**Тема реферату: "КОДИРОВАНИЕ"**

**1. ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ ТЕОРИИ КОДИРОВАНИЯ**

Оптимальным статистическим (экономным) кодированием называется кодирование, при котором обеспечивается распределение времени на передачу отдельных символов алфавита в зависимости от априорных вероятностей их появления:

; (1)



где Cп - пропускная способность канала; pi - априорная вероятность i –й кодовой комбинации; ti -длительность i-й кодовой комбинации.

Оптимальными неравномерными кодами (ОНК) - называются коды, в которых символы алфавита кодируются кодовыми словами минима-льной средней длины.

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 слова, при этом, средняя длина кодовых комбинаций имеет минимально-возможную длину.

При таком кодировании избыточность кода, которая вызвана неравной вероятностью символов алфавита, сводится к минимуму (практически к нулю). Оптимальные коды являются неравномерными блочными кодами, при их построении необходимо обеспечить однозначность декодирования. Префиксным (неприводимым)- называется код, в котором ни одна кодовая комбинация не является началом другой. Для обеспечения этого свойства кодовые комбинации должны записываться от корня кодового дерева.

Возможность однозначного декодирования неравномерного кода обеспечивается выполнением требования разделимости (префиксности) кодовых комбинаций.

При неравномерном кодировании производится сжатие данных. Сжатие данных используется как при хранении данных в памяти, так и при их передаче. Оптимальное кодирование можно использовать только в каналах без помех или в случае низкой требовательности к достоверности передаваемой информации.

Существует много методов оптимального, статистического кодирования. Наиболее часто используют оптимальное кодирование по методу Шеннона - Фано и Хаффмена.

2. КОД ШЕННОНА-ФАНО

Кодирование по методу Шеннона - Фано осуществляется следующим образом:

1. Множество символов, из которых формируются сообщения, записываются в порядке убывания их априорных вероятностей.

2. Дальнейшее построение кода производится методом последовательного деления пополам. Символы сообщения разбиваются на две группы с примерно равными вероятностями (т. к. при отсутствии статистической связи между символами скорость передачи максимальна при условии равной вероятности передачи символов). Если равной вероятности в подгруппах достичь нельзя, то желательно чтобы суммарная вероятность нижней подгруппы была больше верхней.

3. Всем символам верхней группы приписывается кодовый символ 1, а символам нижней - 0. Можно наоборот, т. к. для кодовой реализации безразлично 0 или 1, но с точки зрения мощности, лучше, если в кодовой комбинации меньше единиц.

4. Затем каждая подгруппа аналогичным образом разбивается на подгруппы по возможности с одинаковыми вероятностями. Разбиение осуществляется до тех пор, пока в каждой подгруппе останется по одному символу.

Пример построения кода приведен в таблице 1.

Таблица 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ai | pi | Разбиение | Кодовая комбинация | Длина |
| a1  a2  a3  a4 | 1/2  1/4  1/8  1/8 | }1  0 }1  0 }1  }0 | 1  01  001  000 | τ  2τ  3τ  3τ |

Построенный код является префиксным.

Например: полученная кодовая последовательность 11100001 однозначно декодируется как:

1 1 1 000 1 01 => a1 a1 a1 a4 a1 a2 .

a1 a1 a1 a4 a1 a2

Применяя статистическое кодирование можно получить результат, близкий к идеальному кодированию по Шеннону.

Средняя длина кодовой комбинации, при использовании двоичного кода в качестве вторичного, равна

, (2)



где li - длина i-й комбинации; N -основание первичного кода.

Эффективность ОНК максимальна при

; . (3)



Коэффициент относительной эффективности (коэффициент использования пропускной способности) равен

. (4)



Коэффициент статистического сжатия (уменьшение количества двоичных разрядов на символ сообщения при использовании статистического кодирования по сравнению с обычным кодированием) равен

. (5)



Для рассмотренного примера при длительности символа кодовой комбинации (0 или 1) равной τ средняя длина и средняя длительность кодовой комбинации, соответственно равны:



Энтропия источника равна



При этом: Коэ = 1,75/1,75 = 1; Кcc = 2/1,75 = 1,14.

Скорость передачи информации

, (6)



т. е. коэффициент использования пропускной способности канала равен 1, а значит, имеет место идеальное использование канала (оптимальное статистическое кодирование).

Если подгруппы имеют не одинаковую суммарную вероятность, то коэффициент меньше 1. Для равномерного кода , при этом



.



Недостаток кода ОНК - низкая помехоустойчивость, т. к. потеря одного разряда может означать потерю символа.

3. КОД ХАФФМЕНА

Кодирование по методу Хаффмена осуществляется следующим об-разом:

1. Все подлежащие кодированию символы записываются в порядке убывания их априорных вероятностей. Если некоторые символы имеют одинаковые вероятности, то их располагают рядом в произвольном порядке.

2. Выбирают символы с минимальными вероятностями по 2 и одному приписывают 0, а другому 1.

3. Выбранные символы объединяют в промежуточные символы с суммарной вероятностью.

4. Снова находят пару символов с наименьшими вероятностями и поступают аналогично.

В таблице 2 приведен пример кодирования по методу Хаффмена для источника сообщений с заданными вероятностями символов алфавита:

x1 = 0,4; x2 = x5 = 0,2; x3 = 0,1; x4 = x6 = 0,05.

Таблица 2

|  |  |  |  |
| --- | --- | --- | --- |
| Символ | pi | Граф кода Хаффмена | Код |
| x1  x2  x5  x 3  x 4  x 6 | 0,4  0,2  0,2  0,1  0,05  0,05 | 1  (1,0)  1 0  (0,6)  1 0  (0,4)  1 0  (0,2)  1 0  (0,1)  0 | 1  01  001  0001  00001  00000 |



Энтропия источника равна



Средняя длина кодовой комбинации данного кода



Длина кодовой комбинации примитивного кода определяется соотношением

(7)



Округляя до ближайшего целого в большую сторону, получим l = 3.

Эффективность ОНК максимальна, если .



Коэффициент относительной эффективности равен

.



Коэффициент статистического сжатия равен

.



Неравномерный код можно передавать блоками заданной длины, а на приемной стороне декодировать всю последовательность.

Пример 1. Построить оптимальные неравномерные коды (ОНК) по методу Шеннона-Фано и по методу Хаффмена для передачи сообщений, в которых вероятности символов первичного алфавита равны:

p(a1) =0,1; p(a2) =0,07; p(a3) =0,02; p(a4) =0,17;

p(a5) =0,42; p(a6) =0,09; p(a7) =0,08; p(a8) =0,05.

Оценить эффективность каждого кода, т. е. насколько они близки к оптимальным. Определить емкость (пропускную способность) канала связи для каждого кода, если скорость передачи двоичных символов (V = 1/τ) равна 1000 симв/с, т.е. время передачи одного символа вторичного алфавита (двоичного символа) равна τ = 0,001с = 1мкс.

Решение: Построим код по методу Шеннона-Фано, используя сле-дующий алгоритм:

1. Символы сообщения располагаем в порядке убывания их априорных вероятностей.

2. Исходный ансамбль кодируемых символов разбиваем на две группы с примерно равными вероятностями (лучше, если суммарная вероятность верхней группы меньше).

3. Верхней группе присваиваем символ 1, а нижней 0.

4. Процесс деления повторяем до тех пор, пока в каждой подгруппе останется по одному символу.

Процесс построения кода приведем в таблице 3.

Таблица 3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ai | p(ai) | Разбиение | Код | li | pili |
| a5  a4  a1  a6  a7  a2  a8  a3 | 0,42  0,17  0,10  0,09  0,08  0,07  0,05  0,02 | }0  1 0 }0  }1  1 0 }0  }1  1 }0  }0  }1 | 0  100  101  1100  1101  1110  11110  11111 | 1  3  3  4  4  4  5  5 | 0,42  0,51  0,3  0,36  0,32  0,28  0,25  0,1 |

. .



Могут быть и другие варианты построения кода, но lср при этом не меняется.

Определим энтропию источника сообщений:



= -(0,42⋅ log20,42+0,17 ⋅log20,17+0,1⋅log20,1+0,09⋅log20,09+0,08⋅log20,08+

+0,07⋅log20,07+0,05⋅log20,05+0,02⋅log20,02) = 0,5256+0,4346+0,3322+

+0,3126+0,2915+0,2686+0,2161+0,1129 = 2,49 бит/симв.

Оценим эффективность построенного кода, которая определяется коэффициентами относительной эффективности и статистического сжатия.

Коэффициент относительной эффективности равен



Коэффициент статистического сжатия равен

.



Необходимая пропускная способность канала связи для передачи сообщений оптимальными кодами вычисляется по формуле

.



Для полученного кода она равна

С = 103 ⋅2,54 = 2,54 Кбит/с.

Построим код по методу Хаффмена, используя следующий алгоритм:

1. Символы первичного алфавита располагаем в порядке убывания их вероятностей.

2. Последние два символа объединяем в один, с суммарной вероятностью.

3. Упорядочиваем символы с учетом вновь образованных и последние два символа объединяем в один с суммарной вероятностью. Эту процедуру повторяем до тех пор, пока не останется два символа.

4. Строим кодовое дерево, вершиной которого является 1 (суммарная вероятность всех символов сообщения).

При построении дерева символы, в принципе, можно не упорядочивать.

Процесс построения кода приведен в таблице 4.

Коэффициент статистического сжатия равен

.



Коэффициент относительной эффективности

.



Необходимая пропускная способность канала связи

Кбит/c.



Tаблица 4

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| аi | p(ai) | Кодовое дерево | Код | | li | | pili |
| a5  a4  a1  a6  a7  a2  a8  a3 | 0,42  0,17  0,10  0,09  0,08  0,07  0,05  0,02 | 1 (1,0)  1 1 0  (0,34) (0,58)  0 1 0  (0,24)  1 0  (0,17)  0  1  (0,14)  1 0  (0,7)  0 | 1  011  001  0101  0100  0001  00001  00000 | 1  3  3  4  4  4  5  5 | | 0,42  0,51  0,3  0,36  0,32  0,28  0,25  0,1 | |

. .



Построенные коды Шеннона – Фано и Хаффмена равноценны, т. к. они имеют одинаковые показатели эффективности. Оба кода отличаются от оптимального на 2%.

# 4. ПОМЕХОУСТОЙЧИВОЕ КОДИРОВАНИЕ

Помехоустойчивость- способность системы осуществлять прием информации в условиях наличия помех в линии связи и искажений во внутри аппаратных трактах. Помехоустойчивость обеспечивает надежность и достоверность передаваемой информации (данных). Мы будем в основном рассматривать двоичные коды. Двоичные (коды) данные передаются между вычислительными терминалами, летательными аппаратами, спутниками и т. д.

Передача данных в вычислительных системах чувствительна к малой доле ошибке, т. к. одиночная ошибка может существенно нарушить процесс вычислений.

Наиболее часто ошибки появляются в УВВ, шинах, устройствах памяти. УВВ содержат большое количество элементов, ошибки обуславливаются старением элементов, ухудшением качества электрических соединений, расфазировкой сигналов. Значительная часть ошибок приходится на ОП, вследствие отказа отдельных ИС либо всей ИС, ошибок связанных с флуктуацией напряжения питания и т. д.

В системах со многими пользователями и разделением по времени длинные двоичные сообщения разделяются на пакеты.

Сообщения, представленные длинными последовательностями битов, обычно разбиваются на более короткие последовательности битов, называемые пакетами. Пакеты можно передать по сети как независимые объекты и собирать из них сообщение на конечном пункте. Пакет, снабженный именем и управляющими битами в начале и в конце, называется кадром. Управление линией передачи данных осуществляется по специальному алгоритму, называемому протоколом.

Наличие помех ставит дополнительные требования к методам кодирования. Для защиты информации от помех необходимо вводить в том или ином виде избыточность: повышение мощности сигнала; повторение сообщений; увеличение длинны кодовой комбинации и т. д.

Увеличение мощности сигналов приводит к усложнению и удорожанию аппаратуры, кроме того, в некоторых системах передачи информации имеется ограничение на передаваемую мощность, например, спутниковая связь.

Повторная передача сообщений требует наличия буферов для хранения информации и наличия обратной связи для подтверждения достоверности переданной информации. При этом, значительно падает скорость передачи информации, кроме того этот метод не всегда м. б. использован, например, в система реального времени.

Одним из наиболее эффективных методов повышения достоверности и надежности передачи данных является помехоустойчивое кодирование, позволяющее за счет внесения дополнительной избыточности (увеличение минимального кодового расстояния) в кодовых комбинациях передаваемых сообщений обеспечить возможность обнаружения и исправления одиночных, кратных и групповых ошибок.

Минимальное кодовое расстояние характеризует помехоустойчивость и избыточность сообщений. В зависимости от величины минимального кодового расстояния существуют коды, обнаруживающие и исправляющие ошибки.

Кодовое расстояние - d определяется как количество единиц в результате суммирования по модулю два двух кодовых комбинаций. Минимальное кодовое расстояние d0 - минимальное из кодовых расстояний всех возможных кодовых комбинаций.

Для обнаружения r ошибок минимальное кодовое расстояние равно:

d0 ≥ r+1. (8)

Для обнаружения r ошибок и исправления s ошибок минимальное кодовое расстояние равно:

d0 ≥ r+s+1. (9)

Только для исправления ошибок минимальное кодовое расстояние равно:

d0 ≥ 2s+1. (10)

**5. ОБНАРУЖИВАЮЩИЕ КОДЫ**

Обнаруживающие коды - это коды, позволяющие обнаружить ошибку, но не исправить ее. Простейший способ обнаружения ошибки это добавление к последовательности битов данных еще одного бита-бита проверки на четность (нечетность) значение, которого равно сумме по модулю два исходной последовательности битов. Чаще организуется проверка на нечетность.

В символьном коде ASCII к семи битам кода добавляется восьмой бит проверки на четность - k1.

S1 S2 S3 S4 S5 S6 S7 K1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |

Однобитовая проверка позволяет обнаружить любую единичную ошибку, две ошибки обнаружить нельзя, в общем случае обнаруживается любое нечетное количество ошибок.

Внесение избыточности за счет увеличения длины кодовой комбинации приводит к снижению скорости передачи информации.

Если скорость идеально использует канал, то

. (11)



Если кодовая комбинация длиной n содержит k информационных и m контрольных разрядов (n = k + m), то

.



Для кода ASCII n = 8 и k = 7

,



т. е. введения одного избыточного разряда приводит к уменьшению пропускной способности канала связи на 12,5%.

Чаще всего шумы (молнии, разрыв и т.д.) порождают длинные пакеты ошибок и вероятность четного и нечетного числа ошибок одинакова, а значит и однобитовая проверка не эффективна.

Проверка на четность по вертикали и горизонтали. При этом последовательность битов данных перестраивается в двухмерный массив, и вычисляются биты на четность, как для каждой строки, так и для каждого столбца.

При этом можно обнаружить несколько ошибок, если они не располагаются в одинаковых строках и столбцах.

Чаще всего используется при передаче данных кода ASCII; каждый символ можно считать строкой массива. Такая проверка может не только установить факт ошибки, но и обнаружить ее место, а значит, есть принципиальная возможность ее исправления, хотя это практически не используется.

1 0 1 1 0 1 1 1

0 1 0 0 0 1 0 0

1 0 1 0 0 1 0 1

1 1 0 0 1 0 1 0

0 0 0 1 0 1 0 0

1 0 0 0 1 0 0

После обнаружения ошибок иногда можно повторить передачу сообщений, иногда после обнаружения ошибки предпринимается вторая и даже третья попытка передачи сообщения.

Проверка на четность широко используется на ЭВМ, как на аппаратном, так и на программном уровне.

Например, при считывании с магнитной ленты в случае, когда условие на четность не выполняется, то производится повторное считывание, т. е. если произошла малая потеря намагниченности, то после второй попытки может быть считывание произойдет правильно.

Пример 1. Символы алфавита источника кодируются семиразрядным двоичным кодом с весом кодовых векторов (количеством единиц в кодовой комбинации) w = 3. Определить необходимую мощность кода и его избыточность.

Решение: Мощность семиразрядного кода равна N = 27 = 128.

Так как для кодирования используются только кодовые вектора с весом три , то количество таких векторов в семиразрядном коде равно



Избыточность кода равна R = 1 – log2K/ log2N = 0,265.

**СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ**

1. Семенюк В. В. Экономное кодирование дискретной информации. – СПб.: СПб ГИТМО (ТУ), 2001;
2. Мастрюков Д. Алгоритмы сжатия информации. Ч. 1. Сжатие по Хаффмену //Монитор, 1993. – № 7 – 8 – С. 14 – 20;
3. Мастрюков Д. Алгоритмы сжатия информации. Ч. 2. Арифметическое кодирование //Монитор, 1994 – № 1 – С. 20 – 23;
4. Ф.Дж.Мак-Вильямс, Н.Дж.А.Слоэн, Теория кодов, исправляющих ошибки, Москва, “Связь”, 1979.
5. .Лидл, Г.Нидеррайтер, Конечные поля, Т. 1,2, Москва, “Мир”, 1988.
6. Т.Касами, Н.Токура, Е.Ивадари, Я.Инагаки, Теория кодирования, Москва, “Мир”, 1978.
7. У.Петерсон, Э.Уэлдон, Коды, исправляющие ошибки, Москва, “Мир”, 1976.
8. Э.Берлекэмп, Алгебраическая теория кодирования, Москва, “Мир”, 1971.
9. Дискретная математика и математические вопросы кибернетики. Т.1. /Ю.Л. Васильев, Ф. Я. Ветухновский, В. В. Глаголев, Ю. И. Журавлев, В. И. Левенштейн, С. В. Яблонский. Под общей редакцией С. В. Яблонского и О. Б. Лупанова. – М.: Главная редакция физико – математической литературы изд–ва «Наука», 1974
10. Лидовский В. В. Теория информации: Учебное пособие. — М.: Компания Спутник+, 2004