$I(X) = -\log P_X(X).$$

Тогда энтропия будет определяться как:

$$H(X) = E(I(X)) = -\sum_{i=1}^n p(i) \log p(i).$$

От основания логарифма зависит единица измерения информации и энтропии: [бит](#), [нат](#) или [харгли](#).

Информационная энтропия для независимых случайных событий x с n возможными состояниями (от 1 до n) рассчитывается по формуле:

$$H(x) = -\sum_{i=1}^n p(i) \log_2 p(i).$$

Эта величина также называется *средней энтропией сообщения*. Величина $\log_2 \frac{1}{p(i)}$ называется *частной энтропией*, характеризующей только i -е состояние.

Таким образом, энтропия события x является суммой с противоположным знаком всех произведений относительных частот появления события i , умноженных на их же двоичные логарифмы (основание 2 выбрано только для удобства работы с информацией, представленной в двоичной форме). Это определение для дискретных случайных событий можно расширить для функции распределения [вероятностей](#).

[Шеннон](#) предположил, что прирост информации равен утраченной неопределённости, и задал требования к её измерению:

- мера должна быть непрерывной; то есть изменение значения величины вероятности на малую величину должно вызывать малое результирующее изменение функции;
- в случае, когда все варианты (буквы в приведённом примере) равновероятны, увеличение количества вариантов (букв) должно всегда увеличивать значение функции;
- должна быть возможность сделать выбор (в нашем примере букв) в два шага, в которых значение функции конечного результата должно являться суммой функций промежуточных результатов.

Шеннон показал, что единственная функция, удовлетворяющая этим требованиям, имеет вид:

$$-K \sum_{i=1}^n p(i) \log_2 p(i),$$

где K — константа (и в действительности нужна только для выбора единиц измерения).

Шеннон определил, что измерение энтропии ($H = -p_1 \log_2 p_1 - \dots - p_n \log_2 p_n$), применяемое к источнику информации, может определить требования к минимальной пропускной способности канала, требуемой для надёжной передачи информации в виде закодированных двоичных чисел. Для вывода формулы Шеннона необходимо вычислить [математическое ожидание](#) «количества информации», содержащегося в цифре из источника информации. Мера энтропии Шеннона выражает неуверенность реализации случайной переменной. Таким образом, энтропия является разницей между информацией, содержащейся в сообщении, и той частью информации, которая точно известна (или хорошо предсказуема) в сообщении. Примером этого является избыточность языка — имеются явные статистические закономерности в появлении букв, пар последовательных букв, троек и т. д. См.: [Цепи Маркова](#).

В общем случае **b -арная энтропия** (где b равно 2, 3, ...) источника $\mathcal{S} = (S, P)$ с [исходным алфавитом](#) $S = \{a_1, \dots, a_n\}$ и [дискретным распределением вероятности](#) $P = \{p_1, \dots, p_n\}$ где p_i является вероятностью a_i ($p_i = p(a_i)$) определяется формулой:

$$H_b(\mathcal{S}) = - \sum_{i=1}^n p_i \log_b p_i.$$

Определение энтропии Шеннона связано с понятием [термодинамической энтропии](#). [Больцман](#) и [Гиббс](#) проделали большую работу по статистической термодинамике, которая способствовала принятию слова «энтропия» в информационную теорию. Существует связь между термодинамической и информационной энтропией. Например, [демон Максвелла](#) также противопоставляет термодинамическую энтропию информации, и получение какого-либо количества информации равно потерянной энтропии.

Альтернативное определение

Другим способом определения функции энтропии H является доказательство, что H однозначно определена (как указано ранее), если и только если H удовлетворяет условиям:

1. $H(p_1, \dots, p_n)$ определена и непрерывна для всех p_1, \dots, p_n , где $p_i \in [0, 1]$ для всех $i = 1, \dots, n$ и $p_1 + \dots + p_n = 1$. (Заметьте, что эта функция зависит только от распределения вероятностей, а не от алфавита.)
2. Для целых положительных n , должно выполняться следующее неравенство:

$$H \left(\underbrace{\frac{1}{n}, \dots, \frac{1}{n}}_n \right) < H \left(\underbrace{\frac{1}{n+1}, \dots, \frac{1}{n+1}}_{n+1} \right).$$

3. Для целых положительных b_i , где $b_1 + \dots + b_k = n$, должно выполняться равенство:

$$H \left(\underbrace{\frac{1}{n}, \dots, \frac{1}{n}}_n \right) = H \left(\frac{b_1}{n}, \dots, \frac{b_k}{n} \right) + \sum_{i=1}^k \frac{b_i}{n} H \left(\underbrace{\frac{1}{b_i}, \dots, \frac{1}{b_i}}_{b_i} \right).$$

Свойства

Важно помнить, что энтропия является количеством, определённым в контексте вероятностной модели для [источника данных](#). Например, кидание монеты имеет энтропию $-2(0,5\log_2 0,5) = 1$ [бит](#) на одно кидание (при условии его независимости). У источника, который генерирует строку, состоящую только из букв «А», энтропия равна

$$-\sum_{i=1}^{\infty} \log_2 1 = 0$$

нулю: . Так, например, опытным путём можно установить, что энтропия английского текста равна 1,5 бит на символ, что конечно будет варьироваться для разных текстов. Степень энтропии источника данных означает среднее число битов на элемент данных, требуемых для её зашифровки без потери информации, при оптимальном кодировании.

1. Некоторые биты данных могут не нести информации. Например, структуры данных часто хранят избыточную информацию, или имеют идентичные секции независимо от информации в структуре данных.
2. Количество энтропии не всегда выражается целым числом бит.

Математические свойства

1. *Неотрицательность:* $H(X) \geq 0$.
2. *Ограниченность:* $H(X) \leq \log_2 |X|$. Равенство, если все элементы из X равновероятны.
3. Если X, Y независимы, то $H(XY) = H(X) + H(Y)$.
4. Энтропия — выпуклая вверх функция распределения вероятностей элементов.
5. Если X, Y имеют одинаковое распределение вероятностей элементов, то $H(X) = H(Y)$.

Эффективность

Исходный алфавит, встречающийся на практике, имеет вероятностное распределение, которое далеко от оптимального. Если исходный алфавит имел n символов, тогда он может быть сравнён с «оптимизированным алфавитом», вероятностное распределение которого однородно. Соотношение энтропии исходного и оптимизированного алфавита — это эффективность исходного алфавита, которая может быть выражена в процентах.

Из этого следует, что эффективность исходного алфавита с n символами может быть определена просто как равная его n -арной энтропии.

Энтропия ограничивает максимально возможное сжатие без потерь (или почти без потерь), которое может быть реализовано при использовании теоретически — [типичного набора](#) или, на практике, — [кодирования Хаффмана](#), [кодирования Лемпеля — Зива](#) — [Велча](#) или [арифметического кодирования](#).

Вариации и обобщения

Условная энтропия

Если следование символов алфавита не независимо (например, во французском языке после буквы «q» почти всегда следует «u», а после слова «передовик» в советских газетах обычно следовало слово «производства» или «труда»), количество информации, которую несёт последовательность таких символов (а следовательно и энтропия) очевидно меньше. Для учёта таких фактов используется условная энтропия.

Условной энтропией первого порядка (аналогично для [Марковской модели](#) первого порядка) называется энтропия для алфавита, где известны [вероятности](#) появления одной буквы после другой (то есть вероятности двухбуквенных сочетаний):

$$H_1(\mathcal{S}) = - \sum_i p_i \sum_j p_i(j) \log_2 p_i(j),$$

где i — это состояние, зависящее от предшествующего символа, и $p_i(j)$ — это вероятность j , при условии, что i был предыдущим символом.

Так, для русского языка без буквы «ё» $H_0 = 5$, $H_1 = 4,358$, $H_2 = 3,52$, $H_3 = 3,01$.^[2]

Через частную и общую условные энтропии полностью описываются информационные потери при передаче данных в канале с помехами. Для этого применяются так называемые *канальные матрицы*. Так, для описания потерь со стороны источника (то есть известен посланный сигнал), рассматривают [условную вероятность](#) $p(b_j | a_i)$ получения приёмником символа b_j при условии, что был отправлен символ a_i . При этом канальная матрица имеет следующий вид:

	b_1	b_2	...	b_j	...	b_m
a_1	$p(b_1 a_1)$	$p(b_2 a_1)$...	$p(b_j a_1)$...	$p(b_m a_1)$
a_2	$p(b_1 a_2)$	$p(b_2 a_2)$...	$p(b_j a_2)$...	$p(b_m a_2)$
...
a_i	$p(b_1 a_i)$	$p(b_2 a_i)$...	$p(b_j a_i)$...	$p(b_m a_i)$
...
a_m	$p(b_1 a_m)$	$p(b_2 a_m)$...	$p(b_j a_m)$...	$p(b_m a_m)$

Очевидно, вероятности, расположенные по диагонали описывают вероятность правильного приёма, а сумма всех элементов столбца даст вероятность появления соответствующего символа на стороне приёмника — $p(b_j)$. Потери, приходящиеся на передаваемый сигнал a_i , описываются через частную условную энтропию:

$$H(B | a_i) = - \sum_{j=1}^m p(b_j | a_i) \log_2 p(b_j | a_i).$$

Для вычисления потерь при передаче всех сигналов используется общая условная энтропия:

$$H(B | A) = \sum_i p(a_i) H(B | a_i).$$

$H(B | A)$ означает энтропию со стороны источника, аналогично рассматривается $H(A | B)$ — энтропия со стороны приёмника: вместо $p(b_j | a_i)$ всюду указывается $p(a_i | b_j)$ (суммируя элементы строки можно получить $p(a_i)$, а элементы диагонали означают вероятность того, что был отправлен именно тот символ, который получен, то есть вероятность правильной передачи).

Взаимная энтропия

Взаимная энтропия, или *энтропия объединения*, предназначена для расчёта энтропии взаимосвязанных систем (энтропии совместного появления статистически зависимых сообщений) и обозначается $H(AB)$, где A , как всегда, характеризует передатчик, а B — приёмник.

Взаимосвязь переданных и полученных сигналов описывается вероятностями совместных событий $p(a_i b_j)$, и для полного описания характеристик канала требуется только одна матрица:

$p(a_1 b_1)$	$p(a_1 b_2)$...	$p(a_1 b_j)$...	$p(a_1 b_m)$
$p(a_2 b_1)$	$p(a_2 b_2)$...	$p(a_2 b_j)$...	$p(a_2 b_m)$
...
$p(a_i b_1)$	$p(a_i b_2)$...	$p(a_i b_j)$...	$p(a_i b_m)$
...
$p(a_m b_1)$	$p(a_m b_2)$...	$p(a_m b_j)$...	$p(a_m b_m)$

Для более общего случая, когда описывается не канал, а просто взаимодействующие системы, матрица необязательно должна быть квадратной. Очевидно, сумма всех элементов столбца с номером j даст $p(b_j)$, сумма строки с номером i есть $p(a_i)$, а сумма всех элементов матрицы равна 1. Совместная вероятность $p(a_i b_j)$ событий a_i и b_j вычисляется как произведение исходной и условной вероятности,

$$p(a_i b_j) = p(a_i) p(b_j | a_i) = p(b_j) p(a_i | b_j).$$

Условные вероятности производятся по [формуле Байеса](#). Таким образом имеются все данные для вычисления энтропий источника и приёмника:

$$\begin{aligned} H(A) &= - \sum_i \left(\sum_j p(a_i b_j) \log \sum_j p(a_i b_j) \right), \\ H(B) &= - \sum_j \left(\sum_i p(a_i b_j) \log \sum_i p(a_i b_j) \right). \end{aligned}$$

Взаимная энтропия вычисляется последовательным суммированием по строкам (или по столбцам) всех вероятностей матрицы, умноженных на их логарифм:

$$H(AB) = - \sum_i \sum_j p(a_i b_j) \log p(a_i b_j).$$

Единица измерения — бит/два символа, это объясняется тем, что взаимная энтропия описывает неопределённость на пару символов — отправленного и полученного. Путём несложных преобразований также получаем

$$H(AB) = H(A) + H(B | A) = H(B) + H(A | B).$$

Взаимная энтропия обладает свойством *информационной полноты* — из неё можно получить все рассматриваемые величины.

История

В [1948 году](#), исследуя проблему рациональной передачи информации через зашумлённый коммуникационный канал, [Клод Шеннон](#) предложил революционный вероятностный подход к пониманию коммуникаций и создал первую, истинно математическую, теорию [энтропии](#). Его сенсационные идеи быстро послужили основой разработки двух основных направлений: [теории информации](#), которая использует понятие [вероятности](#) и [эргодическую теорию](#) для изучения статистических характеристик данных и коммуникационных систем, и [теории кодирования](#), в которой используются главным образом алгебраические и геометрические инструменты для разработки эффективных кодов.

Понятие энтропии, как меры случайности, введено Шенноном в его статье «A Mathematical Theory of Communication», опубликованной в двух частях в [Bell System Technical Journal](#) в 1948 году.

Примечания

1. ↑ [Габидулин, Э. М., Пилипчук, Н. И.](#) Лекции по теории информации. — М.: МФТИ, 2007. — С. 16. — 214 с. — ISBN 5-7417-0197-3.
2. ↑ [Лебедев Д. С., Гармаш В. А.](#) О возможности увеличения скорости передачи телеграфных сообщений. — М.: Электросвязь, 1958. — № 1. — С. 68—69.

См. также

- [Взаимная информация](#)
- [Энтропийное кодирование](#)
- [Цепь Маркова](#)

Ссылки

- [Shannon Claude E. A Mathematical Theory of Communication](#)^(англ.)
- Коротчаев С. М. [Энтропия и информация — универсальные естественнонаучные понятия.](#)

Литература

- *Шеннон К.* Работы по теории информации и кибернетике. — М.: Изд. иностр. лит., 1963. — 830 с.
- *Волькенштейн М. В.* Энтропия и информация. — М.: Наука, 1986. — 192 с.
- *Цымбал В. П.* Теория информации и кодирование. — М.: Выща Школа, 1977. — 288 с.