

Тема занятия: Измерение количества информации

Задание:

1. Изучить краткий теоретический материал по теме, рассмотреть примеры решения задач.
2. Решить задачи из раздела «Задачи для самостоятельного решения». Решение задач оформить в тетради для практических занятий.

Краткий теоретический материал

Рассмотрим два подхода к измерению информации – **содержательный** (вероятностный) и **символьный** (алфавитный).

В **содержательном подходе** возможна качественная оценка информации: новая, срочная, важная и т. д. Согласно К. Шеннону, информативность сообщения характеризуется содержащейся в нем полезной информацией – той частью сообщения, которая снимает полностью или уменьшает неопределенность какой-либо ситуации. Неопределенность некоторого события – это количество возможных исходов данного события.

Содержательный подход часто называют субъективным, так как разные люди (субъекты) информацию об одном и том же предмете оценивают по-разному. Но если число исходов не зависит от суждений людей (например, случай бросания кубика или монеты), то информация о наступлении одного из возможных исходов является объективной.

Формулу для вычисления количества информации, учитывающую **неодинаковую** вероятность событий, предложил К. Шеннон в 1948 г.

Количественная зависимость между вероятностью события p и количеством информации I в сообщении о нем выражается формулой Шеннона:

$$I = -\log_2 p$$

Качественную связь между вероятностью события и количеством информации в сообщении об этом событии можно выразить следующим образом: *чем*

меньше вероятность некоторого события, тем больше информации содержит сообщение об этом событии.

Количество информации для событий с различными вероятностями определяется по формуле (эту формулу также называют формулой Шеннона)

$$I = - \sum_{i=1}^n p_i \log_2 p_i$$

Рассмотрим пример. В коробке имеется 100 шаров. Из них 80 белых и 20 черных. Очевидно, вероятность того, что при вытаскивании случайным образом попадет белый шар, больше, чем вероятность попадания черного. Проведем количественную оценку вероятности для каждой ситуации.

Обозначим $p_{\text{ч}}$ – вероятность, что вытащили черный шар, $p_{\text{б}}$ – вероятность, что вытащили белый шар.

Тогда: $p_{\text{ч}} = 20/100 = 0,2$, $p_{\text{б}} = 80/100 = 0,8$. Заметим, что вероятность попадания белого шара в 4 раза больше, чем черного.

Количество информации в сообщении, что вынутый случайным образом шар является черным, вычисляется по формуле

$$I_{\text{ч}} = -\log_2(0,2) = 2,321928 \text{ (бит)}.$$

Количество информации в сообщении, что вынутый случайным образом шар является белым, вычисляется по формуле

$$I_{\text{б}} = -\log_2(0,8) = 0,321928 \text{ (бит)}.$$

Количество информации в сообщении о цвете вынутого случайным образом шара вычисляется по формуле

$$I = -(0,2 \log_2(0,2) + 0,8 \log_2(0,8)) = 0,721928 \text{ (бит)}.$$

Если события **равновероятны** ($p_i = \frac{1}{N}$, где N – число возможных событий), то величина количества информации I вычисляется по формуле Р. Хартли:

$$I = \log_2 N$$

Используя формулу Хартли, можно записать формулу, которая связывает количество возможных событий N и количество информации I :

$$N = 2^I$$

Алфавитный подход основан на том, что всякое сообщение можно закодировать с помощью конечной последовательности символов некоторого алфавита. С позиций информатики носителями информации являются любые последовательности символов, которые хранятся, передаются и обрабатываются с помощью компьютера. Информативность последовательности символов зависит не от содержания сообщения, а определяется минимально необходимым количеством символов для кодирования этой последовательности символов. Алфавитный подход является объективным, т. е. он не зависит от субъекта, воспринимающего сообщение. Смысл сообщения либо учитывается на этапе выбора алфавита кодирования, либо не учитывается вообще.

При алфавитном подходе, если допустить, что *все символы алфавита встречаются в тексте с одинаковой частотой (равновероятно)*, то количество информации, которое несет каждый символ (информационный вес одного символа), вычисляется по формуле

$$I = \log_2 N,$$

где N – **мощность алфавита** (полное количество символов, составляющих алфавит выбранного кодирования).

Тогда мощность алфавита можно вычислить по формуле

$$N = 2^I$$

В алфавите, который состоит из двух символов (двоичное кодирование), каждый символ несет 1 бит ($\log_2 2 = 1$) информации;

из четырех символов – каждый символ несет 2 бита информации ($\log_2 4 = 2$);

из восьми символов – 3 бита ($\log_2 8 = 3$) и т. д.

Один символ из алфавита мощностью 256 несет в тексте 8 битов ($\log_2 256 = 8$) информации.

Если весь текст состоит из k символов, то при алфавитном подходе размер содержащейся в нем информации H определяется по формуле

$$H = k \times I$$

где I – информационный вес одного символа в используемом алфавите.

Примеры решения задач

Задача 1. Два игрока играют в «крестики нолики» на поле размером 4×4 . Определить, какое количество информации I получит второй игрок после первого хода первого игрока.

Решение. Первый игрок может для первого хода выбрать любое поле из 16 возможных ($N = 4 \times 4 = 16$).

Тогда по формуле Хартли:

$$I = \log_2 16 = \log_2 2^4 = 4 \text{ (бита)}.$$

Количество информации I можно также найти из соотношения $N = 2^I$

$$16 = 2^I \quad 2^4 = 2^I \quad \text{Отсюда } I = 4 \text{ бита}.$$

Ответ: $I = 4$ бита.

Задача 2. В группе 24 студента. За экзамен были получены следующие оценки: 3 пятерки, 12 четверок, 6 троек, 3 двойки.

Определить, какое количество информации I содержит сообщение, что студент Романов получил оценку «четыре».

Решение. Вероятность события, что случайным образом выбранный студент получил оценку «четыре», равна $p = \frac{12}{24} = \frac{1}{2}$.

Используя формулу Шеннона, получим

$$I = -\log_2 p = -\log_2 \frac{1}{2} = -\log_2 2^{-1} = \log_2 2 = 1 \text{ бит.}$$

Ответ: $I = 1$ бит.

Задача 3. В коробке лежат красные и синие карандаши, всего в коробке 24 карандаша. Информация о том, что из коробки случайным образом достали синий карандаш, равна 2 битам. Определить, сколько в коробке красных и синих карандашей.

Решение. Обозначим за x число синих карандашей в коробке. Для решения задачи воспользуемся формулой Шеннона: $I = -\log_2 p$. Получим $2 = -\log_2 p$.

Из этого соотношения найдем p – вероятность того, что случайным образом вынутый шар является синим: $p = \frac{1}{4}$. Теперь определим x из соотношения

$$\frac{x}{24} = \frac{1}{4} \quad x = 6.$$

Ответ: В коробке 6 синих и 18 красных карандашей.

Задача 4. Растровое графическое изображение 20×20 точек содержит не более 256 цветов. Сколько памяти потребуется для хранения изображения?

Решение. Для решения воспользуемся формулой $N = 2^i$. Одна точка может иметь один из 256 цветов ($N = 256$). Найдем сколько бит i требуется для ее хранения (битовая глубина) из соотношения: $256 = 2^i$ $i = 8$ бит.

Для хранения изображения 20×20 точек требуется $20 \times 20 \times 8 = 3200$ бит или 400 байт ($3200/8 = 400$).

Ответ: для хранения изображения потребуется 400 байт.

Задача 5. Отправлено SMS-сообщение:

А не могу без тебя жить!
 Мне и в дожди без тебя – сушь,
 Мне и в жару без тебя – стыть,
 Мне без тебя и Москва – глушь.

В мобильном телефоне адресата установлено ограничение размера входящего SMS-сообщения 64 байтами (при превышении этого размера сообщение автоматически делится на части). Каждый символ кодируется 16 битами. На сколько частей будет разбито сообщение?

Решение. Всего символов в сообщении 114. Так как каждый символ кодируется 16 битами (2 байтами), то сообщение занимает $114 \times 2 = 228$ байт. Теперь вычислим, на сколько частей будет разбито сообщение: $228/64 = 3,56$.

Ответ: сообщение будет разбито на 4 части.

Задача 6. Сообщение содержит 4096 символов. Объем сообщения при использовании равномерного кода составил 1/512 Мбайт. Какова мощность алфавита, с помощью которого записано сообщение?

Решение.

Определим, какой объем памяти (в битах) занимает один символ. Для этого переведем 1/512 Мбайт в биты и полученный результат разделим на число символов, содержащееся в сообщении:

$$\frac{1024 \cdot 1024 \cdot 8}{512 \cdot 4096} = \frac{2^{10} \cdot 2^{10} \cdot 2^3}{2^9 \cdot 2^{12}} = \frac{2^{23}}{2^{21}} = 4 \text{ бита.}$$

Мощность алфавита $N = 2^4 = 16$ символов.

Ответ: мощность алфавита 16 символов.

Задача 7. Скорость передачи данных равна 256 000 бит/сек. Передача файла заняла 4 минуты. Определить размер файла в Кбайтах.

Решение. Определим размер файла как произведение скорости передачи на время:

$$256\,000 \times 4 \times 60 \text{ бит} = 256\,000 \times 4 \times 60/8/1024 \text{ Кбайт} = 7500 \text{ Кбайт.}$$

Ответ. размер файла составляет 7500 Кбайт.

Задачи для самостоятельного решения

1. В корзине лежат 16 шаров. Все шары разного цвета. Сколько информации несет сообщение о том, что из корзины выкатился красный шар?
2. В коробке 5 синих и 15 красных шариков. Какое количество информации несет сообщение, что из коробки достали синий шарик?
3. В корзине лежат шары. Все разного цвета. Сообщение о том, что достали синий шар, несет 5 бит информации. Сколько всего шаров в корзине?
4. В коробке находятся кубики трех цветов: красного, желтого и зеленого, причем желтых в два раза больше красных, а зеленых на 6 больше, чем желтых. Сообщение о том, что из коробки случайно вытащили желтый кубик, содержало 2 бита информации. Сколько было зеленых кубиков?
5. Книга состоит из 64 страниц. На каждой странице 256 символов. Какой объем информации содержится в книге, если используемый алфавит состоит из 32 символов?
6. Имеется 2 текста на разных языках. Первый текст использует 32-символьный алфавит и содержит 200 символов, второй – 16-символьный алфавит и содержит 250 символов. Какой из текстов содержит большее количество информации и на сколько бит?
7. Сколько символов в тексте, если мощность алфавита – 64 символа, а объем информации, содержащейся в нем, 1,5 Кбайт?
8. Скорость передачи данных равна 128 000 бит/с. Через данное соединение передают файл размером 625 Кбайт. Определите время передачи файла в секундах.