Лабораторная работа №3

«Оптимальное кодирование. Алгоритм Шеннона-Фано»

Вариант 1

Цель работы: освоить алгоритм Шеннона-Фано. Научиться сжимать сообщения с помощью алгоритма Шеннона-Фано.

1

Условие: Составить коды для каждого символа алфавита с помощью алгоритма Шеннона-Фано.

a	A	B	C	D	E	F
p	0,25	0,05	0,1	0,25	0,05	0,3

Решение:

Сперва упорядочим входную последовательность символов по невозрастанию их вероятностей:

a	F	A	D	C	B	E
p	0,3	0,25	0,25	0,1	0,05	0,05

Далее найдём коды для каждого символа используя следующий алгоритм (Шеннона-Фано):

1. Символы полученного алфавита делим на две части, суммарные вероятности символов которых максимально близки друг к другу.
2. Для текущего разряда первой части присваиваем код «0», для второй – «1».
3. Полученные части рекурсивно делим и их частям назначаем соответствующие двоичные цифры (шаги 1–2).

Адаменко

a	p	Разряды				Код
		1	2	3	4	
F	0,3	0	0			00
A	0,25		1			01
D	0,25	1	0			10
C	0,1		1	0		110
B	0,05			1	0	1110
E	0,05				1	1111

2

Условие: Закодировать сообщение (ADADBDCBBDCABFDAFCEB), используя коды для символов. Вычислить среднюю длину символа. Вычислить энтропию алфавита. Сравнить среднюю длину и энтропию. Сделать выводы.

Решение:

Закодируем сообщение используя найденные выше коды:

ADADBDCBBDCABFDAFCEB =

01100110111010110111011101011001111000100100110111111110

Вычислим среднюю длину символа, то есть среднюю длину его оптимального неравномерного кода:

$$L_{ср} = \Sigma\, L_i \times p_i = 2 \times 0{,}3 + 2 \times 0{,}25 + 2 \times 0{,}25 + 3 \times 0{,}1 + 4 \times 0{,}05 + 4 \times 0{,}05 = \\ = 2{,}3 \text{ бит}$$

Подсчитаем энтропию алфавита:

$$H = -\Sigma\, p_i \times \log_2(p_i) = 0{,}3 \times \log_2(0{,}3) + 0{,}25 \times \log_2(0{,}25) + 0{,}25 \times \log_2(0{,}25) + \\ + 0{,}1 \times \log_2(0{,}1) + 0{,}05 \times \log_2(0{,}05) + 0{,}05 \times \log_2(0{,}05) = 2{,}285 \text{ бит/сим}$$

Эффективность оптимального кодирования тем выше, чем больше средняя длина кода (символа) стремится к энтропии алфавита. Высчитаем коэффициент эффективности:

$$K_э = \frac{H}{L_{ср}} = \frac{2{,}285}{2{,}3} = 0{,}993$$

Поскольку коэффициент эффективности недалёк от единицы, данное кодирование действительно оптимально и эффективно.

Адаменко

3

Условие: Составить список биграмм для данного алфавита. Вычислить вероятность каждой биграммы. Составить коды для каждой биграммы с помощью алгоритма Шеннона-Фано.

Решение:

Составим список биграмм:

AA, AB, AC, AD, AE, AF, BA, BB, BC, BD, BE, BF, CA, CB, CC, CD, CE, CF, DA, DB, DC, DD, DE, DF, EA, EB, EC, ED, EE, EF, FA, FB, FC, FD, FE, FF

Вычислим вероятность каждой биграммы:

$$P(AA) = P(A) \times P(A) = 0{,}25 \times 0{,}25 = 0{,}0625$$
$$P(AB) = P(A) \times P(B) = 0{,}25 \times 0{,}05 = 0{,}0125$$
$$P(AC) = P(A) \times P(C) = 0{,}25 \times 0{,}1 = 0{,}025$$
$$P(AD) = P(A) \times P(D) = 0{,}25 \times 0{,}25 = 0{,}0625$$
$$P(AE) = P(A) \times P(E) = 0{,}25 \times 0{,}05 = 0{,}0125$$
$$P(AF) = P(A) \times P(F) = 0{,}25 \times 0{,}3 = 0{,}075$$
$$P(BA) = P(B) \times P(A) = 0{,}05 \times 0{,}25 = 0{,}0125$$
$$P(BB) = P(B) \times P(B) = 0{,}05 \times 0{,}05 = 0{,}0025$$
$$P(BC) = P(B) \times P(C) = 0{,}05 \times 0{,}1 = 0{,}005$$
$$P(BD) = P(B) \times P(D) = 0{,}05 \times 0{,}25 = 0{,}0125$$
$$P(BE) = P(B) \times P(E) = 0{,}05 \times 0{,}05 = 0{,}0025$$
$$P(BF) = P(B) \times P(F) = 0{,}05 \times 0{,}3 = 0{,}015$$
$$P(CA) = P(C) \times P(A) = 0{,}1 \times 0{,}25 = 0{,}025$$
$$P(CB) = P(C) \times P(B) = 0{,}1 \times 0{,}05 = 0{,}005$$
$$P(CC) = P(C) \times P(C) = 0{,}1 \times 0{,}1 = 0{,}01$$
$$P(CD) = P(C) \times P(D) = 0{,}1 \times 0{,}25 = 0{,}025$$
$$P(CE) = P(C) \times P(E) = 0{,}1 \times 0{,}05 = 0{,}005$$
$$P(CF) = P(C) \times P(F) = 0{,}1 \times 0{,}3 = 0{,}03$$
$$P(DA) = P(D) \times P(A) = 0{,}25 \times 0{,}25 = 0{,}0625$$
$$P(DB) = P(D) \times P(B) = 0{,}25 \times 0{,}05 = 0{,}0125$$
$$P(DC) = P(D) \times P(C) = 0{,}25 \times 0{,}1 = 0{,}025$$
$$P(DD) = P(D) \times P(D) = 0{,}25 \times 0{,}25 = 0{,}0625$$
$$P(DE) = P(D) \times P(E) = 0{,}25 \times 0{,}05 = 0{,}0125$$
$$P(DF) = P(D) \times P(F) = 0{,}25 \times 0{,}3 = 0{,}075$$
$$P(EA) = P(E) \times P(A) = 0{,}05 \times 0{,}25 = 0{,}0125$$
$$P(EB) = P(E) \times P(B) = 0{,}05 \times 0{,}05 = 0{,}0025$$
$$P(EC) = P(E) \times P(C) = 0{,}05 \times 0{,}1 = 0{,}005$$
$$P(ED) = P(E) \times P(D) = 0{,}05 \times 0{,}25 = 0{,}0125$$
$$P(EE) = P(E) \times P(E) = 0{,}05 \times 0{,}05 = 0{,}0025$$

$P(EF) = P(E) \times P(F) = 0{,}05 \times 0{,}3 = 0{,}015$
$P(FA) = P(F) \times P(A) = 0{,}3 \times 0{,}25 = 0{,}075$
$P(FB) = P(F) \times P(B) = 0{,}3 \times 0{,}05 = 0{,}015$
$P(FC) = P(F) \times P(C) = 0{,}3 \times 0{,}1 = 0{,}03$
$P(FD) = P(F) \times P(D) = 0{,}3 \times 0{,}25 = 0{,}075$
$P(FE) = P(F) \times P(E) = 0{,}3 \times 0{,}05 = 0{,}015$
$P(FF) = P(F) \times P(F) = 0{,}3 \times 0{,}3 = 0{,}09$

Составим коды для каждой биграммы с помощью алгоритма Шеннона-Фано (аналогично заданию №2):

aa	p	Разряды									Код
		1	2	3	4	5	6	7	8	9	
FF	0,0900	0	0	0							000
AF	0,0750			1	0						0010
DF	0,0750				1						0011
FA	0,0750		1	0	0						0100
FD	0,0750				1						0101
AA	0,0625			1	0						0110
AD	0,0625				1						0111
DA	0,0625	1	0	0	0						1000
DD	0,0625				1						1001
CF	0,0300			1	0	0					10100
FC	0,0300					1					10101
AC	0,0250				1	0					10110
CA	0,0250					1					10111
CD	0,0250		1	0	0	0					11000
DC	0,0250					1	0				110010
BF	0,0150						1				110011
EF	0,0150				1	0	0				110100
FB	0,0150						1				110101
FE	0,0150					1	0				110110
AB	0,0125						1				110111
AE	0,0125			1	0	0	0				111000
BA	0,0125						1				111001
BD	0,0125					1	0				111010
DB	0,0125						1	0			1110110

aa	p	Разряды									Код
		1	2	3	4	5	6	7	8	9	
DE	0,0125	1	1	1	0	1	1	1			1110111
EA	0,0125				1	0	0				111100
ED	0,0125						1	0			1111010
CC	0,0100							1			1111011
BC	0,0050					1	0	0			1111100
CB	0,0050							1	0		11111010
CE	0,0050								1		11111011
EC	0,0050						1	0	0		11111100
BB	0,0025								1		11111101
BE	0,0025							1	0		11111110
EB	0,0025								1	0	111111110
EE	0,0025									1	111111111

4

Условие: Закодировать сообщение, используя коды для биграмм. Вычислить среднюю длину биграммы. Разделить результат на 2. Сравнить полученное число со средней длиной для посимвольного кодирования. Сделать выводы о целесообразности кодировать сообщения поблочно.

Решение:

Закодируем сообщение используя найденные выше коды биграмм:
ADADBDCBBDCABFDAFCEB =
01110111111010111110101110101011111001110001010111111110

Вычислим среднюю длину биграммы, то есть среднюю длину её оптимального неравномерного кода:

$$L_{ср} = \Sigma\, L_i \times p_i = 3 \times 0{,}09 + 4 \times 0{,}075 + 4 \times 0{,}075 + 4 \times 0{,}075 + 4 \times 0{,}075 +$$
$$+ 4 \times 0{,}0625 + 4 \times 0{,}0625 + 4 \times 0{,}0625 + 4 \times 0{,}0625 + 5 \times 0{,}03 + 5 \times 0{,}03 +$$
$$+ 5 \times 0{,}025 + 5 \times 0{,}025 + 5 \times 0{,}025 + 6 \times 0{,}025 + 6 \times 0{,}015 + 6 \times 0{,}015 +$$
$$+ 6 \times 0{,}015 + 6 \times 0{,}015 + 6 \times 0{,}0125 + 6 \times 0{,}0125 + 6 \times 0{,}0125 + 6 \times 0{,}0125 +$$
$$+ 7 \times 0{,}0125 + 7 \times 0{,}0125 + 6 \times 0{,}0125 + 7 \times 0{,}0125 + 7 \times 0{,}01 + 7 \times 0{,}005 +$$
$$+ 8 \times 0{,}005 + 8 \times 0{,}005 + 8 \times 0{,}005 + 8 \times 0{,}0025 + 8 \times 0{,}0025 + 9 \times 0{,}0025 +$$
$$+ 9 \times 0{,}0025 = 4{,}5975 \text{ бит}$$

Разделим полученное число на 2:

$$\frac{4{,}5975}{2} = 2{,}29875 \text{ бит}$$

Поскольку число получилось меньше, чем при кодировании по символьно, можно сделать вывод, что кодировать поблочно выгоднее, чем посимвольно.

5

Условие: Создать подпрограмму для составления кодов для символов по алгоритму Шеннона-Фано. Подпрограмме передаётся набор символов и их вероятностей.

6

Условие: Создать подпрограмму для кодирования сообщения. Подпрограмме передаётся сообщение, состоящее из символов алфавита и коды для кодирования сообщения, полученные подпрограммой из предыдущего пункта.

7

Условие: Проверить работоспособность подпрограмм, данные из п. 1 и 2 использовать как тестовые.

8

Условие: Модернизировать подпрограммы из п. 5 и 6 для случая поблочного кодирования. Создать программу, на вход которой подаются символы алфавита и их вероятности. Далее пользователь вводит размер блока (от 1 символа). В случае с блоком размера 1, имеет место посимвольное кодирование. Иначе составляются различные возможные комбинации блоков и вычисляются их вероятности. Вычисляется энтропия и средняя длина 1 символа (в поблочном случае вычисляется энтропия и длина блока, и делится на размер блока), результат выводится на экран. Далее пользователь вводит сообщение, программа кодирует его и выводит результат. Выполнить программу для блоков различного размера. Установить размер блока, на котором средняя длина символа минимальна. Сделать выводы.

Решение:

Длина блока	2	3	4	5	6
Ср. длина символа	2,3	2,29	2,28	2,27	2,25

Вывод: чем больше длина блока, тем эффективнее кодирование.