

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Оренбургский государственный университет»

Кафедра промышленной электроники
и информационно-измерительной техники

М.Г. Петрушанский

МЕТОДЫ И СРЕДСТВА ОБРАБОТКИ ИНФОРМАЦИИ

Методические указания

Рекомендовано к изданию редакционно-издательским советом федерального государственного бюджетного образовательного учреждения высшего образования «Оренбургский государственный университет» для обучающихся по образовательным программам высшего образования по направлениям подготовки 11.03.02 Инфокоммуникационные технологии и системы связи и 11.03.04 Электроника и наноэлектроника

Оренбург
2018

УДК 681.3
ББК 32.97
П 31

Рецензент – доцент, кандидат технических наук С.А. Сильвашко

Петрушанский, М.Г.
П 31 Методы и средства обработки информации: методические указания / М.Г. Петрушанский; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2018.

В методических указаниях изложены требования к содержанию и оформлению курсовой работы, предусмотренной рабочими программами дисциплин «Теория информации и теория кодирования» в пятом семестре обучения и «Основы информационной техники» в четвертом семестре обучения, приведены варианты индивидуальных заданий.

Методические указания предназначены для самостоятельной (внеаудиторной) работы обучающихся по образовательным программам высшего образования по направлениям подготовки 11.03.02 Инфокоммуникационные технологии и системы связи и 11.03.04 Электроника и наноэлектроника.

УДК 681.3
ББК 32.97

© Петрушанский М.Г., 2018
© ОГУ, 2018

Содержание

1 Задание на курсовую работу. Требования к оформлению пояснительной записки. Методика оценки работы.....	5
1.1 Задание на курсовую работу.....	5
1.2 Требования к содержанию пояснительной записки и изложению материала.....	8
1.3 Методика оценки работы обучающегося над курсовой работой	9
2 Краткие теоретические сведения и рекомендации по выполнению курсовой работы.....	11
2.1 Системы обработки и передачи информации.....	11
2.2 Источник сообщений	11
2.3 Характеристики источника сообщений	13
2.4 Модель кодера источника сообщений	14
2.5 Обобщенный источник сообщений.....	16
2.6 Дискретные каналы передачи информации.....	17
2.7 Оптимальное правило восстановления принятого сообщения.....	20
Список использованных источников	24

Введение

Курсовая работа на тему «Методы и средства обработки информации» по дисциплине «Теория информации и теория кодирования» направления подготовки 11.03.02 Инфокоммуникационные технологии и системы связи выполняется в пятом семестре обучения, а по дисциплине «Основы информационной техники» направления подготовки 11.03.04 Электроника и нанoeлектроника – в четвертом семестре обучения. В предлагаемой курсовой работе рассматривается математическая модель канала передачи и обработки информации, ряд характеристик которой следует рассчитать. Целью выполнения курсовой работы является закрепление теоретических знаний, полученных при изучении соответствующей дисциплины. В результате выполнения курсовой работы обучающийся должен показать способность применения методов практического расчёта статистических параметров и характеристик средств обработки информации, владение методами кодировки передаваемых сообщений кодом Хаффмана и исследования свойств зашумленного канала обработки и передачи информации.

Методические указания предназначены для организации самостоятельной (внеаудиторной) работы обучающихся по образовательным программам высшего образования по направлениям подготовки 11.03.02 Инфокоммуникационные технологии и системы связи и 11.03.04 Электроника и нанoeлектроника при выполнении курсовой работы и содержат варианты заданий, необходимые для их выполнения теоретические сведения и требования к содержанию пояснительной записки. Перед началом выполнения курсовой работы целесообразно просмотреть включенные в список использованных источников работы теоретического [1-4] и практического [5-6] характера.

1 Задание на курсовую работу. Требования к оформлению пояснительной записки. Методика оценки работы

1.1 Задание на курсовую работу

Имеется некоторая система обработки и передачи дискретных сообщений. Соответствующие разным вариантам характеристики источника сообщений приведены в таблице 1. Канал передачи сообщений, для кодирования которых используется метод Хаффмана, зашумлен, то есть принятый символ не обязательно совпадает с переданным. Предполагается, что канал передачи является стационарным, у него отсутствует память и он характеризуется приведенными в таблице 2 значениями переходных вероятностей в соответствии со степенью зашумленности канала.

При выполнении курсовой работы необходимо решить следующие задачи:

1) разработать общую структурную схему системы обработки и передачи дискретных сообщений, включающую источник сообщений, кодер, модулятор, канал передачи, демодулятор, декодер и получатель сообщений. Показать характерные для разных промежуточных точек разработанной схемы качественные временные диаграммы сигналов, сопровождающиеся текстовыми описаниями. Вид модуляции выбрать самостоятельно;

2) рассчитать количество переданной информации, содержащееся в каждом случайном для приемника элементарном сообщении источника, среднюю энтропию, максимальную среднюю энтропию и избыточность источника информации;

3) построить код Хаффмана для сообщений источника информации;

4) определить вероятности появления символов «0» и «1», передаваемых обобщенным источником по каналу. Найти скорость передачи сообщений по каналу без помех. Рассчитать пропускную способность канала и сравнить ее значение со скоростью передачи сообщений;

5) найти оптимальное по критерию минимума вероятности средней ошибки правило выбора переданного символа по принятому при выполнении условия сильно зашумленного канала;

б) рассчитать среднюю вероятность ошибки при передаче сообщения в случае слабо зашумленного канала;

7) определить вероятность верного приема в условиях слабо зашумленного канала последовательности сообщений, приведенной в таблице 2.

Таблица 1 – Номер варианта, алфавит источника и вероятности сообщений

№	а	б	в	г	д	ж	з	и	к	л	м	н	о	п
1	0,039	0,01	0,052	0,007	0,09	0,06	0,132	0,119	0,042	0,073	0,1	0,09	0,115	0,071
2	0,035	0,11	0,049	0,089	0,001	0,036	0,077	0,11	0,064	0,097	0,06	0,098	0,078	0,096
3	0,091	0,024	0,067	0,082	0,119	0,027	0,078	0,111	0,023	0,022	0,11	0,068	0,153	0,025
4	0,105	0,025	0,105	0,02	0,094	0,036	0,086	0,093	0,016	0,107	0,066	0,055	0,122	0,07
5	0,055	0,1	0,054	0,116	0,087	0,023	0,099	0,059	0,003	0,067	0,062	0,099	0,077	0,099
6	0,02	0,057	0,052	0,025	0,151	0,109	0,047	0,05	0,146	0,038	0,05	0,021	0,136	0,098
7	0,064	0,102	0,034	0,085	0,092	0,074	0,102	0,028	0,026	0,067	0,014	0,014	0,167	0,131
8	0,032	0,058	0,089	0,064	0,086	0,122	0,089	0,081	0,026	0,079	0,035	0,086	0,083	0,07
9	0,099	0,083	0,107	0,077	0,121	0,097	0,089	0,042	0,041	0,014	0,113	0,021	0,011	0,085
10	0,082	0,061	0,07	0,023	0,111	0,124	0,131	0,045	0,019	0,118	0,091	0,011	0,098	0,016
11	0,074	0,102	0,034	0,095	0,092	0,084	0,102	0,018	0,026	0,057	0,04	0,014	0,167	0,131
12	0,065	0,1	0,064	0,116	0,097	0,023	0,089	0,059	0,003	0,057	0,052	0,099	0,077	0,089
13	0,097	0,008	0,007	0,111	0,064	0,036	0,116	0,018	0,01	0,016	0,1	0,16	0,128	0,039
14	0,14	0,083	0,021	0,051	0,014	0,092	0,137	0,084	0,056	0,101	0,038	0,047	0,102	0,034
15	0,107	0,084	0,094	0,104	0,063	0,087	0,009	0,081	0,022	0,071	0,053	0,106	0,035	0,084
16	0,02	0,067	0,052	0,035	0,141	0,109	0,037	0,05	0,136	0,038	0,05	0,021	0,126	0,098
17	0,059	0,019	0,062	0,017	0,08	0,06	0,104	0,11	0,062	0,05	0,11	0,08	0,115	0,072
18	0,099	0,093	0,107	0,087	0,131	0,097	0,079	0,042	0,031	0,014	0,113	0,021	0,011	0,075
19	0,105	0,035	0,105	0,02	0,094	0,046	0,096	0,083	0,016	0,107	0,056	0,045	0,122	0,07
20	0,101	0,011	0,074	0,059	0,079	0,088	0,035	0,099	0,08	0,073	0,09	0,096	0,106	0,009
21	0,045	0,11	0,059	0,099	0,001	0,046	0,067	0,11	0,054	0,097	0,06	0,088	0,078	0,086
22	0,082	0,071	0,07	0,033	0,111	0,134	0,131	0,035	0,019	0,118	0,081	0,011	0,088	0,016
23	0,042	0,058	0,099	0,064	0,096	0,122	0,079	0,081	0,016	0,079	0,025	0,086	0,073	0,07
24	0,091	0,034	0,067	0,092	0,119	0,037	0,068	0,111	0,013	0,022	0,11	0,058	0,153	0,025
25	0,069	0,019	0,052	0,007	0,09	0,06	0,101	0,11	0,062	0,053	0,1	0,09	0,115	0,072

Таблица 2 – Номер варианта, вероятности переходов в зависимости от степени зашумленности канала передачи и вид последовательности сообщений

№	Сильно зашумленный канал		Слабо зашумленный канал		Последовательность сообщений
	$P(x_2 = 1 x_1 = 0)$	$P(x_2 = 1 x_2 = 1)$	$P(x_1 = 0 x_1 = 0)$	$P(x_1 = 0 x_2 = 1)$	
1	0,38	0,62	0,957	0,039	кино
2	0,29	0,71	0,921	0,065	пила
3	0,36	0,59	0,965	0,047	нога
4	0,28	0,69	0,938	0,056	залп
5	0,31	0,59	0,976	0,045	воин
6	0,36	0,61	0,959	0,076	гонг
7	0,32	0,67	0,969	0,062	блок
8	0,24	0,70	0,981	0,075	волк
9	0,27	0,69	0,957	0,049	гном
10	0,32	0,77	0,968	0,078	липа
11	0,21	0,70	0,976	0,081	вода
12	0,35	0,68	0,965	0,054	жало
13	0,27	0,79	0,977	0,035	игла
14	0,32	0,75	0,943	0,021	бомж
15	0,31	0,69	0,978	0,027	дома
16	0,28	0,75	0,925	0,065	зима
17	0,35	0,71	0,973	0,034	лика
18	0,41	0,58	0,945	0,051	плод
19	0,33	0,69	0,955	0,062	зонд
20	0,35	0,61	0,928	0,041	коза
21	0,39	0,59	0,947	0,039	плов
22	0,4	0,67	0,967	0,057	рама
23	0,31	0,71	0,954	0,048	клон
24	0,28	0,62	0,959	0,032	глаз
25	0,34	0,61	0,969	0,078	банк

1.2 Требования к содержанию пояснительной записки и изложению материала

Пояснительная записка к курсовой работе, в соответствии с требованиями СТО 02069024.101-2015 [7], должна содержать следующие структурные элементы:

- титульный лист;
- задание на выполнение курсовой работы с исходными данными в соответствии с номером варианта;
- аннотацию;
- содержание;
- введение;
- основную часть;
- заключение;
- список использованных источников.

Во введении следует определить цель курсовой работы и задачи, подлежащие решению для ее достижения, привести краткую характеристику структуры пояснительной записки по разделам.

Основная часть должна содержать материал, отражающий пути достижения цели курсовой работы. Она должна включать:

- обобщенную структурную схему системы обработки и передачи дискретных сообщений;
- расчет информационных характеристик источника;
- построение кода для сообщений источника;
- расчет статистических характеристик закодированных сообщений;
- определение оптимального по минимуму средней ошибки правило восстановления символа при приеме в условиях сильно зашумленного канала;
- определение ошибок в передаче сообщений по слабо зашумленному каналу.

В заключении необходимо кратко изложить результаты, полученные при выполнении курсовой работы.

Для проведения расчетов рекомендуется использовать математическую систему MathCAD.

1.3 Методика оценки работы обучающегося над курсовой работой

Защита курсовой работы проводится в последнюю неделю четвертого семестра обучения. Дату и время проведения защиты назначает преподаватель (руководитель курсовой работы). Защита, как правило, проводится публично, в присутствии всех обучающихся группы.

На защиту обучающийся должен представить сброшюрованную пояснительную записку и подготовленный короткий доклад на 3 – 5 минут. Доклад должен сопровождаться демонстрацией презентации, выполненной в среде MS PowerPoint.

При выставлении оценки за курсовую работу учитывается:

- качество оформления пояснительной записки (соответствие требованиям СТО 02069024.101-2015 [7]);
- полнота и правильность выполнения задания;
- качество доклада и демонстрационного материала (презентации).

Курсовая работа оценивается по четырехбальной системе: «отлично», «хорошо», «удовлетворительно» и «неудовлетворительно».

Оценка **«отлично»** выставляется за курсовую работу, если пояснительная записка оформлена в полном соответствии с требованиями стандарта СТО 02069024.101-2015 [7], задание выполнено в полном объеме и результаты его выполнения грамотно (с инженерных позиций) представлены в пояснительной записке, во время защиты обучающийся уверенно отстаивал решения, принятые в работе.

Оценка **«хорошо»** выставляется за курсовую работу, если пояснительная записка в основном удовлетворяет требованиям СТО 02069024.101-2015 [7], но при этом имеют место незначительные отклонения от требований этого стандарта, задание выполнено в полном объеме, но его результаты в пояснительной записке

изложены некачественно (для понимания сути представленного материала требуются дополнительные пояснения), во время защиты обучающийся уверенно отстаивал решения, принятые в работе.

Оценка **«удовлетворительно»** выставляется за курсовую работу, если пояснительная записка в основном удовлетворяет требованиям СТО 02069024.101-2015 [7], но не выполнены одновременно несколько требований стандарта (например, форматирование абзаца и размера символов в формулах, форматирование размера шрифта в тексте и оформление библиографической записи и т. п.), задание выполнено не в полном объеме.

Оценка **«неудовлетворительно»** выставляется за курсовую работу, если обучающийся не представил пояснительную записку на дату проведения защиты.

2 Краткие теоретические сведения и рекомендации по выполнению курсовой работы

2.1 Системы обработки и передачи информации

Системой передачи называется совокупность средств, необходимых для передачи сообщений от источника к получателю [3]. Как правило, общая функциональная схема системы передачи включает источник и получатель сообщений. Каналом передачи называют совокупность устройств и линий передачи, которые сигнал проходит последовательно между любыми двумя точками системы передачи.

Курсовая работа посвящена рассмотрению простейшей модели системы обработки и передачи информации, включающей модулятор, передатчик, линию передачи, приемник и демодулятор. Для моделирования работы кодера источника сообщений, включенного в обобщенную модель источника, в работе рассматривается метод кодирования Хаффмана. Декодер должен быть внесен в обобщенную модель получателя сообщений.

Таким образом, рассматриваемая в работе модель системы обработки и передачи информации включает обобщенные модели источника и получателя сообщений и модель канала передачи.

2.2 Источник сообщений

Источник сообщений для получателя можно рассматривать как генератор случайным образом выбранного сообщения, входящего в известный набор возможных для этого генератора сообщений. Для заданного источника сообщений должен быть известен набор возможных сообщений $X = \{x_i\}$ и определены вероятности этих сообщений $\{p(x_i)\}$, то есть должно быть задано множество упорядоченных пар $\{x_i, p(x_i)\}$. Если число возможных сообщений конечно, то источник и включающий его канал передачи информации называются дискретными.

Предполагается, что в единицу времени τ источник выдает только одно сообщение. Источник сообщений называется источником без памяти, если каждое очередное сообщение выдается без учета воспроизведенного на предыдущем интервале времени τ сообщения. Такое допущение существенно упрощает решение задач, связанных с обработкой и передачей информации.

Последовательность сообщений, «генерируемую» источником, можно представить в виде случайного процесса с дискретным набором состояний. На рисунке 1 приведен пример одной из возможных реализаций случайного процесса, содержащего сообщения x_1 , x_2 , x_3 и x_4 [5].

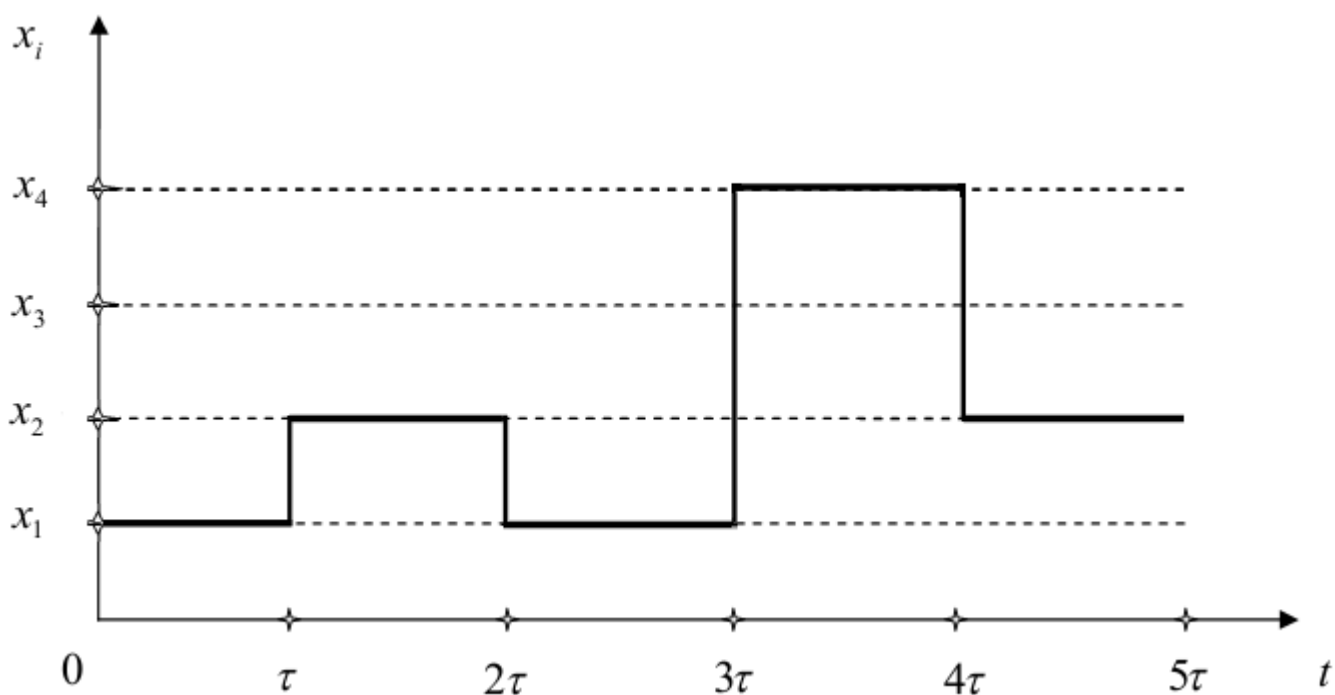


Рисунок 1 – Пример случайной последовательности сообщений источника

Если мощность множества $\{x_i\}$ равна M , то на временном интервале $(l \cdot \tau)$, где l – целое положительное число, можно построить M^l различных последовательностей сообщений возможных реализаций случайного процесса. Для рассмотренного выше примера $M = 4$, $l = 5$, поэтому число возможных реализаций случайного процесса будет равно $4^5=1024$.

2.3 Характеристики источника сообщений

Источник передает получателю одно сообщение из набора возможных. Получателю известен весь этот набор, но неизвестно, какое конкретно сообщение передавалось в данном временном интервале. В этом состоит неопределенность состояния получателя, которая исчезает при наличии уверенности, что сообщение принято и верно понято. Поскольку правильно принятое сообщение содержит в себе столько информации, сколько было неопределенности у получателя, то количество информации можно характеризовать ее энтропией, то есть мерой неопределенности сообщения.

Если источник сообщений $\{x_i, p(x_i)\}$, $1 \leq i \leq M$, имеющий алфавит из M букв, передал случайную для приемника букву-сообщение x_j , то количество переданной информации определяется, согласно Шеннону, формулой

$$H(x_j) = -\log_2 p(x_j). \quad (1)$$

Чем больше вероятность $p(x_j)$ передачи случайного сообщения x_j , тем меньшее количество информации при этом передается. Величина $H(x_j)$ называется случайной энтропией, поскольку соответствует передаче одного из случайных сообщений. Источник сообщений в целом характеризуют средней энтропией, определяемой следующим выражением:

$$H(X) = -\sum_{i=1}^M p(x_i) \log_2 p(x_i), \quad (2)$$

где $X = \{x_i\}$ – алфавит источника, включающий M сообщений.

Средняя энтропия $H(X)$ имеет максимальное значение, если все M сообщений равновероятны:

$$p(x_i) = \frac{1}{M}. \quad (3)$$

В этом случае максимальная средняя энтропия источника с таким «оптимизированным» алфавитом равна:

$$H_{\max}(X) = -\sum_{i=1}^M \frac{1}{M} \log_2 \frac{1}{M} = \log_2 M. \quad (4)$$

Если источник имеет алфавит с далеким от равномерного вероятностным распределением сообщений, то для такого источника вводят понятие избыточности источника [2]:

$$R = \frac{H_{\max}(X) - H(X)}{H_{\max}(X)} = 1 - \frac{H(X)}{H_{\max}(X)}. \quad (5)$$

Избыточность характеризует степень «неплотности упаковки» передаваемой источником информации при заданном наборе сообщений.

2.4 Модель кодера источника сообщений

Для снижения избыточности источника R применяют эффективное кодирование, при котором используют коды переменной длины, обратно зависящей от вероятности появления сообщения в передаваемой последовательности. Если, например, длина кода сообщения пропорциональна отрицательному логарифму вероятности передачи этого сообщения, то сообщениям с большей вероятностью будут соответствовать коды с меньшей длиной.

Одним из методов эффективного кодирования, обеспечивающих однозначное построение кода, является метод Хаффмана. При его использовании сначала строится оптимальное кодовое дерево, на базе которого затем строится соответствующее отображение «код – сообщение».

Алгоритм построения кода Хаффмана включает следующие пункты:

- 1) подсчет вероятностей $p(x_i)$ появления сообщений $\{x_i\}$ алфавита источника в исходной последовательности сообщений, если эти вероятности не известны заранее;
- 2) выписывание в столбец таблицы возможных сообщений источника в порядке уменьшения вероятности их появления;
- 3) объединение в получившемся упорядоченном списке двух последних сообщений в одно и запись его вместо этих двух объединенных сообщений с вероятностью, равной сумме вероятностей объединенных;
- 4) ранжирование полученного измененного списка по вероятностям сообщений и внесение их в следующий столбец таблицы;
- 5) повторение двух предыдущих пунктов до тех пор, пока в списке не останется всего одно сообщение с вероятностью, равной единице;
- 6) построение, в соответствии с полученной таблицей, кодового дерева в виде графа.

Переходы по дереву начинают из вершины графа, соответствующей сообщению с вероятностью единица. При этом ребрам графа присваиваются определенные символы кода, например, «1» при выходе из узла вверх и «0» при выходе из узла вниз. Путь по кодовому графу из вершины к одному из исходных сообщений дает кодовую комбинацию символов для этого сообщения. Для декодирования используется то же самое кодовое дерево.

Пусть, например, имеется источник сообщений с алфавитом мощностью $M = 9$, заданный приведенным в таблице 4 множеством пар $\{x_i, p(x_i)\}$ [5].

Таблица 4 – Алфавит источника и вероятности сообщений

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
$p(x_i)$	0,20	0,15	0,15	0,12	0,10	0,10	0,08	0,06	0,04

На рисунке 2 изображены все объединения, выполненные в соответствии с рассмотренным выше алгоритмом Хаффмана.










x_i	$p(x_i)$	1	2	3	4	5	6	7	8	Код
x_1	0,20									11
x_2	0,15									001
x_3	0,15									011
x_4	0,12									010
x_5	0,10									101
x_6	0,10									100
x_7	0,08									0001
x_8	0,06									00001
x_9	0,04									00000

Рисунок 2 – Схема реализации алгоритма Хаффмана

Для перехода из вершины графа к сообщению x_1 нужно два раза последовательно выйти из узла вверх, поэтому сообщению x_1 соответствует кодовая комбинация «11». Для перехода из вершины графа к сообщению x_5 нужно один раз выйти из узла вверх, из следующего узла вниз, а затем снова вверх. Это соответствует кодовой комбинации «101». Аналогично определяются кодовые комбинации для всех остальных сообщений.

2.5 Обобщенный источник сообщений

Характеристики обобщенного источника сообщений, включающего источник сообщений и соответствующий кодер, отличаются от характеристик исходного источника. Мощность алфавита полученного обобщенного источника равна двум, поскольку для формирования кодовых комбинаций используются только символы «0» и «1». Вероятности появления этих символов можно определить по известным вероятностям появления сообщений исходного источника и полученным кодовым комбинациям сформированного для этого источника кода Хаффмана. Для каждой из

этих кодовых комбинаций можно найти частоты появления символов «1» и «0». Известна и вероятность появления этого кодового слова. Например, для рассмотренного выше случая построения кода частота появления символа «1» при передаче кода, соответствующего сообщению x_3 , составляет $v_3(1) = \frac{2}{3}$, а вероятность передачи самого сообщения x_3 равна $p(x_3) = 0,15$. Поэтому вклад сообщения x_3 в вероятность передачи символа «1» равен $v_3(1) \cdot p(x_3) = \frac{2}{3} \cdot 0,15 = 0,1$.

Полная вероятность передачи символа «1» в последовательности сообщений обобщенного источника можно определить с помощью следующего выражения:

$$p(1) = \sum_{i=1}^M (v_i(1) \cdot p(x_i)), \quad (6)$$

где $v_i(1)$ – частота появления символа «1» при передаче кода, соответствующего сообщению x_i . Аналогично можно рассчитать вероятность передачи символа «0»:

$$p(0) = \sum_{i=1}^M (v_i(0) \cdot p(x_i)), \quad (7)$$

где $v_i(0)$ – частота появления символа «0» при передаче кода, соответствующего сообщению x_i . Отклонение найденных величин $p(1)$ и $p(0)$ от значения 0,5 характеризует степень отличия сформированного кода от эффективного.

2.6 Дискретные каналы передачи информации

Канал передачи информации называют дискретным по входу или выходу, если входные или выходные символы образуют конечное множество.

Любой реальный канал передачи информации содержит источники помех и шумов. Для рассматриваемой модели канала передачи указывается вероятность

того, что под действием шумов и помех передаваемый символ будет искажен настолько, что на выходе канала передачи он будет воспринят как некоторый другой символ из алфавита источника. Дискретный канал передачи называют каналом без памяти, если любой передаваемый символ искажается в этом канале независимо от значения ранее переданного символа.

Дискретный канал без памяти считается заданным, если известно множество $\{P(x_i|x_j)\}$ условных вероятностей $P(x_i|x_j)$ приема символа x_i вместо переданного символа x_j . Мощность алфавита рассматриваемого в курсовой работе обобщенного источника сообщений, в состав которого входит кодер, равна двум. Поэтому матрица переходов $\|P(x_i|x_j)\|$ состоит из четырех элементов, соответствующих разным комбинациям индексов $i=1, i=2, j=1$ и $j=2$. При этом таблица 2 содержит только два элемента матрицы. Два оставшихся элемента можно найти из условия

$$\sum_i P(x_i|x_j) = 1, \quad (8)$$

означающего, что на выходе канала передачи обязательно появится один из допустимых символов.

Канал передачи называется слабо зашумленным, если значение $P(x_i|x_i)$ близко к единице, то есть верное распознавание является почти достоверным событием: был передан символ x_i и принятый символ воспринят как x_i . И наоборот, канал передачи называется сильно зашумленным, если значение $P(x_i|x_i)$ существенно отличается от единицы.

Канал передачи называется симметричным, если входной алфавит совпадает с выходным и вероятности приема всех ошибочных символов равны, то есть

$$P(x_i|x_j) = \begin{cases} \frac{p_{\text{ош}}}{M-1}, & \text{если } i \neq j, \\ 1 - p_{\text{ош}}, & \text{если } i = j \end{cases}, \quad (9)$$

где $p_{\text{ош}}$ – вероятность появления ошибки.

Симметричный канал передачи называется стационарным, если вероятность ошибки не зависит от времени. В стационарном симметричном канале без памяти ошибки приема разных символов являются статистически независимыми и для него вероятность $P_n(r)$ возникновения r ошибок при передаче n символов можно определить с помощью следующего выражения:

$$P_n(r) = C_n^r \cdot p_{\text{ош}}^r \cdot (1 - p_{\text{ош}})^{n-r}, \quad (10)$$

где C_n^r – биномиальный коэффициент.

Из формулы (10) можно определить вероятность верного приема последовательности n символов:

$$P_n(0) = (1 - p_{\text{ош}})^n, \quad (11)$$

вероятность приема последовательности n символов, в которой имеется хотя бы одна ошибка:

$$P_n(r \geq 1) = 1 - P_n(0) = 1 - (1 - p_{\text{ош}})^n, \quad (12)$$

вероятность появления в последовательности n символов m и более ошибок и другие характеристики канала передачи.

2.7 Оптимальное правило восстановления принятого сообщения

Передача информации по зашумленному каналу приводит к искажению сигнала шумами и помехами, что может привести к недостоверному распознаванию переданного сообщения. В процессе выполнения курсовой работы необходимо найти оптимальное по критерию минимума вероятности средней ошибки правило выбора переданного символа x_j по принятому x_i . Поскольку символ x_j является лишь наиболее вероятным, а не истинно переданным, он будет обозначаться как \hat{x}_j .

Канал передачи задается парой множеств $\{p(x_i)\}$ и $\{p(x_i|x_j)\}$, первое из которых задает вероятность возникновения символа x_j на входе канала, а второе определяет вероятности перехода, то есть приема символа x_i , если в действительности передавался символ x_j . В соответствии с методом правдоподобия выбранному символу \hat{x}_j должно соответствовать максимальное значение условной вероятности $p(x_j|x_i)$:

$$\hat{x}_j = \arg \max_j \{p(x_j|x_i)\}. \quad (13)$$

Указанная условная вероятность принципиально отличается от определяющей канал передачи вероятности перехода. Она задается при условии известного принятого символа, а переходная вероятность – при условии известного переданного символа. Эти вероятности связаны друг с другом следующим выражением:

$$p(x_i, x_j) = p(x_i) \cdot p(x_j|x_i) = p(x_j) \cdot p(x_i|x_j), \quad (14)$$

из которого можно найти:

$$p(x_j|x_i) = \frac{p(x_j) \cdot p(x_i|x_j)}{p(x_i)}. \quad (15)$$

Подставляя соотношение (15) в формулу (13), можно получить, что с наибольшей достоверностью переданный символ может быть определен с помощью выражения

$$\hat{x}_j = \arg \max_j \left\{ \frac{p(x_j) \cdot p(x_i|x_j)}{p(x_i)} \right\}. \quad (16)$$

В этом случае вероятность ошибки определения переданного символа при принятом символе x_i будет минимальной:

$$p_{\text{ош}}(x_i) = 1 - \frac{p(\hat{x}_j) \cdot p(x_i|\hat{x}_j)}{p(x_i)}. \quad (17)$$

Средняя вероятность ошибки, полученная суммированием по всем возможным принимаемым символам, составляет

$$P_{\text{ош}} = \sum_i (p(x_i) \cdot p_{\text{ош}}(x_i)) = \sum_i \left(p(x_i) \cdot \left(1 - \frac{p(\hat{x}_j) \cdot p(x_i|\hat{x}_j)}{p(x_i)} \right) \right) = 1 - \sum_i (p(\hat{x}_j) \cdot p(x_i|\hat{x}_j)). \quad (18)$$

Из соотношения (18) следует, что для минимизации ошибки нужно каждому принятому символу x_i ставить в соответствие такой символ \hat{x}_j , чтобы выражение

$$p(\hat{x}_j) \cdot p(x_i|\hat{x}_j) = p(x_i, \hat{x}_j) \quad (19)$$

принимало максимальное значение. Это правило можно реализовать с помощью алгоритма, который иллюстрирует следующий пример [5].

Пусть для символов x_1 , x_2 и x_3 определены вероятности передачи по каналу $p(x_1) = 0,3$, $p(x_2) = 0,5$, $p(x_3) = 0,2$ и матрица моделирующая шум и помехи переходных вероятностей $p_{i,j} = p(x_i|x_j)$:

$$\|p_{i,j}\| = \begin{vmatrix} 0,7 & 0,45 & 0,3 \\ 0,2 & 0,45 & 0,35 \\ 0,1 & 0,1 & 0,45 \end{vmatrix}. \quad (20)$$

Для заданных значений вероятностей выполняются построения, представленные на рисунке 3.

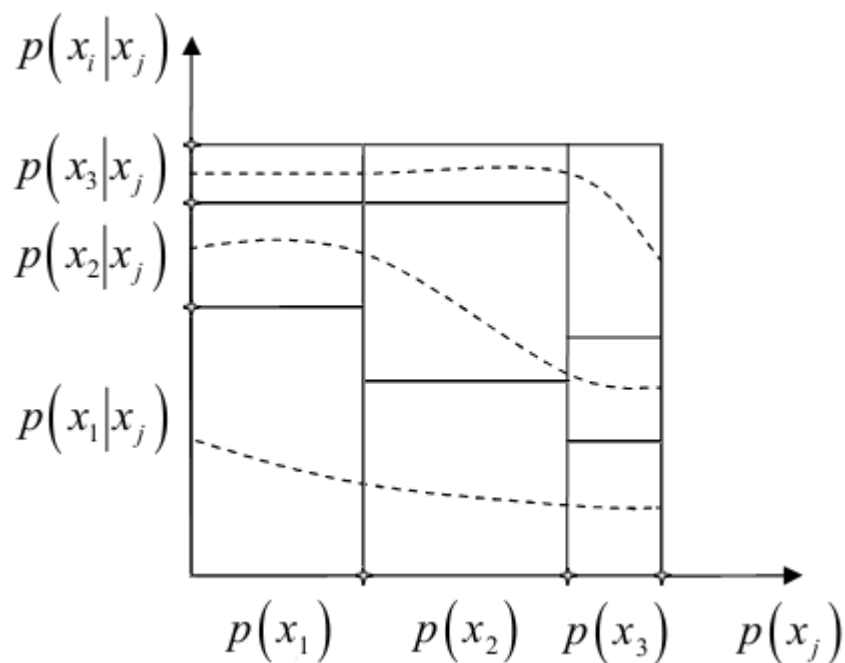


Рисунок 3 – Определение оптимального правила выбора переданного символа

По горизонтали на рисунке 3 отложены отрезки, соответствующие величинам $p(x_1)$, $p(x_2)$ и $p(x_3)$, а по вертикали в каждом столбце отложены значения условных вероятностей $p(x_i|x_j)$. Площади получившихся прямоугольников равны вероятностям

$$p(x_i, x_j) = p(x_j) \cdot p(x_i | x_j). \quad (21)$$

Согласно выражению (18), в каждой выделенной пунктиром на рисунке 3 «строчке», соответствующей принятому символу x_i , нужно найти прямоугольник с максимальной площадью. Его номер j по горизонтали покажет наиболее вероятный переданный источником символ \hat{x}_j .

Нижняя строка на рисунке 3 относится к случаю приема символа x_1 . Для переданного символа x_1 вероятность $p(x_1, x_1) = 0,21$, для символа x_2 – $p(x_1, x_2) = 0,225$, для символа x_3 – $p(x_1, x_3) = 0,06$. Из перечисленных трех значений совместной вероятности $p(x_i, x_j)$ наибольшее достигается для пары « x_1 (принятый символ), x_2 (переданный символ)»: $p(x_1, x_2) = 0,225$. Это означает, что при приеме символа x_1 нужно считать, что передавался символ x_2 . Кажущийся парадокс объясняется существенно большей вероятностью передачи $p(x_2) = 0,5$ символа x_2 по сравнению с вероятностью передачи $p(x_1) = 0,3$ символа x_1 . В рассмотренном примере оптимальное по критерию минимума вероятности средней ошибки правило выбора переданного символа x_j по принятому x_i можно записать в следующем виде:

$$\begin{aligned} x_1 &\rightarrow x_2, \\ x_2 &\rightarrow x_2, \\ x_3 &\rightarrow x_3. \end{aligned} \quad (22)$$

Список использованных источников

1 Темников, Ф.Е. Теоретические основы информационной техники: учебное пособие для вузов / Ф.Е. Темников, В.А. Афонин, В.И. Дмитриев. – М.: Энергия, 1979. – 512 с.

2 Колесник, В.Д. Курс теории информации: учебное пособие для вузов / В.Д. Колесник, Г.Ш. Полтырев. – М.: Наука, 1982. – 416 с.

3 Котоусов, А.С. Теория информации: учебное пособие для вузов / А.С. Котоусов. – М.: Радио и связь, 2003. – 80 с.

4 Стратонович, Р.Л. Теория информации / Р.Л. Стратонович. – М.: Сов. радио, 1975. – 424 с.

5 Кузнецов, В.Л. Теория информации. Пособие по выполнению курсовой работы / В.Л. Кузнецов. – М.: МГТУГА, 2009. – 21с.

6 Кавчук, С.В. Сборник примеров и задач по теории информации. Руководство для практических занятий на базе Mathcad 6.0 Plus / С.В. Кавчук. – Таганрог: Изд-во ТРТУ, 2002. – 64 с.

7 СТО 02069024.101 – 2015. Работы студенческие. Общие требования и правила оформления. Введ. 2016-02-08. – Оренбург: ОГУ, 2015. – 85 с.