Курс: Теория информации и кодирования

Тема: Кодирование

Содержание

1. Кодирование. Основные понятия и определения

2. Классификация кодов

3. Способы представления кодов

3.1 Матричное представление кодов

3.2 Представление кодов в виде кодовых деревьев

3.3 Представление кодов в виде многочленов

3.4 Геометрическое представление кодов

Список литературы

## 1. Кодирование. Основные понятия и определения

Рассмотрим основные понятия, связанные с кодированием информации. Для передачи в канал связи сообщения преобразуются в сигналы. Символы, при помощи которых создаются сообщения, образуют первичный алфавит, при этом каждый символ характеризуется вероятностью его появления в сообщении. Каждому сообщению однозначно соответствует сигнал, представляющий определенную последовательность элементарных дискретных символов, называемых кодовыми комбинациями. ***Кодирование*** - это преобразование сообщений в сигнал, т.е. преобразование сообщений в кодовые комбинации. ***Код*** - система соответствия между элементами сообщений и кодовыми комбинациями. ***Кодер*** - устройство, осуществляющее кодирование. ***Декодер* -** устройство, осуществляющее обратную операцию, т.е. преобразование кодовой комбинации в сообщение. ***Алфавит*** - множество возможных элементов кода, т.е. элементарных символов (кодовых символов) *X = {xi},* где *i = 1, 2,..., m.* Количество элементов кода - *m* называется его ***основанием***. Для двоичного кода *xi = {0, 1}* и *m = 2.* Конечная последовательность символов данного алфавита называется ***кодовой комбинацией*** (кодовым словом). Число элементов в кодовой комбинации - *n* называется ***значностью*** (длиной комбинации). Число различных кодовых комбинаций (*N = mn*) называется ***объемом*** или мощностью кода.

Если *N0* - число сообщений источника, то *N ≥ N0*. Множество состояний кода должно покрывать множество состояний объекта. Полный равномерный *n* - значный код с основанием *m* содержит *N = mn* кодовых комбинаций. Такой код называется ***примитивным.***

## 2. Классификация кодов

Коды можно классифицировать по различным признакам:

1. По основанию (количеству символов в алфавите): ***бинарные*** (двоичные m=2) и ***не бинарные*** (m ≠ 2).

2. По длине кодовых комбинаций (слов):

***равномерные*** - если все кодовые комбинации имеют одинаковую длину;

***неравномерные*** - если длина кодовой комбинации не постоянна.

3. По способу передачи:

***последовательные*** и ***параллельные;***

***блочные -*** данные сначала помещаются в буфер, а потом передаются в канал и бинарные ***непрерывные***.

4. По помехоустойчивости:

***простые*** (примитивные, полные) - для передачи информации используют все возможные кодовые комбинации (без избыточности);

***корректирующие*** (помехозащищенные) - для передачи сообщений используют не все, а только часть (разрешенных) кодовых комбинаций.

5. В зависимости от назначения и применения условно можно выделить следующие типы кодов:

***Внутренние коды -*** этокоды, используемые внутри устройств. Это машинные коды, а также коды, базирующиеся на использовании позиционных систем счисления (двоичный, десятичный, двоично-десятичный, восьмеричный, шестнадцатеричный и др.). Наиболее распространенным кодом в ЭВМ является двоичный код, который позволяет просто реализовать аппаратно устройства для хранения, обработки и передачи данных в двоичном коде. Он обеспечивает высокую надежность устройств и простоту выполнения операций над данными в двоичном коде. Двоичные данные, объединенные в группы по 4, образуют шестнадцатеричный код, который хорошо согласуется с архитектурой ЭВМ, работающей с данными кратными байту (8 бит).

***Коды для обмена данными*** ***и их передачи по каналам связи***. Широкое распространение в ПК получил код ASCII (American Standard Code for Information Interchange). ASCII - это 7-битный код буквенно-цифровых и других символов. Поскольку ЭВМ работают с байтами, то 8-й разряд используется для синхронизации или проверки на четность, или расширения кода. В ЭВМ фирмы IBM используется расширенный двоично-десятичный код для обмена информацией EBCDIC (Extended Binary Coded Decimal Interchange Code).

В каналах связи широко используется телетайпный код МККТТ (международный консультативный комитет по телефонии и телеграфии) и его модификации (МТК и др.).

При кодировании информации для передачи по каналам связи, в том числе внутри аппаратным трактам, используются коды, обеспечивающие максимальную скорость передачи информации, за счет ее сжатия и устранения избыточности (например: коды Хаффмана и Шеннона-Фано), и коды обеспечивающие достоверность передачи данных, за счет введения избыточности в передаваемые сообщения (например: групповые коды, Хэмминга, циклические и их разновидности).

***Коды для специальных применений*** - это коды, предназначенные для решения специальных задач передачи и обработки данных. Примерами таких кодов является циклический код Грея, который широко используется в АЦП угловых и линейных перемещений. Коды Фибоначчи используются для построения быстродействующих и помехоустойчивых АЦП.

Основное внимание в курсе уделено кодам для обмена данными и их передачи по каналам связи.

**ЦЕЛИ КОДИРОВАНИЯ:**

1) Повышение эффективности передачи данных, за счет достижения максимальной скорости передачи данных.

2) Повышение помехоустойчивости при передаче данных.

В соответствии с этими целями теория кодирования развивается в двух основных направлениях:

1. ***Теория экономичного (эффективного, оптимального) кодирования*** занимается поиском кодов, позволяющих в каналах без помех повысить эффективность передачи информации за счет устранения избыточности источника и наилучшего согласования скорости передачи данных с пропускной способностью канала связи.

2. ***Теория помехоустойчивого кодирования*** занимается поиском кодов, повышающих достоверность передачи информации в каналах с помехами.

## 3. Способы представления кодов

В зависимости от применяемых методов кодирования, используют различные математические модели кодов, при этом наиболее часто применяется представление кодов в виде: кодовых матриц; кодовых деревьев; многочленов; геометрических фигур и т.д.

## 3.1 Матричное представление кодов

Используется для представления равномерных *n -* значных кодов. Для примитивного (полного и равномерного) кода матрица содержит *n* - столбцов и *2n* - строк, т.е. код использует все сочетания. Для помехоустойчивых (корректирующих, обнаруживающих и исправляющих ошибки) матрица содержит *n* - столбцов (*n = k+m*, где *k-*число информационных, а *m* - число проверочных разрядов) и *2k* - строк (где *2k* - число разрешенных кодовых комбинаций). При больших значениях *n* и *k* матрица будет слишком громоздкой, при этом код записывается в сокращенном виде. Матричное представление кодов используется, например, в линейных групповых кодах, кодах Хэмминга и т.д.

## 3.2 Представление кодов в виде кодовых деревьев

***Кодовое дерево*** - связной граф, не содержащий циклов. ***Связной граф*** - граф, в котором для любой пары вершин существует путь, соединяющий эти вершины. Граф состоит из узлов (вершин) и ребер (ветвей), соединяющих узлы, расположенные на разных уровнях. Для построения дерева равномерного двоичного кода выбирают вершину называемую корнем дерева (истоком) и из нее проводят ребра в следующие две вершины и т.д.

Пример кодового дерева для полного кода приведен на рис.1.

 1 0

 1 0 1 0

 1 0 1 0 1 0 1 0

 111 110 101 100 011 010 001 000

Рис.1. Дерево для полного двоичного кода при *n* = 3

Дерево помехоустойчивого кода строится на основе дерева полного кода путем вычеркивания запрещенных кодовых комбинаций. Для дерева неравномерного кода используется взвешенный граф, при этом на ребрах дерева указываются вероятность переходов. Представление кода в виде кодового дерева используется, например, в кодах Хаффмена.

## 3.3 Представление кодов в виде многочленов

Представление кодов в виде полиномов основано на подобии (изоморфизме) пространства двоичных *n* - последовательностей и пространства полиномов степени не выше *n - 1*.

Код для любой системы счисления с основанием *Х* может быть представлен в виде:

*G (x) = an-1 xn-1+ an-2 xn-2+... + a1 x+ a0 =,*

где *аi -* цифры данной системы счисления (в двоичной 0 и 1);

*х* - символическая (фиктивная) переменная, показатель степени которой соответствует номерам разрядов двоичного числа-

Например: Кодовая комбинация 1010110 может быть представлена в виде:

*G (x) =1⋅x6+0⋅x5+1⋅x4+0⋅x3+1⋅x2+1⋅x1+0⋅x0 =x6+x4+x2+x=10101*

При этом операции над кодами эквивалентны операциям над многочленами. Представление кодов в виде полиномов используется например, в циклических кодах.

## 3.4 Геометрическое представление кодов

Любая комбинация *n* - разрядного двоичного кода может быть представлена как вершина *n* - мерного единичного куба, т.е. куба с длиной ребра равной 1. Для двухэлементного кода (*n = 2*) кодовые комбинации располагаются в вершинах квадрата. Для трехэлементного кода

(*n = 3*) - в вершинах единичного куба (рис.2).

В общем случае *n* мерный куб имеет *2n* вершин, что соответствует набору кодовых комбинаций *2n*.

*n = 2 n = 3*

Рис.2. Геометрическая модель двоичного кода

**Геометрическая интерпретация кодового расстояния**. Кодовое расстояние - минимальное число ребер, которое необходимо пройти, чтобы попасть из одной кодовой комбинации в другую. Кодовое расстояние характеризует помехоустойчивость кода.

## Список литературы

1. Кловский Д.Д. Теория передачи сигналов. -М.: Связь, 1984.
2. Кудряшов Б.Д. Теория информации. Учебник для вузов Изд-во ПИТЕР, 2008. - 320с.
3. Рябко Б.Я., Фионов А.Н. Эффективный метод адаптивного арифметического кодирования для источников с большими алфавитами // Проблемы передачи информации. - 1999. - Т.35, Вып. - С.95 - 108.
4. Семенюк В.В. Экономное кодирование дискретной информации. - СПб.: СПбГИТМО (ТУ), 2001
5. Дмитриев В.И. Прикладная теория информации. М.: Высшая школа, 1989.
6. Нефедов В.Н., Осипова В.А. Курс дискретной математики. М.: МАИ, 1992.
7. Колесник В.Д., Полтырев Г.Ш. Курс теории информации. М.: Наука, 2006.