§ 15. Различные подходы к измерению информации

15.1. Содержательный подход

Сегодня информация является одним из основных ресурсов человечества. Поэтому так важны ответы на вопросы, как много информации мы получили, передали, обработали, создали.

При физических измерениях величину сравнивают с эталоном, а с чем сравнивать информацию?

Известно несколько подходов к измерению количества информации.

При **содержательном подходе** измерение информации происходит с точки зрения ее содержания, т. е. определяется, в какой мере пришедшая информация (знания) уменьшает незнание. Человек получает знания посредством сообщений. Чем больше пополняет наши знания сообщение, тем большее количество информации в нем заключено (пример 15.1).

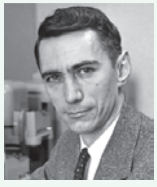
Основателем такого подхода к измерению информации является К. Шенон, который ввел приведенное ниже определение.

Сообщение, которое уменьшает неопределенность знания в два раза, несет **1 бит** информации.

Неопределенность знания о результате некоторого события — количество возможных результатов.

Если в некотором сообщении содержатся сведения о том, что произошло одно из N равновероятных событий, то количество информации i, содержащееся в сообщении, можно определить из формулы Хартли: $N = 2^i$ (пример 15.2).

Клод Элвуд Шеннон (1916—2001) — американский инженер, криптоаналитик и математик. Является основателем теории информации.



Пример 15.1. У вас сегодня контрольная по математике. Учитель обычно дает 2 варианта заданий. До контрольной вы не знаете свой вариант, поэтому неопределенность знания равна 2. Если вариантов на контрольной 4, то неопределенность знания равна 4.

Ральф Винтон Лайон Хартли (1888—1970) — американский ученый-электронщик. Предложил генератор Хартли, преобразование Хартли и сделал вклад в теорию информации, введя в 1928 г. логарифмическую меру информации: $i = \log_2 N$.

Пример 15.2. Количество информации, которую вы получите, узнав свой вариант контрольной работы, можно рассчитать по формуле Хартли.

Если вариантов два, то $2 = 2^i$, следовательно, $i = 1$. Вы получите 1 бит информации.

Если вариантов 4, то $4 = 2^i$, следовательно, $i = 2$. Вы получите 2 бита информации.

Правообладатель Народная асвета

Если вариантов 6, то $6 = 2^i$, следовательно, $i \approx 2{,}58$. Вы получите 2,58 бита информации. Для получения значения i в этом случае нужно посчитать значение $i = \log_2 6$ (например, на калькуляторе или по таблице[1]).

Пример 15.3. При компьютерном наборе текста на русском языке обычно используется 32 буквы (буква «ё» применяется очень редко). Тогда, согласно формуле Хартли, $32 = 2^5$, одна буква русского алфавита несет 5 бит информации.

Пример 15.4. Единицы измерения объемов информации.

Килобайт (Кбайт):
2^{10} = 1024 байта

Мегабайт (Мбайт):
2^{20} = 1024 килобайта = 1 048 576 байт

Гигабайт (Гбайт):
2^{30} = 1024 мегабайта = 1 073 741 824 байта

Терабайт (Тбайт):
2^{40} = 1024 гигабайта = 1 099 511 627 776 байт

Петабайт (Пбайт):
2^{50} = 1024 терабайта =
= 1 125 899 906 842 624 байта

Эксабайт (Эбайт):
2^{60} = 1024 петабайта =
= 1 152 921 504 606 846 976 байт

Зеттабайт (Збайт):
2^{70} = 1024 эксабайта =
= 1 180 591 620 717 411 303 424 байта

Йоттабайт (Йбайт):
2^{80} = 1024 зеттабайта =
= 1 208 925 819 614 629 174 706 176 байт

С точки зрения теории измерений единицы измерения количества информации, в названии которых есть части «кило», «мега» и др., некорректны. Эти приставки используются в метрической системе мер, в которой в качестве множителей кратных единиц применяется коэффициент 10^n, где n = 3, 6, 9 и т. д.

15.2. Алфавитный подход

Если человек получает текстовое сообщение, то количество информации может быть измерено количеством символов в нем. Однако каждый символ алфавита тоже несет какое-то количество информации. Если предположить, что все символы алфавита встречаются в тексте с одинаковой частотой (равновероятно), то количество информации i, которое несет каждый символ, вычисляется по формуле Хартли: $N = 2^i$, где N — мощность алфавита (пример 15.3). Под мощностью алфавита понимают количество символов в нем.

Алфавитный (объемный) подход используется, если для преобразования, хранения и передачи информации применяют технические средства.

При использовании двоичного алфавита один символ несет 1 единицу информации — **1 бит.**

Для измерения объемов информации применяют производные единицы измерения (пример 15.4).

Для двоичного представления текстов в компьютере часто используется восьмиразрядный код. С его помощью можно закодировать алфавит из 256 символов. Один символ из алфавита мощностью $256 = 2^8$ несет в тексте 8 бит информации. Такое количество информации называется **байтом.**

Объем текста измеряется в байтах. При восьмиразрядном кодировании 1 символ = 1 байт, и информационный объем текста определяется количе-

[1] Таблица двоичных логарифмов: http://sokolova-aa.ru/cribs/tablitsa-dvoichnykh-logarifmov-tselykh-chisel-ot-1-do-64 (дата доступа: 28.07.2019).

Правообладатель Народная асвета

ством символов в нем. Если весь текст состоит из K символов, то при алфавитном подходе объем V содержащейся в нем информации равен: $V = K \cdot i$, где i — информационный вес одного символа в используемом алфавите.

15.3. Вероятностный подход

В жизни различные события происходят с разной вероятностью. Событие «летом идет снег» маловероятно, а у события «осенью идет дождь» вероятность велика. Если в коробке 10 красных шаров и 40 зеленых, то вероятность достать не глядя зеленый шар больше, чем вероятность достать красный.

Для количественного измерения вероятности используют следующий подход: если общее количество возможных исходов какого-либо события равно N, а K из них — те, в которых мы заинтересованы, то вероятность интересующего нас события может быть посчитана по формуле $p = \frac{K}{N}$ (пример 15.5).

Чем меньше вероятность события, тем больше информации содержит сообщение о том, что это событие произошло.

Вероятностный подход применяется для измерения количества информации при наступлении событий, имеющих разную вероятность. Связь между вероятностью события и количеством информации в сообщении о нем выражается формулой $\frac{1}{p} = 2^i$, где p — вероятность события, а i — количество информации (пример 15.6).

В 1999 г. утвердили ряд новых приставок для единиц измерения количества информации: киби (kibi), меби (mebi), гиби (gibi), теби (tebi), пети (peti), эксби (exbi)[1]. В настоящее время они используются наравне с «кило», «мега» и др.

Пример 15.5. В коробке 16 красных шаров и 48 зеленых. Какова вероятность достать зеленый шар не глядя? Красный?

Всего в коробке $N = 16 + 48 = 64$ шара. Нас интересует зеленый шар. Благоприятный исход — достать любой из 48 зеленых шаров. Поэтому $p = \frac{48}{64} = 0{,}75$. Аналогично получим вероятность достать красный шар: $p = \frac{16}{64} = 0{,}25$. Значит, вероятность вытащить зеленый шар в 3 раза больше, чем вытащить красный шар.

* Если произошло несколько разновероятностных событий, то количество информации можно определять по формуле Шеннона, предложенной им в 1948 г: $I = -(p_1 \log_2 p_1 + p_2 \log_2 p_2 + \ldots\ p_N \log_2 p_N)$, где I — количество информации; N — количество возможных событий; p_i — вероятность i-го события.

Легко заметить, что если вероятности p_1, ..., p_N равны между собой, то каждая из них равна $1/N$ и формула Шеннона превращается в формулу Хартли.

Пример 15.6. Какое количество информации несет сообщение «Из коробки достали красный шар» для примера 15.5?

По формуле $\frac{1}{p} = 2^i$ получаем $\frac{1}{0{,}25} = 2^i \Rightarrow 4 = 2^i$. Тогда $i = 2$, т. е. мы получили 2 бита информации.

[1] https://ru.wikipedia.org/wiki/Двоичные_приставки (дата доступа: 28.07.2019).

Правообладатель Народная асвета

Пример 15.7. Выпадение каждой грани кубика равновероятно. Поэтому количество информации от одного результата бросания находится из уравнения $2^i = 6$. Тогда $2^i = 6 < 8 = 2^3$, $i = 3$ бита.

Можно рассуждать и так:

$i = \log_2 6 = 2{,}585$ бита ≈ 3 бита.

Пример 15.8. 11 Кбайт = 11 * 1024 = = 11264 байт. Поскольку количество байт равно количеству символов, то использована восьмибитная кодовая таблица. Мощность алфавита: $2^8 = 256$.

Пример 15.9. Нужно закодировать 65 равновероятностных значений. По формуле Харли: $2^i = 65 < 128 = 2^7$, $i = 7$ бит.

Пример 15.10. Пусть в ящике x желтых мячей. Тогда вероятность достать желтый мяч равна $\frac{x}{32}$. Подставляем в формулу, связывающую вероятность с количеством информации:

$$\frac{1}{p} = 2^i \Rightarrow \frac{1}{x/32} = 2^4 \Rightarrow \frac{32}{x} = 16 \Rightarrow x = 2.$$

В ящике 2 желтых мяча.

Пример 15.11*. Всего в коробке $10 + 8 + 6 = 24$ кубика. Вероятности доставания кубиков: $p_{кр} = \frac{10}{24}$, $p_{зел} = \frac{8}{24}$, $p_{ж} = \frac{6}{24}$. Количество информации по формуле Шеннона:

$$I = -(p_{кр}\log_2 p_{кр} + p_{зел}\log_2 p_{зел} + p_{ж}\log_2 p_{ж}) =$$
$$= \left(\frac{10}{24}\log_2\frac{10}{24} + \frac{8}{24}\log_2\frac{8}{24} + \frac{6}{24}\log_2\frac{6}{24}\right) \approx$$
$$\approx -(0{,}42 \cdot (-1{,}26) + 0{,}33 \cdot (-1{,}58) +$$
$$+ 0{,}25 \cdot (-2)) = 1{,}5506$$

15.4. Решение задач на определение объема информации

Пример 15.7. При игре в кости используется кубик с шестью гранями. Сколько бит информации получает игрок при каждом бросании кубика? Ответ округлить в бóльшую сторону до ближайшего целого количества бит.

Пример 15.8. Объем сообщения равен 11 Кбайт. Сообщение содержит 11 264 символа. Какова мощность алфавита?

Пример 15.9. Измеряется температура воздуха, которая может быть целым числом от –30 до 34 градусов. Какое наименьшее целое количество бит необходимо, чтобы закодировать одно измеренное значение?

Пример 15.10. В ящике находится 32 теннисных мяча, среди которых есть мячи желтого цвета. Наудачу вынимается один мяч. Сообщение «Извлечен мяч желтого цвета» несет 4 бита информации. Сколько желтых мячей в ящике?

Пример 15.11*. В коробке лежат кубики. Известно, что среди них 10 красных, 8 зеленых, 6 желтых. Вычислите вероятность доставания кубика каждого цвета. Сколько информации несет сообщение, что достали кубик любого цвета? Для решения задачи воспользоваться формулой Шеннона.

? **1.** В чем сущность содержательного подхода к измерению информации?
2. Что обозначает 1 бит информации при алфавитном подходе к измерению информации?
3. Когда применяют вероятностный подход к измерению информации?

Правообладатель Народная асвета

Упражнения

1 Сколько различных звуковых сигналов можно закодировать с помощью 6 бит?

2 Какое количество информации несет сообщение о том, что человек живет в первом или втором подъезде, если в доме 16 подъездов?

3 Сообщение о том, что ваш друг живет на 10-м этаже, несет 4 бита информации. Сколько этажей в доме?

4 Азбука Морзе позволяет кодировать символы для радиосвязи, задавая комбинацию точек и тире. Сколько различных символов (цифр, букв, знаков пунктуации и т. д.) можно закодировать, используя код Морзе длиной не менее пяти и не более шести сигналов (точек и тире)?

5 В ящике находится 32 теннисных мяча, среди которых есть мячи черного цвета. Наудачу вынимается один мяч. Сообщение «Извлечен мяч НЕ черного цвета» несет 3 бита информации. Сколько черных мячей в ящике?

6 К празднику надували белые и синие шарики. Белых шариков 24. Сообщение о том, что лопнул синий шарик, несет 2 бита информации. Сколько всего надули шариков?

7 В школьной библиотеке 32 стеллажа с книгами, на каждом — по 8 полок. Пете сообщили, что нужный учебник находится на 2-й полке 4-го стеллажа. Какое количество информации получил Петя?

8 Для регистрации на некотором сайте пользователю нужно придумать пароль, состоящий из 10 символов. В качестве символов можно использовать десятичные цифры и шесть первых букв латинского алфавита, причем буквы используются только заглавные. Пароли кодируются посимвольно. Все символы кодируются одинаковым и минимально возможным количеством бит. Для хранения сведений о каждом пользователе в системе отведено одинаковое и минимально возможное целое число байт. Какой объем будет занимать информация о паролях 1000 пользователей?

9* В некоторой стране автомобильный номер длиной 6 символов составляют из заглавных букв (задействовано 30 различных букв) и десятичных цифр в любом порядке. Каждый такой номер в компьютерной программе записывается минимально возможным и одинаковым целым количеством байт (при этом используют посимвольное кодирование, и все символы кодируются одинаковым и минимально возможным количеством бит). Определите объем памяти в байтах, отводимый этой программой для записи 50 номеров.

10 В озере обитают 12 500 окуней, 25 000 пескарей, а карасей и щук по 6250. Какое количество информации несет сообщение о том, что поймали пескаря? Сколько информации мы получим, когда поймаем какую-нибудь рыбу?

11* Какое сообщение содержит большее количество информации?

1. Бабушка испекла 16 пирожков. Лера съела один пирожок.
2. Бабушка испекла 12 пирожков с капустой, 12 пирожков с повидлом. Маша съела один пирожок.
3. Бабушка испекла 16 пирожков с капустой, 24 пирожка с повидлом. Миша съел один пирожок.

Правообладатель Народная асвета