

Лабораторная работа №5 «Изучение методов Шеннона – Фано и Хаффмана по построению эффективных кодов»

1 Цель работы

- изучить возможности эффективного кодирования информации по методам Шеннона – Фано и Хаффмана.

2 Основные теоретические сведения

Кодирование, при котором обеспечивается минимальная средняя длина кодовых слов, называется **эффективным** (оптимальным). В эффективном коде символу, встречающемуся чаще всего, присваивается наиболее короткая кодовая комбинация.

Задачи эффективного кодирования заключаются в следующем:

1) Запоминание максимального количества информации в ограниченной памяти.

2) Обеспечение максимальной пропускной способности канала связи.

Эффективное кодирование базируется на теореме Шеннона о кодировании при отсутствии помех, согласно которой минимальная средняя длина кодовых слов определяется соотношением:

$$\bar{L}_{\min} = \frac{H}{\log_2 K};$$

где H – энтропия источника сообщений, K – основание кода. Для двоичного кода, очевидно, что $L_{\min} = H$.

Эффективностью кода x называется отношение L_{\min} к реально достигнутой в данном коде средней длине кодовых слов \bar{L} :

$$\chi = \frac{\bar{L}_{\min}}{\bar{L}} = \frac{H}{\bar{L} \log_2 K}.$$

Средняя длина кодовых комбинаций может быть найдена следующим образом:

$$\bar{L} = \sum_{i=1}^m p(s_i) L_i;$$

где L_i - длина кодовой комбинации, соответствующей символу s_i из алфавита размером m .

Для случая отсутствия статистической взаимосвязи между символами метод построения эффективных кодов впервые был предложен Шенноном и Фано.

Для двоичного кода метод Шеннона-Фано сводится к следующему:

1) Буквы алфавита располагаются в порядке убывания вероятностей.

2) Алфавит букв разбивается на две группы таким образом, чтобы суммарные вероятности букв обеих групп были по возможности равны. Первой группе присваивается символ 1, второй символ – 0.

3) Каждую из образованных групп вновь делят на две части с приблизительно равными суммарными вероятностями и присваивают им 1 и 0. Таким образом, получают вторые цифры кода.

4) Процесс повторяется до тех пор, пока в каждой подгруппе не останется по одной букве.

Рассмотрим пример построения кода Шеннона-Фано для алфавита из шести символов (таблица 1). Кодовое дерево, соответствующее полученному коду Шеннона-Фано, представлено на рисунке 1.

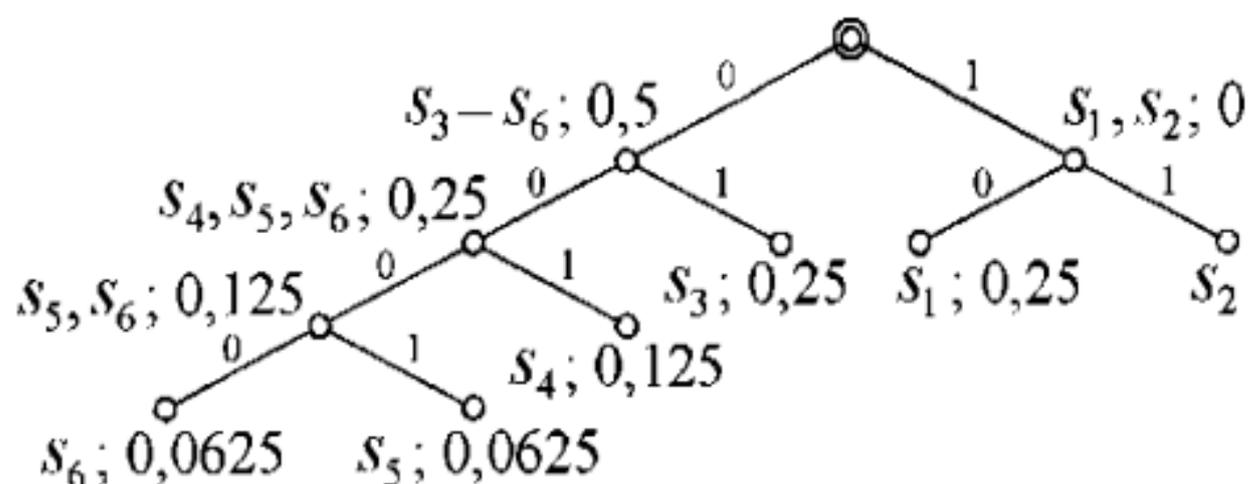


Рисунок 1 – Кодовое дерево для кода Шеннона-Фано

Таблица 1 – Получение эффективного кода по методу Шеннона – Фано

Символы исходного алфавита	$P_{i,j}$	Разделение символов на группы				Кодовые слова	
		1	2	3	4		
S_1	0,25	} I-1	I-1			11	
S_2	0,25		II-0			10	
S_3	0,25	} II-0	III-1			01	
S_4	0,125		} IV-0		IV-1		001
S_5	0,0625					V-1	0001
S_6	0,0625				- 0	VI-0	0000

Метод Шеннона-Фано не всегда приводит к однозначному построению кода. От указанного недостатка свободен метод Хаффмана.

Для двоичного кода метод Хаффмана сводится к следующему:

1) Буквы алфавита выписываются в столбец в порядке убывания вероятностей.

2) Две последние буквы объединяются в одну вспомогательную букву, которой приписывается суммарная вероятность.

3) Вероятности букв, участвующих в объединении и полученная суммарная вероятность вновь располагаются в порядке убывания вероятностей в дополнительном столбце, а две последние буквы объединяются.

4) Процесс продолжается до тех пор, пока не будет получена единственная вспомогательная буква с суммарной вероятностью, равной 1.

Для получения кодовой комбинации, соответствующей данной букве необходимо проследить путь перехода по строкам и столбцам таблицы.

Рассмотрим пример построения кода Хаффмана для алфавита из восьми символов (таблица 2). Результат показан на рисунке 2 в виде кодового дерева соответствующего коду Хаффмана.

Таблица 2 – Получение эффективного кода по методу Хаффмана

Символы	Вероятности	Вспомогательные столбцы			
		1	2	3	4
s_1	0,4	0,4	0,4	0,6	1
s_2	0,25	0,25	0,35	0,4	
s_3	0,2	0,2	0,25		
s_4	0,1	0,15			
s_5	0,05				

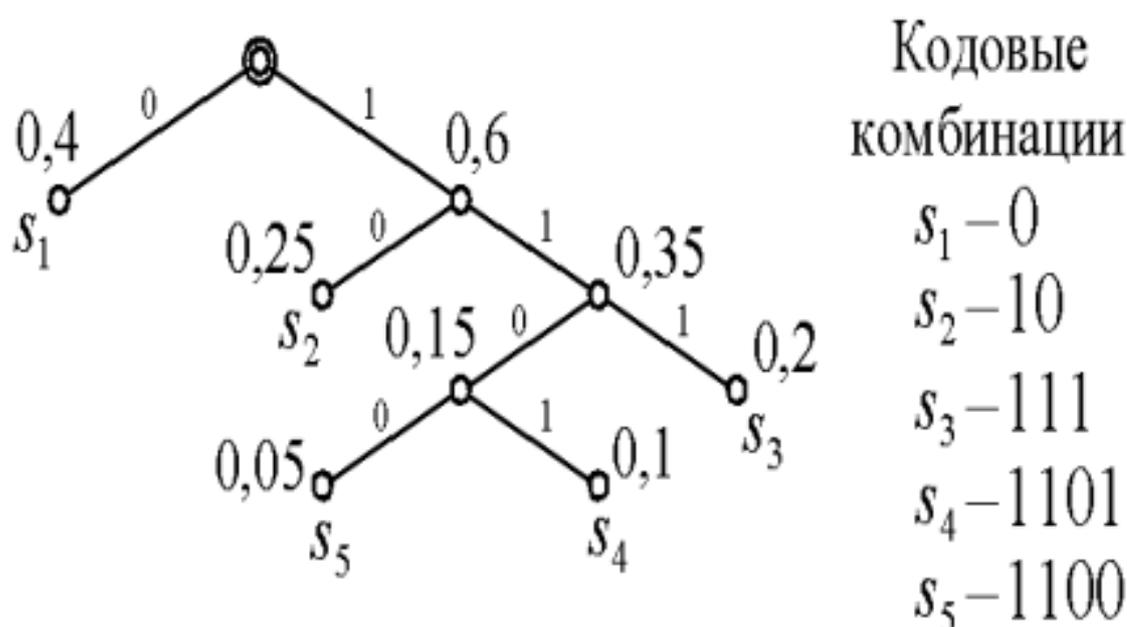


Рисунок 2 – Пример кодового дерева для кода Хаффмана

3 Порядок выполнения работы

Данная лабораторная работа предполагает выполнение следующих этапов:

- 1) Изучить методические указания к лабораторной работе.
- 2) Пройти собеседование с преподавателем и получить задание для выполнения работы.
- 3) На основе заданного первичного алфавита и вероятностей появления символов этого алфавита (табл. 3.3) получить в форме таблицы двоичный код Шеннона-Фано.
- 4) Построить кодовое дерево для полученного кода Шеннона-Фано.

5) Определить эффективность кода, полученного по методу Шеннона-Фано.

6) Выполнить пункты 3, 4, 5 для метода Хаффмана.

7) Сравнить эффективности методов Шеннона-Фано и Хаффмана. Сделать выводы о полученных результатах.

8) Оформить и защитить отчет по выполнению лабораторной работы.

4 Варианты заданий

Таблица 3 – Вероятности появления символов для различных вариантов

Вар.	Символы алфавита источника сообщений											
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
1	0,14	0,06	0,05	0,08	0,13	0,04	0,01	0,09	0,15	0,02	0,11	0,12
2	0,11	0,05	0,09	0,10	0,12	0,03	0,02	0,08	0,15	0,07	0,14	0,04
3	0,13	0,07	0,05	0,06	0,15	0,04	0,11	0,02	0,12	0,16	0,08	0,01
4	0,02	0,11	0,12	0,01	0,09	0,15	0,08	0,13	0,04	0,14	0,06	0,05
5	0,07	0,14	0,04	0,02	0,08	0,15	0,10	0,12	0,03	0,11	0,05	0,09
6	0,16	0,08	0,01	0,11	0,02	0,12	0,06	0,15	0,04	0,13	0,07	0,05
7	0,01	0,09	0,15	0,02	0,11	0,12	0,14	0,06	0,05	0,08	0,13	0,04
8	0,02	0,08	0,15	0,07	0,14	0,04	0,13	0,07	0,05	0,06	0,15	0,04
9	0,11	0,02	0,12	0,16	0,08	0,01	0,07	0,05	0,13	0,06	0,15	0,04
10	0,06	0,05	0,14	0,13	0,04	0,08	0,15	0,01	0,09	0,12	0,02	0,11
11	0,09	0,11	0,05	0,03	0,10	0,12	0,15	0,02	0,08	0,14	0,04	0,07
12	0,05	0,13	0,07	0,15	0,04	0,06	0,02	0,12	0,11	0,04	0,07	0,14
13	0,12	0,02	0,11	0,09	0,15	0,01	0,13	0,04	0,08	0,06	0,05	0,14
14	0,14	0,04	0,07	0,15	0,02	0,08	0,03	0,10	0,12	0,09	0,11	0,05
15	0,04	0,07	0,14	0,02	0,12	0,11	0,15	0,04	0,06	0,05	0,13	0,07
16	0,05	0,07	0,13	0,04	0,15	0,06	0,12	0,02	0,11	0,01	0,08	0,16
17	0,1	0,03	0,05	0,09	0,14	0,04	0,01	0,08	0,16	0,04	0,12	0,14
18	0,12	0,07	0,08	0,11	0,16	0,01	0,04	0,06	0,13	0,09	0,1	0,03
19	0,11	0,08	0,07	0,04	0,14	0,05	0,13	0,02	0,1	0,15	0,09	0,02
20	0,03	0,12	0,14	0,02	0,08	0,15	0,1	0,11	0,03	0,12	0,05	0,05
21	0,08	0,13	0,05	0,01	0,06	0,14	0,11	0,13	0,06	0,1	0,04	0,09
22	0,15	0,09	0,02	0,13	0,02	0,1	0,08	0,16	0,01	0,12	0,06	0,06
23	0,03	0,11	0,16	0,05	0,14	0,15	0,09	0,01	0,04	0,07	0,13	0,02
24	0,06	0,01	0,12	0,09	0,16	0,02	0,11	0,03	0,08	0,05	0,15	0,12