

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Нижегородский государственный архитектурно-строительный
университет»
(ННГАСУ)

Факультет инженерно-экологических систем и сооружений
Кафедра информационных систем и технологий

Отчет по лабораторной работе
По дисциплине: Теории информации, данные, знания
Тема: Изучение методов эффективного кодирования
Вариант 4

Выполнил студент группы ИС-29:

Гордеев Н. С.

(подпись)

Проверил кандидат технических наук, доцент:

Родькина О.Я.

(подпись)

Нижний Новгород

2023 г.

Цель работы и задание

Изучение принципа эффективного кодирования источника дискретных сообщений на примере нескольких методов и сравнение их с равномерным кодированием.

Задание:

1. Изучить принцип эффективного кодирования источника дискретных сообщений методами Шеннона - Фано и Хаффмена.
2. Осуществить кодирование каждого сообщения алфавита (см. таблицу 1) последовательно тремя различными способами, используя двоичный код:
 - равномерный;
 - код Шеннона – Фано;
 - код Хаффмена, в соответствии с заданным вариантом.

| |
|------|
| 4 |
| 0,28 |
| 0,04 |
| 0,16 |
| 0,02 |
| 0,13 |
| 0,07 |
| 0,30 |

3. Определить значения $H_{\max}(x)$, $H(x)$ и $l_{p.k.}$ –длину кода при равномерном кодировании;
4. Для каждого из эффективных методов кодирования рассчитать значения \bar{l} - среднюю длину кода элементарного сообщения.
5. Для каждого из эффективных методов кодирования рассчитать значения K_{cc} и $K_{оз}$

Основная идея методов эффективного кодирования

Равномерный код – такой код, когда все символы какого-либо алфавита кодируются кодами одинаковой длины.

Код Шеннона – Фано.

Прежде всего, буквы (или любые сообщения, подлежащие кодированию) исходного алфавита записывают в порядке убывающей вероятности. Упорядоченное таким образом множество букв разбивают так, чтобы суммарные вероятности этих подмножеств были примерно равны. Всем знакам (буквам) верхней половины в качестве первого символа присваивают кодовый элемент 1, а всем нижним 0. Затем каждое подмножество снова разбивается на два подмножества с соблюдением того же условия равенства вероятностей и с тем же условием присваивания кодовых элементов в качестве второго символа. Такое разбиение продолжается до тех пор, пока в подмножестве не окажется только по одной букве кодируемого алфавита

Код Хаффмана.

Буквы алфавита сообщений выписывают в основной столбец таблицы в порядке убывания вероятностей. Две последние буквы объединяют в одну вспомогательную букву, которой приписывают суммарную вероятность. Вероятности букв, не участвующих в объединении и полученная суммарная вероятность снова располагаются в порядке убывания вероятностей в дополнительном столбце, а две последние объединяются. Процесс продолжается до тех пор, пока не получат единственную вспомогательную букву с вероятностью, равной единице.

Двоичное кодирование равномерным методом.

N1

Равномерный метод

$$H_{\text{MAX}} = \log_2 N$$

l.p.k. = 3 бита

$$H_{\text{MAX}} = \log_2 4 = 2,807$$

$$H(X) = -\sum_{i=1}^n p_i \cdot \log_2 p_i = 0,28 \log_2 0,28 + 0,04 \log_2 0,04$$

$$+ 0,76 \log_2 0,76 + 0,02 \log_2 0,02 + 0,73 \log_2 0,73 + 0,07 \log_2 0,07$$

$$+ 0,30 \log_2 0,30 = 2,38 \text{ бита}$$

| a_1 | a_2 | a_3 | a_4 | a_5 | a_6 | a_7 |
|-------|-------|-------|-------|-------|-------|-------|
| 000 | 001 | 010 | 011 | 100 | 101 | 110 |

Двоичное кодирование методом Шеннона-Фано. Расчет показателей эффективности.

№2

| a_i | p_i | № разбиения | | | | | l_i |
|-------|-------|-------------|---|---|---|---|-------|
| | | 1 | 2 | 3 | 4 | 5 | |
| a_4 | 0,30 | 0 | 0 | | | | 2 |
| a_1 | 0,28 | 0 | 1 | | | | 2 |
| a_3 | 0,16 | 1 | 0 | | | | 2 |
| a_5 | 0,13 | 1 | 1 | 0 | | | 3 |
| a_6 | 0,07 | 1 | 1 | 1 | 0 | | 4 |
| a_2 | 0,04 | 1 | 1 | 1 | 1 | 0 | 5 |
| a_4 | 0,02 | 1 | 1 | 1 | 1 | 1 | 5 |

$$\bar{l} = \sum_{i=1}^k p(a_i) \cdot l_i =$$

$$= 0,30 \cdot 2 + 0,28 \cdot 2 +$$

$$+ 0,16 \cdot 2 + 0,13 \cdot 3 +$$

$$+ 0,07 \cdot 4 + 0,04 \cdot 5 +$$

$$+ 0,02 \cdot 5 = 2,45$$

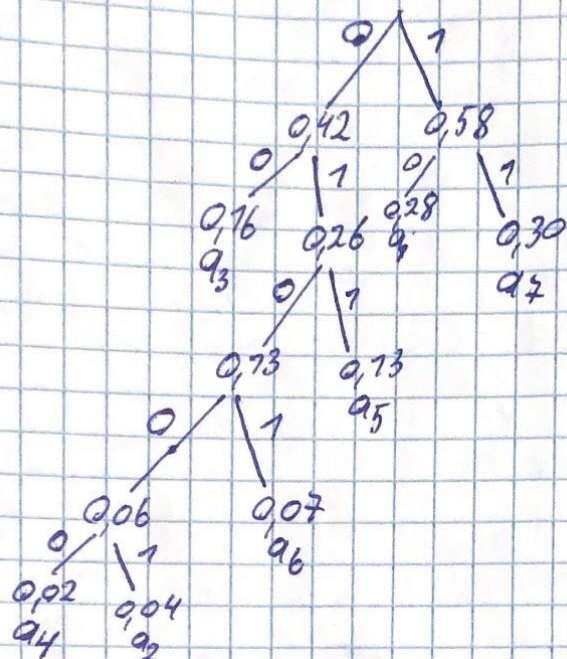
$$K_{ссл} = \frac{l_{р.к}}{\bar{l}} = \frac{3}{2,45} \approx 1,22$$

$$K_{од} = \frac{H(A)}{\bar{l}} = \frac{2,38}{2,45} \approx 0,97$$

Двоичное кодирование методом Хаффмана. Расчет показателей эффективности.

№ 3

| a_i | p_i | Выполняем таблицу | | | | | | Код | l_i |
|-------|-------|-------------------|------|------|------|------|---|-------|-------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | | |
| a_7 | 0,30 | 0,30 | 0,30 | 0,30 | 0,42 | 0,58 | 1 | 11 | 2 |
| a_1 | 0,28 | 0,28 | 0,28 | 0,28 | 0,30 | 0,42 | | 10 | 2 |
| a_3 | 0,16 | 0,16 | 0,16 | 0,26 | 0,28 | | | 00 | 2 |
| a_5 | 0,13 | 0,13 | 0,13 | 0,16 | | | | 012 | 3 |
| a_6 | 0,07 | 0,07 | 0,13 | | | | | 0101 | 4 |
| a_2 | 0,04 | 0,06 | | | | | | 01001 | 5 |
| a_4 | 0,02 | | | | | | | 01000 | 5 |



$$\bar{l} = 2,45$$

$$K_{cc} = \frac{k_{р.к}}{\bar{l}} = \frac{9}{2,45} \approx 3,67$$

$$K_{од} = \frac{H(A)}{\bar{l}} = \frac{2,38}{2,45} \approx 0,97$$

Выводы об эффективности использования выбранных кодов для данного источника.

Вычисления по данной лабораторной работе показали, что для кодирования данного источника нужно использовать коды Хаффмана или Шеннона-Фано.