Министерство образования и науки Российской Федерации
Нижнекамский химико-технологический институт (филиал)
Федерального государственного бюджетного образовательного учреждения
высшего профессионального образования
«Казанский национальный исследовательский технологический университет»

ТЕОРИЯ ИНФОРМАЦИИ

МЕТОДИЧЕСКИЕ УКАЗАНИЯ
ДЛЯ СТУДЕНТОВ ЗАОЧНОЙ ФОРМЫ ОБУЧЕНИЯ

Нижнекамск
2013

УДК 004
Л 40

Печатаются по решению редакционно-издательского совета Нижнекамского химико-технологического института (филиала) ФГБОУ ВПО «КНИТУ».

Рецензенты:
Галеев Э.Р., кандидат технических наук,
Саримов Н.Н., кандидат физико-математических наук.

Лежнева, Н.В.
Л 40 Теория информации : методические указания для студентов заочной формы обучения / Н.В. Лежнева, В.В. Гетман. – Нижнекамск : Нижнекамский химико-технологический институт (филиал) ФГБОУ ВПО «КНИТУ», 2013. - 32 с.

Приведены программа курса, контрольные задания и методические указания по следующим разделам курса «Теория информация»: оптимальное кодирование, помехоустойчивое кодирование. Изложены основные теоретические положения и методики построения эффективных кодов: Шеннона-Фано, Хаффмена, а также корректирующих кодов. Показано применение теории при решении практических задач, приведены практические задачи для самостоятельного решения.

Предназначены для студентов заочной формы обучения, по направлению «Информатика и вычислительная техника».

УДК 004

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Для повышения качества системы передачи информации необходимо согласование статистических свойств источников сообщений и каналов связи.

Из основной теоремы Шеннона о кодировании для дискретного канала связи без помех следует, что при преобразовании сообщений в статистически независимые и равновероятные символы можно повысить скорость передачи вплоть до пропускной способности этого канала. Технически это реализуется кодером источника, обеспечивающим кодирование, при котором за счет устранения избыточности снижается среднее число символов, требующихся для выражения одного знака сообщения, называемое *эффективным или оптимальным*. Оно позволяет при отсутствии помех снизить время передачи или объем запоминающего устройства, что повышает эффективность системы, а при наличии помех в канале связи – преобразовать входную информацию в такую последовательность символов, которая наилучше подготовлена для дальнейших преобразований (максимальна сжата).

При статистическом согласовании источников дискретных сообщений с каналами связи с помехами для повышения достоверности передачи сообщений при минимальном сокращении скорости передачи по каналу вводят дополнительную избыточность для того, чтобы она максимально способствовала устранению вредного действия помехи с определенными статистическими свойствами. Количество избыточной информации, необходимое для обеспечения достоверной передачи безызбыточных, равно потерям информации вследствие воздействия помехи. Техническая реализация возможности повышения достоверности передачи осуществляется кодером канала. Такое кодирование называется *помехоустойчивым*.

I. Программа курса

Тема 1. Система передачи информации

Структура системы передачи информации. Ее основные понятия и определения [1-3, 8, 13].

Тема 2. Математические модели сигналов

Понятие сигнала и его модели. Формы представления детерминированных сигналов. Спектры периодических и непериодических сигналов. Распределение энергии сигнала в спектре. Случайный процесс как модель сигнала. Частотное и спектральное представление случайных процессов [1-4, 6, 8-11, 13].

Тема 3. Преобразование непрерывных сигналов в дискретные

Преимущества цифровой формы представления сигналов. Дискретизация и восстановление аналогового сигнала. Критерии качества восстановления. Равномерная дискретизация. Теорема Котельникова. Адаптивная дискретизация. Квантование сигнала при отсутствии и наличии помех [1-3, 8, 10,12, 13].

Тема 4. Количественная оценка информации

Энтропия как мера неопределенности выбора. Свойства энтропии. Условная энтропия и ее свойства. Количество информации как мера снятой неопределенности [1-3, 6, 7-11, 13].

Тема 5. Информационные характеристики источников сообщений и каналов связи

Информационные характеристики дискретных источников сообщений. Информационные характеристики дискретных каналов связи. Информационные характеристики источников непрерывных сообщений. Информационные характеристики непрерывных каналов связи. Согласование статистических свойств источников сообщений и каналов связи [1-3, 7-11, 13].

Тема 6. Кодирование информации при передаче по дискретному каналу связи без помех

Кодирование как процесс выражения информации в цифровом виде. Технические средства представления информации в цифровой форме. Кодирование как средство криптографического закрытия информации. Эффективное кодирование [1, 4, 9, 10, 11, 13].

Тема 7. Кодирование информации при передаче по дискретному каналу связи с помехами

Основная теорема Шеннона о кодировании для дискретного канала связи с помехами. Разновидности помехоустойчивых кодов. Блоковые коды. Построение циклических кодов. Выбор образующего многочлена по заданному объему кода и заданной корректирующей способности. Коды Боуза- Чоудхури- Хоквингема. Итеративные коды [1, 2, 5, 9, 10, 11, 13].

II. Контрольные задания

2.1. Контрольные задания по теме «Оптимальное кодирование»

1. Определить избыточность сообщений, построенных из алфавита с распределением вероятностей появления символов в сообщениях, приведенным в таблице 1.

Таблица 1. Исходные данные для задания 1

Вариант	Вероятности
1	p_a=0.03, p_b=0.26, p_c=0.09, p_d=0.05, p_e=0.16, p_f=0.1, p_g=0.09, p_h=0.22
2	p_a=0.1, p_b=0.05, p_c=0.04, p_d=0.01, p_e=0.1, p_f=0.03, p_g=0.07, p_h=0.5
3	p_A=0.1, p_B=0.25, p_C=0.35, p_D=0.3

4	p_a=0.125, p_b=0.125, p_c=0.125, p_d=0.125, p_e=0.125, p_f=0.125, p_g=0.125, p_h=0.125
5	p_a=0.5, p_b=0.25, p_c=0.098, p_d=0.052, p_e=0.04, p_f=0.03, p_g=0.019, p_h=0.011
6	p_a=0.25, p_b=0.25, p_c=0.125, p_d=0.125, p_e=0.0625, p_f=0.0625, p_g=0.0625, p_h=0.0625
7	p_A=0.25, p_B=0.25, p_C=0.25, p_D=0.1, p_E=0.1, p_F=0.05
8	p_a=0.03, p_b=0.02, p_c=0.13, p_d=0.18, p_e=0.13, p_f=0.15, p_g=0.16, p_h=0.11, p_i=0.02, p_j=0.07
9	$p_А$=0.17, $p_Б$=0.17, $p_С$=0.11, $p_Г$=0.09, $p_Д$=0.07, $p_Е$=0.03, $p_Ж$=0.04, $p_И$=0.02, $p_К$=0.31, $p_Л$=0.03
10	p_a=0.18, p_b=0.18, p_c=0.18, p_d=0.18, p_e=0.1, p_f=0.09, p_g=0.09
11	$p_А$=0.5, $p_Б$=0.15, $p_С$=0.12, $p_Г$=0.1, $p_Д$=0.04, $p_Е$=0.04, $p_Ж$=0.03, $p_И$=0.02
12	p_A=0.5, p_B=0.15, p_C=0.12, p_D=0.1, p_E=0.04, p_F=0.04, p_G=0.03, p_H=0.02
13	p_a=0.19, p_b=0.19, p_c=0.19, p_d=0.19, p_e=0.08, p_f=0.08, p_g=0.08
14	p_a=0.3, p_b=0.18, p_c=0.15, p_d=0.15, p_e=0.07, p_f=0.06, p_g=0.05, p_i=0.04
15	$p_А$=0.38, $p_Б$=0.24, $p_С$=0.18, $p_Г$=0.1, $p_Д$=0.06, $p_Е$=0.02, $p_Ж$=0.02

2. Произвести кодирование по методу Шеннона-Фано для алфавита, приведенного в предыдущем задании. Вычислить энтропию и среднюю длину кодового слова.

3. Произвести кодирование двоичным кодом по методу Хаффмена для алфавита, приведенного в задании 1. Определить избыточность полученного кода.

4. Алфавит состоит из трех букв A, B, C с вероятностями, приведенными в таблице 2. Произвести кодирование отдельных букв и двухбуквенных сочетаний по методам Шеннона-Фано и Хаффмена. Сравнить избыточность и эффективность полученных кодов.

Таблица 2. Исходные данные для задания 4

Вариант	Вероятности
1	p_A=0.7, p_B=0.2, p_C=0.1
2	p_A=0.5, p_B=0.3, p_C=0.2
3	p_A=0.6, p_B=0.2, p_C=0.2
4	p_A=0.8, p_B=0.1, p_C=0.1
5	p_A=0.5, p_B=0.1, p_C=0.4
6	p_A=0.4, p_B=0.2, p_C=0.4
7	p_A=0.85, p_B=0.1, p_C=0.05
8	p_A=0.6, p_B=0.3, p_C=0.1
9	p_A=0.55, p_B=0.25, p_C=0.2
10	p_A=0.65, p_B=0.15, p_C=0.2
11	p_A=0.75, p_B=0.15, p_C=0.1
12	p_A=0.7, p_B=0.15, p_C=0.15
13	p_A=0.25, p_B=0.25, p_C=0.5
14	p_A=0.6, p_B=0.25, p_C=0.15
15	p_A=0.75, p_B=0.2, p_C=0.05

2.2. Контрольные задания по теме «Корректирующие коды»

1. Определить величину кодового расстояния между двумя двоичными кодовыми комбинациями, приведенными в таблице 3.

Таблица 3. Кодовые комбинации

Вариант	Кодовые комбинации
1	1101101, 1001011
2	1001010, 1010101

3	0110011, 1000010
4	1111000,0001000
5	1000000, 1100110
6	1100001, 0101110
7	1000111, 0001010
8	1100101, 1011001
9	1010110, 0111001
10	1101111, 1001100
11	1000110, 1011110
12	0010011, 0110011
13	1010110, 1000111
14	1011111, 0111100
15	1011100, 1010100

2. а) Определить величину кодового расстояния, обеспечивающего исправление s-кратных ошибок (варианты 1-7);
б) Определить наименьшее количество проверочных символов, а также количество информационных символов, необходимых для исправления s-кратных ошибок, если число символов в кодовой комбинации равно n (варианты 8-15).
Значения s и n приведены в таблице 4.

Таблица 4. Исходные данные для задания 2

Вариант	s	n
1	1	
2	2	
3	3	
4	4	
5	5	
6	6	
7	7	
8	2	10

9	3	10
10	3	14
11	2	11
12	2	14
13	3	11
14	3	12
15	2	12

3. Перемножить 3 многочлена, приведенных в таблице 5, в алгебре циклических кодов. Проделать аналогичную операцию и для двоичных эквивалентов.

Таблица 5. Многочлены

Вариант	Многочлены
1	$x+1$, x^2+x+1, x^3+x+1
2	x^2+1, x^3+1, x
3	x^3+x^2+x+1, $x+1$, x^2+x
4	x^4+1, x^3+x, x^2+1
5	x^2+x+1, x, x^4+1
6	x^3+x^2+x+1, x^2+1, $x+1$
7	x^4+x^3, x^4+1, x^2+x+1
8	x^3+x, x^2+x+1, x^4+x^2
9	x^4+x^3, x^3+x, $x+1$
10	x^2+x+1, x^4+1, x^3+x
11	x^4+x+1, x^3+1, x^2+x+1
12	x^4+1, x^3+x^2+1, x^2+x
13	x^5+1, x^2+x, $x+1$
14	x^4+x^3+1, x^3+x+1, x^2+1
15	x^4+x+1, x^3+x, x^3+x+1

4. Найти остаток от деления многочлена 1 на многочлен 2, приведенных в таблице 6. Проделать аналогичную операцию для двоичных эквивалентов.

Таблица 6. Многочлены

Вариант	Многочлен 1	Многочлен 2
1	$x^7+x^6+x^4+x+1$	x^4+x+1
2	x^5+x^4+x	x^4+x^2
3	$x^6+x^5+x^4+1$	x^3+x^2+x
4	$x^7+x^6+x^5$	x^3+x+1
5	$x^8+x^7+x^3+1$	x^4+x^2
6	$x^9+ +x+1$	x^5+x^4+x
7	$x^8+x^4+x^3+x$	$x^4+x^3+x^2$
8	$x^7+x^5+x^4$	x^3+x^2+x
9	$x^9+x^8+x^7$	x^5+x^4+1
10	x^8+x^6+1	x^4+x^3+1
11	x^7+x^4+x	x^4+1
12	$x^9+x^7+x^5$	x^5+x^2+x
13	$x^8+x^6+x^4+1$	x^6+x^2+x
14	x^7+x^6+x	x^3+x^2+x
15	$x^8+x^7+x^6+x^5$	x^4+x^2+x

5. Закодировать в циклическом коде заданные кодовые комбинации (таблица 7), если образующий многочлен g(x) имеет вид, приведенный в таблице 7.

Таблица 7. Исходные данные для задания 5

Вариант	Кодовые комбинации	g(x)
1	1001	x^3+x+1
2	1010	x^3+x^2+x
3	1100	x^3+x^2+x
4	1011	x^3+x^2+x
5	0111	x^3+x^2+x
6	1110	x^3+x+1
7	0110	x^3+x+1
8	1001	x^3+x^2+x

9	1010	x^3+x+1
10	1100	x^3+x^2+x
11	1011	x^3+x^2+x
12	1111	x^3+x+1
13	1110	x^3+x^2+x
14	0110	x^3+x^2+x
15	0111	x^3+x+1

6. Закодировать заданный многочлен (таблица 8) с проверкой на четность.

Таблица 8. Многочлен

Вариант	Многочлен
1	$x^7+ x^4+x^3+x+1$
2	$x^8+ x^5+x^2+x$
3	$x^6+x^5+x^4+1$
4	$x^7+ x^6+x^2+x$
5	$x^9+ x^6+x^5+x^3+1$
6	$x^8+x^7+ x^6+ 1$
7	$x^7+ x^5+x^3+x^2$
8	$x^6+x^5+x^4+x$
9	$x^7+ x^4+x^3+x+1$
10	$x^8+ x^7+x^5+x^3$
11	$x^6+x^3+x^2+x+1$
12	$x^7+ x^6+x^5+x^2+1$
13	$x^8+ x^4+x^3+x^2+x$
14	$x^9+x^8+ 1$
15	$x^8+ x^6+x^5+x^2$

7. Закодировать в циклическом коде заданную кодовую комбинацию h(x) (таблица 9).

Таблица 9. Кодовая комбинация

Вариант	h(x)

1	100011
2	110011
3	100101
4	100110
5	100111
6	110111
7	101001
8	101100
9	100010
10	101101
11	100001
12	110101
13	101110
14	101111
15	111100

8. Проверить принятую кодовую комбинацию h(x) (таблица 10) на наличие одиночной ошибки. При обнаружении ошибки исправить ее.

Таблица 10. Принятая кодовая комбинация

Вариант	h(x)
1	$x^{14}+x^{11}+x^{8}+x^{6}+x^{3}+x^{2}+x$
2	$x^{13}+x^{12}+x^{9}+x^{8}+x^{5}+1$
3	$x^{12}+x^{10}+x^{8}+x^{7}+x^{5}+x$
4	$x^{14}+x^{13}+x^{10}+x^{8}+x^{5}+x^{4}$
5	$x^{13}+x^{10}+x^{9}+x^{8}+x^{2}+x+1$
6	$x^{12}+x^{11}+x^{10}+x^{8}+x^{6}+x^{4}+x$
7	$x^{13}+x^{8}+x^{7}+x^{6}+x^{5}+x^{4}+1$
8	$x^{14}+x^{11}+x^{9}+x^{7}+x^{5}+x$
9	$x^{12}+x^{11}+x^{8}+x^{7}+x^{6}+x+1$
10	$x^{13}+x^{5}+x^{4}+x^{3}+x^{2}+x+1$
11	$x^{14}+x^{10}+x^{8}+x^{5}+x^{3}+x$

12	$x^{12}+ x^{9}+x^{7}+x^{6}+x^{4}+x^{2}+1$
13	$x^{13}+ x^{10}+x^{9}+x^{8}+x^{5}+x^{4}+x$
14	$x^{14}+ x^{13}+x^{12}+x^{11}+ x+1$
15	$x^{12}+ x^{11}+x^{9}+x^{8}+x^{6}+x^{5}$

9. Заданные кодовые комбинации (таблица 11) закодировать в циклическом коде. В правильно закодированные кодовые комбинации внести одиночные ошибки и по образующей матрице определить местоположение ошибки.

Таблица 11. Кодовые комбинации

Вариант	Кодовые комбинации		
	1	2	3
1	110100	100011	1001
2	1101010	110000	1010
3	11010000	1000	11010
4	110101	100100	1011
5	1101011	11000	1100
6	1101000	1101	10100
7	11010011	1001	11011
8	110110	100100	1101
9	1101100	110010	1110
10	1100111	1110	10101
11	11010010	1010	11100
12	110111	100101	1111
13	1101101	110011	1001
14	1100110	0111	10110
15	11101001	1100	11101

III. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ КОНТРОЛЬНОЙ РАБОТЫ

3.1. План выполнения контрольной работы

1. Изучение методики эффективного и помехоустойчивого кодирования
2. Приобретение практических навыков эффективного и помехоустойчивого кодирования путем решения практических заданий по вариантам.
3. Составление отчета по контрольной работе. Содержание отчета:
   - формулировка задачи по варианту;
   - описание методики решения задачи;
   - решение задачи;
   - выводы по проделанной работе;
   - список использованных источников
4. Подготовка ответов на контрольные вопросы.

3.2. ОПТИМАЛЬНОЕ КОДИРОВАНИЕ

Процесс составления кода в виде совокупности символов называется *кодированием*. Различают равномерные (например, код Бодо) и неравномерные (например, код Морзе) коды. Кодирование считается оптимальным, если на передачу сообщения затрачивается минимальное время. Если на передачу каждого элементарного символа (например, 0 или 1) тратится одно и то же время, то оптимальным будет такой код, при котором на передачу сообщения будет затрачено минимальное количество символов.

В общем случае необходимое число элементарных символов при кодировании составляет

$$k = \frac{\log_2 n}{\log_2 m},$$

где n - количество символов первичного алфавита, m - объем вторичного алфавита.

При двоичном кодировании m=2 и k = $\log_2 n$.

Учитывая статистические свойства источника сообщений, можно минимизировать среднее число двоичных символов, требуемых для выражения одного знака сообщения, что при отсутствии шума в канале связи позволяет сократить время передачи или объем запоминающего устройства.

Эффективное кодирование сообщений для передачи их по дискретному каналу связи без помех базируется на основной теореме Шеннона:

1) При любой производительности источника сообщений $\bar{I}(Z)$, меньшей пропускной способности канала $C_д$, т.е. при $\bar{I}(Z) = C_д - \varepsilon$, где ε- сколь угодно малая положительная величина, существует способ кодирования, позволяющий передавать по каналу связивсе сообщения, вырабатываемые источником.
2) Не существует способа кодирования, обеспечивающего передачу сообщений без их неограниченного накопления, если $\bar{I}(Z) > C_д$.

Шеннон доказал, что для дискретного канала связи без помех можно создать систему эффективного кодирования дискретных сообщений, для которой среднее число двоичных символов на букву алфавита будет сколь угодно близко к энтропии источника этих сообщений, но не менее ее.

Принципы построения оптимальных кодов:

1) Каждая кодовая комбинация должна содержать максимальное количество информации, что обеспечивает максимальную скорость передачи данных;

2) Символам первичного алфавита, имеющим наибольшую вероятность появления в сообщении, присваиваются более короткие кодовые комбинации, при этом средняя длина кодовых комбинаций имеет минимально возможную длину.

При таком кодировании избыточность кода, которая вызвана разными вероятностями знаков алфавита, сводится к минимуму (практически к нулю).

Экономичность кода характеризуется средней длиной кодовой комбинации:

$$k_c = \sum_{i=1}^{n} k_i p_i,$$

где k_i- количество двоичных разрядов у i-го кодируемого элемента, p_i- вероятность появления i-го элемента. Каждый элементарный символ кода при оптимальном кодировании должен нести максимальную информацию. При двоичном кодировании она равна 1 дв.ед. на один элементарный символ:

$$I = \frac{H}{k_c} = 1 \text{ дв.ед},$$

где H- энтропия источника сообщений.

В некоторых случаях для повышения экономичности кода выгоднее прибегать к увеличению числа кодируемых элементов, объединяя их в блоки и определяя для каждого блока свою кодовую комбинацию.

При учете вероятностей появления букв в текстах передаваемую информацию можно значительно сжать, сократить. Сжатие возможно при наличии избыточности. *Избыточностью сообщения* называется мера относительного удлинения сообщения; она определяется по формуле:

$$D = 1 - \frac{H}{H_{max}},$$

где H_{max}- максимально возможная энтропия. Наиболее эффективным способом снижения избыточности сообщения является построение эффективных или оптимальных кодов.

Оптимальные коды – это коды практически с нулевой избыточностью. Оптимальные коды имеют минимальную среднюю длину кодовых слов k_c. Верхняя и нижняя границы k_c для оптимального кодирования определяются из неравенства:

$$\frac{H}{\log m} \le k_c \le \frac{H}{\log m} + 1.$$

Пример 1: Сообщения составляются из буки алфавита a, b, c, d. Вероятности появления букв алфавита в текстах равны соответственно 0.2, 0.3, 0.4, 0.1. Найти избыточность сообщений, составленных из букв данного алфавита.

Решение:

Для алфавита из четырех букв максимальная энтропия $H_{max} = \log_2 m = \log_2 4 = 2$ дв. ед.. Средняя энтропия на символ сообщения

$$H = -\sum_{i=1}^{n} p_i \log p_i = -0.2\log 0.2 - 0.3\log 0.3 - 0.4\log 0.4 -$$

$$-0.1\log 0.1 = 1.84 \text{ бит.}$$

Следовательно, избыточность $D = 1 - \frac{H}{H_{max}} = 1 - \frac{1.84}{2} = 0.08$ дв.ед.

При построении оптимальных кодов наибольшее распространение нашли методики Шеннона-Фано и Хаффмена.

3.2.1. Методика Шеннона-Фано

Согласно методике Шеннона-Фано построение оптимального двоичного кода осуществляется следующим образом:

1) буквы алфавита сообщений упорядочиваются в порядке убывания их вероятностей;

2) последовательность букв разбивается на две по возможности равновероятные группы;

3) всем буквам первой группе присваивается символ "1", второй - "0",

4) каждая из полученных подгрупп, в свою очередь, разбивается на подгруппы с одинаковыми суммарными вероятностями с присвоением "1" и "0";

5) данный процесс продолжается до тех пор, пока в каждой из подгрупп не останется по одной букве.

Пример 2: Построить эффективный код по методу Шеннона-Фано для алфавита, состоящего из букв a1, a2, a3, a4, a5, вероятности появления которых равны 0.4, 0.3, 0.15, 0.1, 0.05 соответственно.

Решение: Построение кода Шеннона-Фано приведено в таблице 12.

Таблица. 12. Построение кода Шеннона-Фано

<table>
<tr><td rowspan="2">Буква</td><td rowspan="2">P</td><td colspan="4">Разряды</td><td rowspan="2">Код. комб-и</td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td>a1</td><td>0,4</td><td>1</td><td>-</td><td>-</td><td>-</td><td>1</td></tr>
<tr><td>a2</td><td>0,3</td><td rowspan="4">0</td><td>1</td><td>-</td><td>-</td><td>01</td></tr>
<tr><td>a3</td><td>0,15</td><td rowspan="3">0</td><td>1</td><td>-</td><td>001</td></tr>
<tr><td>a4</td><td>0,1</td><td rowspan="2">0</td><td>1</td><td>0001</td></tr>
<tr><td>a5</td><td>0,05</td><td>0</td><td>0000</td></tr>
</table>

Среднее число символов, приходящихся на букву:

$$k_c = \sum_{i=1}^{5} k_i p_i = 1 \cdot 0.4 + 2 \cdot 0.3 + 3 \cdot 0.15 + 4 \cdot 0.1 + 4 \cdot 0.05 =$$

=2.05 бит.

Энтропия $H = -\sum_{i=1}^{5} p_i \log p_i = 2.01$ бит.

Таким образом, средняя длина получилась достаточно близкой к предельному значению. Длина кода при равномерном кодировании k = 3 бита. Код Шеннона-Фано по сравнению с равномерным позволил сократить среднюю длину кодовых комбинаций для данного примера на 31,6%, т.к. избыточность $D = 1 - \frac{2.05}{3}$ = 0.316.

Методика Шеннона-Фано не всегда приводит к однозначному построению кода, т.к при разбиении на подгруппы можно сделать большей по вероятности как верхнюю, так и нижнюю подгруппу. От указанного недостатка свободна методика Хаффмена, позволяющая осуществлять однозначное построение кода с наименьшим для данного распределения вероятностей средним числом символов на букву.

3.2.2. Методика Хаффмена

Код Хаффмсна, в отличие от кода Шеннона-Фано, строится в виде кодового дерева следующим образом:

1) буквы алфавита сообщений располагаются в порядке убывания их вероятностей;

2) значения вероятностей двух последних букв суммируются, при этом верхней приписывается символ "1", а нижней - "0". Полученное суммарное значение вероятности располагается на такой воображаемой горизонтали, чтобы не нарушился общий порядок убывания вероятностей;

3) младшие значения вероятностей вновь суммируются с приписыванием "1" и "0". Этот процесс продолжают до тех пор, пока полученная суммарная вероятность не станет равной единице;

4) код соответствующей буквы составляется из символов, встречающихся при прохождении кодового дерева но линиям "направо-вверх-направо...". При этом кодовая комбинация записывается справа налево.

Пример 3: Построить эффективный код по методу Хаффмена для алфавита, приведенного в примере 1.

Решение: Построение кода Хаффмена приведено на рисунке 2.

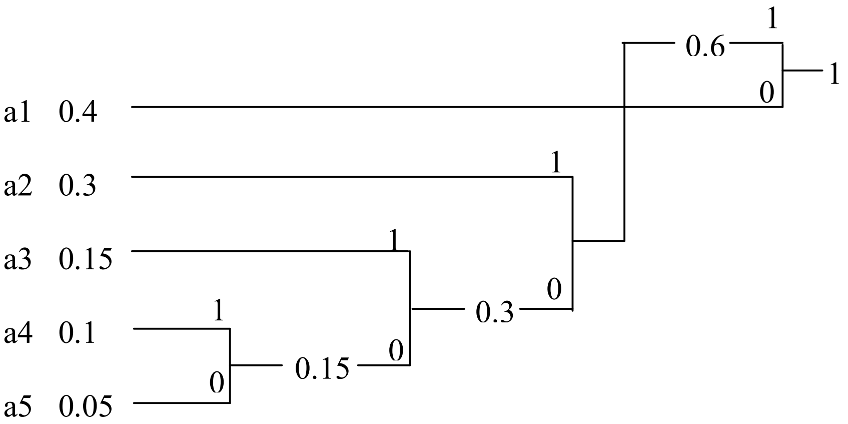

Рис.1. Построение кода Хаффмена.

В результате получили следующие кодовые комбинации:

a1 0
a2 11
a3 101
a4 1001
a5 1000

Средняя длина кода, построенного по методике Хаффмена, равна $k_c = 2.05$ бит и совпала с длиной, полученной по методике Шеннона-Фано.

3.2.3. Кодирование блоками

Повысить эффективность кодирования можно, строя код не для символов, а для блоков из n символов, при этом вероятность блока определяется произведением вероятностей символов, входящих в блок, т.к. знаки статистически не связаны.

Пример 4: Пусть сообщения образованы с помощью алфавита, состоящего из букв z_1 и z_2, с вероятностями появления 0.9 и 0.1 соответственно.

Поскольку вероятности букв существенно различны, то последовательность из таких букв будет обладать избыточностью. При побуквенном кодировании никакого эффекта не получиться. На передачу каждой буквы требуется один символ: 1 или 0, в то время как энтропия равна 0.47 дв.ед.

При кодировании блоков, содержащих по две буквы, получим коды, приведенные в таблице 13.

Таблица 13. Кодирование двухбуквенных блоков

Блоки	Вероятности	Кодовые комбинации
z_1 z_1	0.81	1
z_1 z_2	0.09	01
z_2 z_1	0.09	001
z_2 z_2	0.01	000

Среднее число символов на блок равно 1.29, а среднее число символов на букву 0.645.

Кодирование блоков, содержащих по три буквы, дает еще больший эффект (таблица 14).

Таблица 14. Кодирование трехбуквенных блоков

Блоки	Вероятности	Кодовые комбинации
z_1 z_1 z_1	0.729	1
z_2z_1 z_1	0.081	011
z_1z_2 z_1	0.081	010
z_1z_1 z_2	0.081	001
z_2 z_2 z_1	0.009	00011
z_2 z_1 z_2	0.009	00010
z_1z_2 z_2	0.009	00001
z_2 z_2 z_2	0.001	00000

Среднее число символов на блок равно 1.59, а среднее число символов на букву 0.53, что всего на 12% больше энтропии. Теоретический минимум H(Z)=0.47 дв.ед. может быть достигнут при кодировании блоков, содержащих бесконечное число букв:

$$\lim_{n\to\infty} k_c = H(Z).$$

3.3. КОРРЕКТИРУЮЩИЕ КОДЫ

Помехоустойчивыми, или *корректирующими*, называются коды, позволяющие обнаружить или обнаружить и исправить ошибки в кодовых комбинациях.

Корректирующая способность кода обусловлена наличием избыточных символов r = n - k, где k - число информационных символов, n - число символов в закодированной последовательности. Корректирующая способность кода также тесно связана с кодовым (хэмминговым) расстоянием d, представляющим собой число символов, которыми отличаются комбинации данного кода. Для нахождения кодового расстояния между двумя комбинациями двоичного кода нужно определить число единиц в сумме этих комбинаций по модулю два. Так, складывая две комбинации по модулю два

$$\begin{array}{r} 11010111 \\ \oplus\underline{10001101} \\ 01011010 \end{array},$$

определяем, что кодовое расстояние d = 4.

Минимальное расстояние, взятое по всем парам разрешенных комбинаций кода, называется минимальным кодовым расстоянием d_{min}.

Для обнаружения всех ошибок кратности до r включительно и исправления ошибок кратности до s включительно, причем r ≥ s, минимальное кодовое расстояние выбирается из соотношения d_{min} = r +s + 1. Если код предназначен для исправления всех обнаруженных ошибок кратности до s включительно, то $d_{min} = 2s + 1$. Каждый конкретный корректирующий код не гарантирует исправления любой комбинации ошибок. Коды предназначены для исправления комбинаций ошибок, наиболее вероятных для заданного канала связи и наиболее опасные по последствиям. Если характер и уровень помех отличается от предполагаемых, эффективность применения кода резко снижается. Применение корректирующего кода не может гарантировать безошибочного приема, но позволяет повысить вероятность получения на выходе правильного результата.

В настоящее время существует множество кодов, позволяющих обнаруживать и исправлять ошибки различной кратности. Наибольшее применение нашли циклические коды.

Циклическими кодами называют специальную группу кодов, для построения которых могут быть использованы циклические свойства квадратных матриц, а также коды, которые описываются неприводимыми, образующими (порождающими) многочленами (полиномами). Например, для кодовой комбинации 101101 полиномиальное представление следующее:

$h(x) = 1{\cdot}x^5 + 0{\cdot}x^4 + 1{\cdot}x^3 + 1{\cdot}x^2 + 0{\cdot}x^1 + 1 = x^5 + x^3 + x^2 + 1.$

Циклические коды относятся к систематическим (n, k) - кодам, в которых контрольные r и информационные k (, где $n = k + r$) разряды расположены на строго определенных местах.

3.3.1. Построение циклических кодов

Циклический код получают следующим образом: заданный многочлен h(x) сначала умножается на одночлен $x^{n-k} = x^r$, затем делится на образующий многочлен g(x). В результате получим:

$$\frac{h(x)x^{n-k}}{g(x)} = Q(x) + \frac{R(x)}{g(x)}$$

или $F(x) = Q(x) \bullet g(x) = x^{n-k}h(x) + R(x)$.

Таким образом, циклический код можно построить умножением кодовой комбинации h(x), являющейся заданной, на одночлен x^{n-k} и добавлением к этому произведению остатка R(x).

Образующий полином g(x) является сомножителем при разложении двучлена x^n+1.

Сомножителями разложения двучлена являются неприводимые полиномы (приложение).

Образующий полином выбирают следующим образом. По заданной кодовой комбинации (известному количеству информационных разрядов k) определяют число контрольных символов из соотношения:

$$r = [\log (n + 1)], \qquad (1)$$

где [ ] – округление до ближайшего большего целого

или по эмпирической формуле:

$$r = [\log\{(k + 1) + [\log(k + 1)]\} \qquad (2)$$

Соотношения между n, k и r приведены в таблице 15.

Таблица 15. Зависимости между n, k и r

n	3	5	6	7	9..15	17..31	33..63	65..127
k	1	2	3	4	5..11	12..26	27..57	28..120
r	2	3	3	3	4	5	6	7

Из таблицы неприводимых полиномов (приложение) выбирают самый короткий многочлен g(x) со степенью, равной числу контрольных символов r, его и принимают за образующий.

Пример: Пусть требуется закодировать кодовую комбинацию вида 1101, что соответствует $h(x) = x^3 + x^2 + 1$.

Решение: По формуле (2) определяем число контрольных символов при количестве информационных $k=4$: $r = [\log\{(4 + 1) + [\log(4 + 1)]\}] = 3$. Из таблицы неприводимых полиномов (приложение) выбираем многочлен третьей степени: $g(x) = x^3 + x + 1$, т.е. 1011.

Умножим $h(x)$ на x^r:

$h(x)x^r = (x^3 + x^2 + 1)x^3 = x^6 + x^5 + x^3 \to 1101000$

Разделим полученное произведение на образующий полином $g(x)$:

$$\frac{h(x)x^r}{g(x)} = \frac{x^6 + x^5 + x^3}{x^3 + x + 1} = x^3 + x^2 + x + 1 + \frac{1}{x^3 + x + 1} \to$$

$$\to 1111 + \frac{001}{1011}.$$

При делении необходимо учитывать, что вычитание производится по модулю 2.

Полученный остаток суммируем с $h(x)x^r$. В результате получим закодированное сообщение:

$F(x) = (x^3 + x^2 + x + 1)(x^3 + x + 1) = (x^3 + x^2 + 1)x^3 + 1 \to$ 1101001.

В полученной кодовой комбинации циклического кода информационные символы $h(x) = 1101$, а контрольные $R(x) = 001$. Закодированное сообщение делится на образующий полином без остатка.

Сообщение, которое закодировано, является одной из комбинаций 4-разрядного кода, так как весь ансамбль сообщений (вся группа) содержит $N=2^4 =16$ сообщений. Это значит, что если все сообщения передаются в закодированном виде, то каждое из них необходимо кодировать так же, как и комбинацию $h(x) = 1101$. Однако выполнять дополнительные пятьнадцать расчетов (в общем случае 2^n-1) нет необходимости.

Это можно сделать проще, путем составления образующей (порождающей) матрицы.

Образующая матрица составляется на основе единичной транспонированной $I_{k,k}$, к которой справа дописывается матрица дополнений $C_{n,r}$:

$$H_{n,k} = \| I_{k,k}\ C_{n,r} \|.$$

Матрица дополнений получается из остатков от деления единицы с нулями на образующий многочлен g(x). Комбинации единиц с нулями представляют собой векторы ошибок: 00...01, 00... 10, 00... 100, 10…00. Каждому вектору ошибок будет соответствовать свой остаток (опознаватель):

$$\begin{array}{l}
\;1100000\ldots \;\big|\,\underline{1011} \\
\oplus\underline{1011} \\
\quad 011 \quad 1\text{-й остаток} \\
\quad 110 \quad 2\text{-й остаток} \\
\quad 1100 \\
\;\oplus\,\underline{1011} \\
\qquad 111 \quad 3\text{-й остаток} \\
\qquad 1110 \\
\quad\;\oplus\underline{1011} \\
\qquad\; 101 \quad 4\text{-й остаток}
\end{array}
\qquad
H_{7,4} = \begin{pmatrix} 0001 & 011 \\ 0010 & 110 \\ 0100 & 111 \\ 1000 & 101 \end{pmatrix}.$$

Получены четыре комбинации циклического кода, что равно количеству информационных разрядов, а так как в четырехразрядном двоичном коде всего $N = 2^4 = 16$ комбинаций, то остальные 11 ненулевых комбинаций находятся суммированием по модулю 2 всевозможных сочетаний строк образующей матрицы. Например, необходимо из исходных кодов 1101 и 1010 получить циклические помехозащищенные коды. Они получаются суммированием соответствующих строк образующей матрицы:

1. 1+3+4 = 1101001;
2. 2+4 = 1010011.

3.3.2. Декодирование циклических кодов

Для обнаружения и исправления ошибок принятая комбинация делится на образующий многочлен g(x). Если остаток R(x) = 0, значит комбинация принята без ошибок. Наличие ненулевого остатка свидетельствует о том, что комбинация принята искаженной. Значение остатка совпадет с одним из опознавателей матрицы $H_{n,k}$ который и укажет на местоположение ошибки по вектору ошибок.

Пример: Принятая кодовая комбинация имеет вид: $F(x)=x^6+x^5+x^2+x+1 \rightarrow 1100111$. Определить имеется ли в ней ошибка. При наличии ошибки найти ее местоположение.

Решение: По формуле (1) определяем количество контрольных символов при n=7: r = [log (7 + 1)]= 3, следовательно, количество информационных символов k= 7- 3 = 4. Из таблицы неприводимых полиномов (приложение) выбираем самый короткий многочлен третьей степени: $g(x)=x^3+x+1$, т.е. 1011. Делим принятую кодовую комбинацию на образующий многочлен:

$$
\begin{array}{l}
1100111 \,\big|\, \underline{1011} \\
\oplus\underline{1011} \\
\quad 1111 \\
\oplus\ \underline{1011} \\
\qquad 1001 \\
\quad \oplus\underline{1011} \\
R(x) = 101
\end{array}
$$

Ненулевой остаток R(x)=101 свидетельствует о наличии ошибки в принятой комбинации кода. Местоположение ошибки определяем по образующей матрице $H_{7,4}$. Остаток совпадает с опознавателем матрицы $H_{7,4}$, расположенным в четвертой строке. По соответствующему вектору ошибок определяем, что ошибка в шестом разряде, т.е. переданная кодовая комбинация имела следующий вид: $x^5+x^2+x+1 \rightarrow 0100111$.

IV. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Оптимальные коды.
2. Коды Шеннона-Фано.
3. Коды Хаффмена.
4. Избыточность сообщения.
5. Экономичность кода.
6. Помехоустойчивые коды.
7. Чем обусловлена корректирующая способность кода?
8. Разновидности помехоустойчивых кодов.
9. Что подразумевают под кратностью ошибки?
10. Как определяется минимальное кодовое расстояние?
11. Соотношение между минимальным кодовым расстоянием и числом обнаруживаемых и исправляемых ошибок.
12. Циклические коды.
13. Построение циклического кода.
14. Декодирование циклического кода.
15. Образующий полином.
16. Образующая матрица.

Литература

1. Пании В.В. Основы теории информации/В.В. Панин. –М.: Бином. Лаборатория знания, 2009.
2. Духин А.А. Теория информации/ А.А. Духин. –М.: Гелиос, 2007.
3. Литвинская О.С. Основы теории передачи и информации/ О.С.Литвинская, Н.И.Чернышев. –М.: КноРус, 2010.
4. Хэмминг Р.В. Теория кодирования и теория информации/ Хэмминг Р.В.. –М.:Радио и связь, 1983.
5. Шеннон К. Работы по теории информации и кибернетике/ К. Шеннон. –М.: Радио и связь, 1963.
6. Вентцель, Л.А.Овчаров. Теория случайных процессов и ее инженерные приложения/ Е.С. Вентцель, –М.: Высш. школа, 2010. – 448 с.
7. Вентцель Е.С. Теория вероятностей и ее инженерные приложения/ Е.С. Вентцель, Л.А.Овчаров. –М.: КноРус, 2010. – 480 с.
8. Игнатьв В.А. Теория информации и передачи сигналов/ В.А.Игнатьев. –М.: Радио и связь, 1991.
9. Березкин Е.Ф. Основы теории информации и кодирования : лабораторный практикум / Е.Ф. Березкин, Ю.Н.Федосеев. –М.: Изд. МИФИ, 1997.
10. Коган Н.М. Прикладная теория информации/ Н.М. Коган. –М.: Радио и связь. 1981.
11. Цымбал В.П. Задачи по теории информации и кодированию: учебное пособие / В.П.Цымбал. – Киев.: Вища школа, 1976. –276 с.
12. Шувалов В.П. Передача дискретных сообщений/ В.П.Шувалов. –М.: Радио и связь, 1990.
13. Дмитриев В.И. Прикладная теория информации/ В.И.Дмитриев. –М.: Высш. школа, 1989. – 320 с.

Приложение

Таблица неприводимых полиномов

g(x)	**полином**
$g(x)$	$x+1$
$g\left(x^2\right)$	x^2+x+1
$g\left(x^3\right)$	x^3+x+1
$g\left(x^3\right)$	x^3+x^2+1
$g_1\left(x^4\right)$	x^4+x+1
$g_2\left(x^4\right)$	x^4+x^3+1
$g_3\left(x^4\right)$	$x^4+x^3+x^2+x+1$
$g_1\left(x^5\right)$	x^5+x^2+1
$g_2\left(x^5\right)$	x^5+x^3+1
$g_3\left(x^5\right)$	$x^5+x^4+x^3+x^2+1$
$g_4\left(x^5\right)$	$x^5+x^4+x^2+x+1$
$g_5\left(x^5\right)$	$x^5+x^3+x^2+x+1$
$g_6\left(x^5\right)$	$x^5+x^4+x^3+x+1$
$g_1\left(x^6\right)$	x^6+x+1

$g_2\left(x^6\right)$	x^6+x^3+1
$g_3\left(x^6\right)$	x^6+x^5+1
…	…
$g_1\left(x^7\right)$	x^7+x+1
$g_2\left(x^7\right)$	x^7+x^3+1
$g_3\left(x^7\right)$	$x^7+x^3+x^2+x+1$
…	…
$g_1\left(x^8\right)$	$x^8+x^4+x^3+x+1$
$g_2\left(x^8\right)$	$x^8+x^4+x^3+x^2+1$
…	…
$g_1\left(x^9\right)$	x^9+x+1
$g_2\left(x^9\right)$	x^9+x^4+1

Учебное издание

Лежнева Наталья Викторовна
кандидат технических наук, доцент

Гетман Валерия Владимировна
кандидат технических наук, доцент

ТЕОРИЯ ИНФОРМАЦИИ

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ СТУДЕНТОВ ЗАОЧНОЙ ФОРМЫ ОБУЧЕНИЯ

Корректор Габдурахимова Т.М.
Худ. редактор Федорова Л.Г.

Сдано в набор 24.06.2013
Подписано в печать 19.09.2013.
Бумага писчая. Гарнитура Таймс.
Усл. печ. л. 2. Тираж 100.
Заказ №38.

НХТИ (филиал) ФГОУ ВПО «КНИТУ»,
г. Нижнекамск, 423570, ул.30 лет Победы, д.5а.