

Лабораторная работа №4 «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.

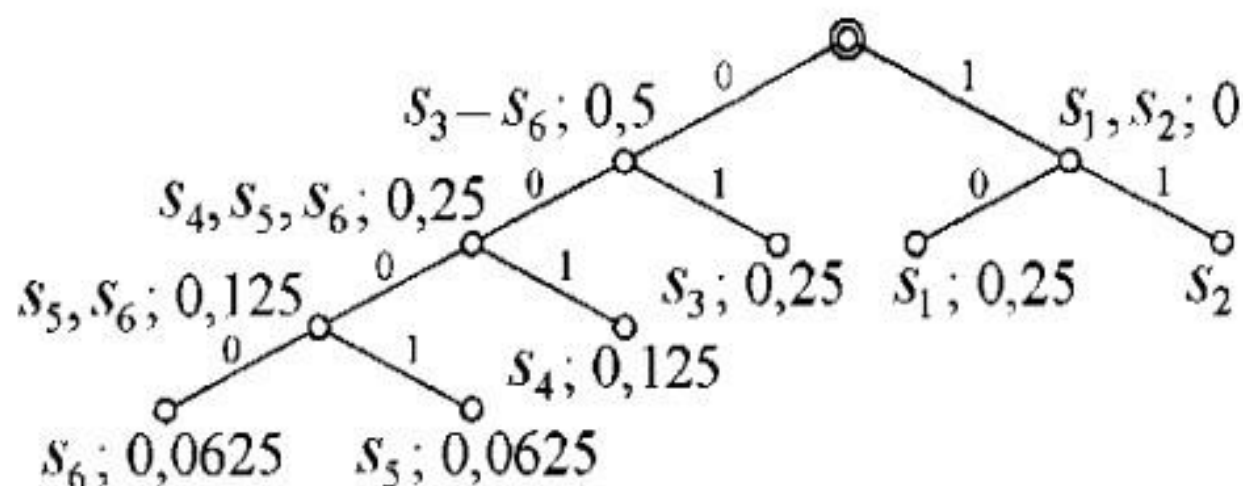


Рисунок 1 – Кодовое дерево для кода Шеннона-Фано

Таблица 1 – Получение эффективного кода по методу Шеннона – Фано

Символы исходного алфавита	P_{iS_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				- 0	V-1	0001
S_6	0,0625					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.

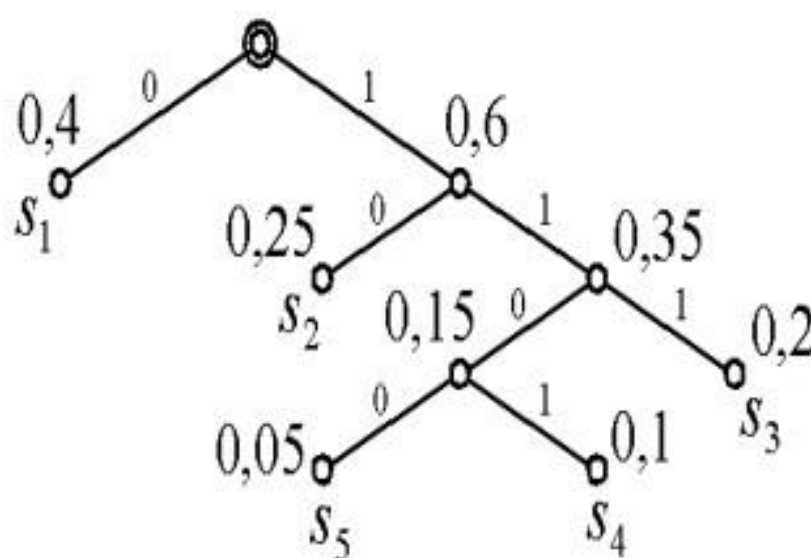
4) Процесс продолжается до тех пор, пока не будет получена единственная вспомогательная буква с суммарной вероятностью, равной 1.

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				



Кодовые комбинации

$s_1 - 0$

$s_2 - 10$

$s_3 - 111$

$s_4 - 1101$

$s_5 - 1100$

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

Задание 1.

1. И прыгают скороговорки, как караси на сковородке.
2. Карл у Клара украл кораллы, а Клара у Карла украла кларнет.
3. Мерчендайзеры соврали — сорван сэмплинг самоваров!
4. Ядро потребителей пиастров — пираты, а пиратов — пираньи.
5. Полосу про паласы заменили двумя полуполосами про пылесосы.
6. Невелик на ситиборде бодибилдера бицепс.
7. Скреативлен креатив не по-креативному, нужно перекреативить!
8. Брейнштурм: гам, гром, ор ртов, пир рифм, вдруг — бум! Блеск!
9. Выборка по уборщицам на роллс-ройсах нерепрезентативна.
10. Банкиров ребрендрили-ребрендрили-ребрендрили, да не выребрендировали.
11. В Каннах львы только ленивым венки не вили.
12. В Кабардино-Балкарии валокордин из Болгарии.
13. Деидеологизировали-деидеологизировали, и додеидеологизировались.
14. Их пестициды не перепестицидят наши по своей пестицидности.
15. Кокосовары варят в скорококосоварках кокосовый сок.
16. Рапортовал, да не дорапортовал; дорапортовал, да зарапортовался.
17. Регулировщик лигуриец регулировал в Лигурии.
18. У нас во дворе-подворье погода размокропогодилась.
19. У Сени и Сани в сенях сом с усами.
20. Самшит, самшит, как ты крепко сшит.
21. Черной ночью черный кот прыгнул в черный дымоход.
22. Тридцать три корабля лавировали-лавировали, да не вылавировали.