Предисловие

Классическая теория информации является одной из базовых дисциплин, необходимых для подготовки специалистов различных специальностей и направлений. Успешное освоение данной дисциплины невозможно без дополнения теоретического материала практическими заданиями.

Цель практикума — помочь найти практические применения основных положений теории информации: научить определять энтропийные характеристики информации, информационные потери в каналах связи с помехами, вычислять скорость передачи информации, пропускную способность каналов связи, строить оптимальные коды, обнаруживать и исправлять ошибки при различных методах передачи и обработки информации и т. д.

Материал учебного пособия разбит на семь глав. В каждую главу включены следующие разделы: принятые обозначения; необходимые определения, правила и свойства; основные формулы и алгоритмы для вычислений; пояснения и примечания; примеры решения задач; задачи для самостоятельного решения. Учебное пособие снабжено приложениями и списком литературы.

Следует отметить, что обозначения приведены только для основных параметров и переменных. Другие обозначения расшифровываются по мере их появления в тексте, там же приводятся пояснения индексов.

Если обозначения повторяются и несут одну и ту же смысловую нагрузку, то в последующих главах, где они встречаются, их расшифровка опускается со ссылкой на предыдущий текст.

Предлагаемое учебное пособие может быть использовано в качестве:

1) задачника или методического пособия по решению задач для вузов;

2) основы для построения ознакомительного курса по теории информации;

3) материала для самостоятельного изучения теории информации;

4) базиса (в сочетании с [1–5]) для теоретического курса.

1 ПОНЯТИЯ: ЭНТРОПИЯ И КОЛИЧЕСТВО ИНФОРМАЦИИ

1.1. Принятые обозначения

n — число символов в сообщении;

N — общее число неповторяющихся сообщений;

$\bar{H}$ — неопределенность, приходящаяся на символ первичного алфавита;

I — количество информации;

k — число сообщений;

m — число символов в алфавите;

$p(a_i)$ — безусловные вероятности появления символов в алфавите.

1.2. Необходимые определения, правила, свойства

Первичный алфавит — набор символов (качественных признаков), которыми записывают передаваемое сообщение.

Вторичный алфавит — набор символов, с помощью которых сообщение переводят в код.

Сообщение — последовательность выбранных символов, отображающих определенную информацию.

Информация — интересующая получателя сообщения совокупность сведений о каких-либо явлениях, объектах, событиях.

Энтропия — количество информации, приходящееся на одно элементарное сообщение источника, вырабатывающего статистически независимые сообщения.

Количество информации — логарифмическая мера вероятности появления конкретного сообщения.

1.3. Основные формулы и алгоритмы для вычислений

Общее число неповторяющихся сообщений, составленное из m символов комбинированием по n равновероятных и взаимонезависимых символов в сообщении,

$$N = m^n.$$

Неопределенность, приходящаяся на символ первичного (кодируемого) алфавита,

$$\bar{H} = \log m.$$

Количество информации

$$I = k\bar{H} \text{ бит.}$$

Количество информации в k сообщениях первичного алфавита (символы равновероятны и взаимонезависимы)

$$I = k \log_2 m \text{ бит.}$$

Энтропия на символ первичного алфавита для неравновероятных алфавитов

$$H = \sum_{i=1}^{m} p(a_i) \log_2 \frac{1}{p(a_i)} = -\sum_{i=1}^{m} p(a_i) \log_2 p(a_i) \text{ бит/символ.}$$

Количество информации в сообщении, составленном из k неравновероятных символов,

$$I = -k \sum_{i=1}^{m} p(a_i) \log_2 p(a_i) \text{ бит.}$$

1.4. Пояснения и примечания

Количество информации определяется исключительно характеристиками первичного алфавита.

Единицей измерения энтропии считают такое ее количество, которое содержит некоторое стандартное сообщение. Система, о которой передают сообщение, должна иметь вполне определенное число состояний (исходов), не менее двух, так как неопределенность системы, имеющей одно состояние, равна нулю. Единицы измерения энтропии определяют выбором основания логарифма в члене $\log_a p$.

Если в качестве основания логарифма взять число 2 и рассмотреть систему с двумя равновероятными состояниями, то неопределенность данной системы

$$H(X) = \log_2 n = \log_2 2 = 1 \text{ дв. ед.},$$

где n — число равновероятных состояний; X — набор состояний системы.

Эту единицу энтропии называют двоичной (бинарной) единицей и обозначают bit (бит).

Следовательно, за единицу измерения энтропии принимают энтропию простейшей системы X, которая имеет два равновозможных состояния.

Если взять логарифм по основанию 10, а в качестве физической системы рассмотреть систему с десятью равновероятными состояниями, то будет получена одна десятичная единица (дес. ед.) количества энтропии. Условное название этой единицы — dit (дит).

Иногда применяют натуральную единицу (нат. ед.) энтропии (условно — nit (нит)), если в качестве основания логарифма выбрано число e.

1.5. Примеры решения задач

Пример 1. Определите число способов передачи положения фигур на доске с нанесенными квадратными клетками размером 4×4 (клетки). Для каждого случая определите количество информации.

Решение. Пронумеруем все клетки доски размером 4×4 и осуществим передачу номеров клеток. При этом число качественных признаков $m = 16$. Пусть для передачи номера клетки достаточно одного сообщения. Вычислим количество информации:

$$I = n \log_2 m = 1 \log_2 16 = 4 \text{ бит.}$$

Укажем на доске необходимую клетку и передадим ее координаты по горизонтали и вертикали. Для этого достаточно четырех качественных признаков (четыре номера по горизонтали и четыре по вертикали), передавать будем два сообщения. Вычислим количество информации:

$$I = 2 \log_2 4 = 2 \cdot 2 = 4 \text{ бит.}$$

При передаче номера по горизонтали и вертикали двоичным кодом потребуется два качественных признака, которые комбинируют по два элемента в сообщении. Вычислим количество информации:

$$I = 2 \log_2 2^2 = 2 \cdot 2 \log_2 2 = 4 \text{ бит.}$$

Пример 2. Пусть алфавит состоит из букв А, Б, В, Г. При этом безусловные вероятности появления букв пусть будут равны соответственно $p(A) = p(\text{Б}) = 0{,}35$; $p(\text{В}) = 0{,}21$; $p(\Gamma) = 0{,}12$. Вычислите количество информации на символ сообщения, которое можно составить из данного алфавита.

Решение. Энтропия данного алфавита — количество информации на символ алфавита. В нашем случае символы алфавита неравновероятны, поэтому энтропия равна

$$H = -\sum_{i=1}^{m} p(a_i) \log_2 p(a_i) =$$
$$= -(2 \cdot 0{,}35 \log_2 0{,}35 + 0{,}21 \log_2 0{,}21 + 0{,}12 \log_2 0{,}12) =$$
$$= -(2 \cdot 0{,}35(-1{,}51457) + 0{,}21(-2{,}25154) + 0{,}12(-3{,}05889) =$$
$$= 1{,}9 \text{ бит/символ.}$$

Пример 3. Пусть задан равномерный трехзначный троичный код. Определите количество информации при получении 6 сообщений.

Решение. В нашей задаче $m = 3$ — число качественных признаков; $n = 3$ — число комбинаций качественных признаков в коде; $N = m^n = 3^3$ — число сообщений обозначенного в условии задачи кода; $H = \log_2 N = 3\log_2 3$ — энтропия одного закодированного сообщения. Далее определим количество информации в 6 сообщениях:

$$I = 6 \cdot 3\log_2 3 = 18 \cdot 1{,}584 = 28{,}52933 \text{ бит.}$$

Пример 4. Пусть информация хранится в 262144 стандартных ячейках. Определите число вариантов передачи сведений о том, из какой ячейки можно извлечь необходимые данные. Вычислите количество информации для каждого случая. Осуществите выбор геометрического построения хранилища, которое будет передавать информацию с минимальным числом качественных признаков.

Решение. Из условия задачи следует, что ячейки имеют одинаковые размеры, пронумеруем их и выполним передачу номера ячейки (см. пример 1):

$$I = n\log_2 m = 1\log_2 262144 = 18 \text{ бит.}$$

Как в примере 1, передадим номер ячейки по горизонтали и вертикали, предварительно расположив их квадратом. При этом число сообщений равно двум. Тогда

$$I = n\log_2 m = 2\log_2 \sqrt{262144} = 2 \cdot 9 = 18 \text{ бит.}$$

Второй способ расположения ячеек — куб. В данном случае будем передавать три координаты. При этом число сообщений равно трем. Тогда

$$I = 3\log_2 \sqrt[3]{262144} = 3 \cdot 6 = 18 \text{ бит.}$$

Минимальное число качественных признаков при передаче информации будет обеспечивать геометрическое построение хранилища в виде куба.

Пусть информация хранится в 128 шкафах с каждой ячейкой в виде куба. Тогда в сообщении будем передавать отдельно номер ячейки в шкафу и номер шкафа. Далее определим общее количество информации:

$$I = I_1 + I_2.$$

Последним действием определим минимальное число качественных признаков:

$$I_1 = \log_2 128 = 7 \text{ бит;} \quad I_2 = \log_2 2048 = 11 \text{ бит;}$$

$$I = I_1 + I_2 = 7 + 11 = 18 \text{ бит.}$$