

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное
учреждение
высшего образования
Национальный исследовательский университет ИТМО
Факультет Систем Управления и Робототехники

Лабораторная работа №1
по курсу «Прикладная теория информации»

«Алгоритмы эффективного кодирования»

Выполнили: Московский К.А.
Алексеева Ю.В.

Проверил: Краснов А.Ю.

Санкт-Петербург
2021 г.

1 Введение

Основной задачей эффективного кодирования является формирование таких кодов, которые обеспечивали бы наибольшую эффективность использования имеющихся ресурсов канальной среды. Одним из способов решения такой задачи является сокращение среднего числа разрядов кода на символ. Рассматриваемые далее методы эффективного кодирования базируются на основной теореме К. Шеннона.

2 Цель работы

Применить алгоритмы эффективного кодирования к искомому алфавиту с целью сокращения среднего числа разрядов кода на символ.

3 Условие

ИДИ генерирует алфавит:

$$X = \{x_i; i = \overline{1, 4}\}$$

со следующими вероятностями появления символов:

$$p(x_1) = 0.9, p(x_2) = 0.25, p(x_3) = p(x_4) = 0.025$$

4 Ход работы

4.1 Алгоритм эффективного кодирования К. Шеннона — Р. Фэнно

- Энтропия заданного ИДИ:

$$\begin{aligned} H(X) &= - \sum_{i=1}^4 p(x_i) \log_2 p(x_i) = \\ &= -(0.9 \times (-0.152) + 0.25 \times (-2) + 2 \times 0.025 \times (-5,322)) \\ &= 1.126297 \dots \approx 1.126 \text{ (бит/символ)} \end{aligned}$$

Таблица 1: Таблица формирования кода символов ИДИ

	$p(x_i)$	коды $K\{x_i\}$			длина $l\{K\{x_i\}\}$
x_1	0.7	1			1
x_2	0.25	0	1		2
x_3	0.0025	0	0	1	3
x_4	0.0025	0	0	0	3

Составим таблицу формирования эффективного кода символов ИДИ. (Таблица 1) В таблицу необходимо вписать символы кодируемого алфавита в порядке уменьшения вероятности их появления на выходе ИДИ. Старшим разрядам кодовых комбинаций символов из первой группы присваиваются значения 1, а символов из второй подгруппы - 0.:

- Среднее на символ число двоичных разрядов:

$$l_{sr}\{K(X)\} = \sum_{i=1}^4 p(x_i) l\{K(X)\} = 0.7 \times 1 + 0.25 \times 2 + \\ + 2 \times 0.0025 \times 3 = 1.35 \text{ (бит/символ)}$$

- Критерий корректного формирования эффективного кода:

$$\delta = \frac{l_{sr}\{K(X)\} - H(X)}{H(X)} = \frac{1.35 - 1.126}{1.126} = 0.198617 \approx 0.199$$

Полученная δ не входит в диапазон $0.05 \div 0.1$, значит необходимо модифицировать исходный ИДИ путем введения мультипликативно агрегированных символов (Таблица 2)

- Среднее на блок символов число двоичных разрядов полученного эффективного кода:

$$l_{sr}\{K(\tilde{X})\} = \sum_{i=1}^4 p(\tilde{x}_k) l\{K(\tilde{x}_k)\} = \dots = 2.34 \text{ (бит/блок)}$$

Таблица 2: Таблица эффективного кодирования модифицированного ИДИ

	$p(x_i)$	коды $K\{x_i\}$							длина $l\{K\{x_i\}\}$
$\tilde{x}_1 = x_1x_1$	0.49	1							1
$\tilde{x}_2 = x_1x_2$	0.175	0	1	1					3
$\tilde{x}_3 = x_2x_1$	0.175	0	1	0					3
$\tilde{x}_4 = x_2x_2$	0.065	0	0	1					3
$\tilde{x}_5 = x_1x_4$	0.0175	0	0	0	1	1	1		6
$\tilde{x}_6 = x_3x_1$	0.0175	0	0	0	1	1	0		6
$\tilde{x}_7 = x_4x_1$	0.0175	0	0	0	1	0	1		6
$\tilde{x}_8 = x_1x_3$	0.0175	0	0	0	1	0	0		6
$\tilde{x}_9 = x_2x_3$	0.0065	0	0	0	0	1	1	1	7
$\tilde{x}_{10} = x_2x_4$	0.0065	0	0	0	0	1	1	0	7
$\tilde{x}_{11} = x_3x_2$	0.0065	0	0	0	0	1	0	1	7
$\tilde{x}_{12} = x_4x_2$	0.0065	0	0	0	0	1	0	0	7
$\tilde{x}_{13} = x_3x_3$	0.00065	0	0	0	0	0	1	1	7
$\tilde{x}_{14} = x_3x_4$	0.00065	0	0	0	0	0	1	0	7
$\tilde{x}_{15} = x_4x_3$	0.00065	0	0	0	0	0	0	1	7
$\tilde{x}_{16} = x_4x_4$	0.00065	0	0	0	0	0	0	0	7

- Среднее на символ число двоичных разрядов полученного эффективного кода:

$$l_{sr}\{K(X)\} = \frac{l_{sr}\{K(\tilde{X})\}}{a} = \frac{2.34}{2} = 1.17 \text{ (бит/символ)}$$

Оценим степень близости средней длины эффективного кода на символ к энтропии исходного ИДИ:

$$\delta = \frac{l_{sr}\{K(X)\} - H(X)}{H(X)} = \frac{1.17 - 1.126}{1.126} = 0.0388017 \approx 0.039 \approx \delta_l = 0.05 \div 0.1$$

Неравенство выполнено даже в лучшую сторону, можно сделать вывод о том, что полученный эффективный код обеспечивает приемлимое решение поставленной задачи. Очевидно, что рассмотренный алгоритм является вариативным, так как деление символов на группы можно осуществлять различными способами, что повлияет на итоговое решение задачи эффективного кодирования.