|  |  |
| --- | --- |
| Череповецкий государственный университет  Кафедра «Математического и программного обеспечения ЭВМ» | |
| ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ  по дисциплине «Теория информации»  Эффективное кодирование по методу Шеннона-Фано и Хаффмана | |
|  | Принял:  преподаватель Е.Н. Руденко    подпись, дата  Выполнил:  студент гр. 1ПИб-02-1оп-22  Остапенко Степан Ярославович  подпись, дата |
| Череповец, 2023 | |

**Реферат**

Предметом исследования являются теорема Шеннона, а также методика Шеннона-Фано и методика Хаффмана.

Цель работы – исследование и практическое применение методов Шеннона-Фано и Хаффмана для эффективного кодирования информации.

В ходе работы осваиваются методы Шеннона-Фано и Хаффмана и используются для решения задач, требующих эффективного кодирования информации.

**Ход работы**

**Задача 1**

Имеется алфавит символов и их вероятности, с которыми они встречаются в тексте. Построить таблицу кодов символов методом Шеннона-Фано. Закодировать сообщение «вилка» и раскодировать заданную последовательность кодов.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| а | в | л | И | е | С | К |
| 0,3 | 0,2 | 0,15 | 0,1 | 0,1 | 0,08 | 0,07 |

Решение:

Построим таблицу кодов методом Шеннона-Фано:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Символ | вероятности | символы кода | | | | |
| 1 | 2 | 3 | 4 | код |
| а | 0,3 | 0 | 0 |  |  | 00 |
| в | 0,2 | 1 | 01 |
| л | 0,15 | 1 | 0 | 0 | 100 |
| и | 0,1 | 1 | 101 |
| е | 0,1 | 1 | 0 | 110 |
| с | 0,08 | 1 | 0 | 1110 |
| к | 0,07 | 1 | 1111 |

Закодируем сообщение «вилка»:

в = 01

и = 101

л = 100

к = 1111

а = 00

Результат: 01101100111100

Раскодируем сообщение «01101100111100»

1. 01 = в

2. 101 = и

3. 100 = л

4. 1111 = к

5. 00 = а

Результат: «вилка».

**Задача 2**

Построить таблицу кодов символов методами Шеннона-Фано и Хаффмана.

Пусть A{a1, a2, а3, a4, a5, a6, a7}, P=(0,20; 0,20; 0,19; 0,12; 0,11; 0,09; 0,09).

Решение:

Составим таблицу вероятностей появления символов:

|  |  |
| --- | --- |
| Символ | Вероятность |
| а1 | 0,2 |
| а2 | 0,2 |
| а3 | 0,19 |
| а4 | 0,12 |
| а5 | 0,11 |
| а6 | 0,09 |
| а7 | 0,09 |

Воспользуемся методом Шеннона-Фано:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Символ | Вероятности | Символы кода | | | |
| 1 | 2 | 3 | Код |
| а1 | 0,2 | 0 | 0 |  | 00 |
| а2 | 0,2 | 1 | 0 | 010 |
| а3 | 0,19 | 1 | 011 |
| а4 | 0,12 | 1 | 0 | 0 | 100 |
| а5 | 0,11 | 1 | 101 |
| а6 | 0,09 | 1 | 0 | 110 |
| а7 | 0,09 | 1 | 111 |

Воспользуемся методом Хаффмана:

Таблицы кодов:

Метод Фано:

|  |  |
| --- | --- |
| Символ | Код |
| а1 | 00 |
| а2 | 010 |
| а3 | 011 |
| а4 | 100 |
| а5 | 101 |
| а6 | 110 |
| а7 | 111 |

Метод Хаффмана:

|  |  |
| --- | --- |
| Символ | Код |
| а1 | 10 |
| а2 | 11 |
| а3 | 000 |
| а4 | 010 |
| а5 | 011 |
| а6 | 0010 |
| а7 | 0011 |

**Задача 3**

Построить оптимальный неравномерный код методом Хаффмана.

Данные: Pa1=0,22, Pa2=0,58, Pa3=0,01, Pa4=0,03, Pa5=0,16.

Решение:

Воспользуемся методом Хаффмана:



Таблица кодов:

|  |  |
| --- | --- |
| Символ | Код |
| а1 | 10 |
| а2 | 0 |
| а3 | 1111 |
| а4 | 1110 |
| а5 | 110 |

**Задача 4**

Построить оптимальный код по методам Шеннона-Фано и Хаффмана.

Определить энтропию сообщения, сравнить среднюю длину кодового слова, построенного разными методами.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X1 | Х12 | ХЗ | X4 | X5 | X6 | X7 | X8 | X9 |
| 0,35 | 0,15 | 0,13 | 0,09 | 0,09 | 0,08 | 0,05 | 0,04 | 0,02 |

Решение:

Воспользуемся методом Шеннона-Фано:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Символ | Вероятности | Символы кода | | | | | |
| 1 | 2 | 3 | 4 | 5 | Код |
| X1 | 0,35 | 0 | 0 |  |  |  | 00 |
| X2 | 0,15 | 1 | 01 |
| X3 | 0,13 | 1 | 0 | 0 | 100 |
| X4 | 0,09 | 1 | 101 |
| X5 | 0,09 | 1 | 0 | 0 | 1100 |
| X6 | 0,08 | 1 | 1101 |
| X7 | 0,05 | 1 | 0 | 1110 |
| X8 | 0,04 | 1 | 0 | 11110 |
| X9 | 0,02 | 1 | 11111 |

Энтропия сообщения H = = 2,75483

Средняя длина слова nср = = 2,84

Воспользуемся методом Хаффмана:

Энтропия сообщения H = = 2,75483

Средняя длина слова nср = = 2,82

Результат: средняя длина слова, закодированного методом Хаффмана, оказалась меньше средней длины слова, закодированного методом Фано, на 2,84-2,82 = 0,02.

Вопросы для защиты лабораторной работы

1. Что понимают под кодированием сообщения?

Под кодированием сообщения понимается процесс преобразования символов или символьных последовательностей в определенные коды, которые позволяют представить информацию в компактной форме для передачи, хранения или обработки данных.

1. Какие коды называются равномерными?

Равномерные коды - это коды, в которых каждому символу или символьной последовательности присваивается одинаковое количество битов. В таких кодах длина кодовых слов одинакова для всех символов, что позволяет достичь равномерного распределения информации.

1. Как строится код Шеннона-Фано?

Код Шеннона-Фано строится путем разделения символов на две группы на основе их вероятностей. Рекурсивно процесс разделения повторяется для каждой группы до достижения достаточно точного разделения. Затем каждая группа получает битовое значение, присваиваемое префиксу кода.

1. Как определяется число элементарных сигналов, приходящихся на одну букву сообщения?

Число элементарных сигналов, приходящихся на одну букву сообщения, определяется как двоичный логарифм обратной вероятности данной буквы. Формула для определения числа элементарных сигналов: log(1/p), где p - вероятность данной буквы.

1. Сформулировать основную теорему о кодировании.

Основная теорема о кодировании утверждает, что для любого источника с информационной энтропией H, существует код, который может быть построен средствами метода Хаффмана или метода Шеннона-Фано, и его средняя длина кодового слова будет равна или меньше H.

1. Что называется декодированием сообщения?

Декодирование сообщения - это процесс преобразования закодированной информации обратно в исходный вид или формат, чтобы оно стало понятным и удобным для чтения или использования.

1. Что называется блочным кодированием?

Блочное кодирование - это метод кодирования, в котором информация разделяется на блоки фиксированной длины, а затем каждый блок обрабатывается независимо от остальных блоков. Такой подход обеспечивает более эффективное использование ресурсов и устойчивость передачи данных.

1. Объяснить принцип построения кода Хаффмана.

Принцип построения кода Хаффмана основан на использовании переменной длины кодовых слов для символов, при котором частота появления символа в источнике информации определяет его код. Часто встречающимся символам присваиваются более короткие кодовые слова, а реже встречающимся символам - более длинные кодовые слова. Это позволяет достичь эффективного сжатия информации.

1. Назначение и цели эффективного кодирования.

Эффективное кодирование имеет следующие назначение и цели:

* Уменьшение объема передаваемых данных или занимаемого места для хранения информации.
* Повышение скорости передачи данных и снижение нагрузки на сеть.
* Обеспечение защиты информации путем использования кодов, которые трудно восстановить без специального ключа или знания правил декодирования.
* Оптимизация использования ресурсов, таких как пропускная способность канала связи или объем памяти, путем эффективного представления информации с помощью минимального числа битов или символов.

**Заключение**

Исследовал методы Шеннона-Фано и Хаффмана для эффективного кодирования информации и применил их на практике.

Область применения результатов, полученных в ходе лабораторной работы, может быть широкой и охватывать различные области, где эффективное кодирование информации является важным аспектом. На основе полученных результатов можно разработать эффективные методы сжатия информации, что пригодится, например, в телекоммуникации для передачи данных лучшего качества при использовании канала с ограниченной пропускной способностью.

Знать о том, как производится эффективное кодирование по методам Шеннона-Фано и Хаффмана нужно хотя бы для того, чтобы иметь представление о том, как производится сжатие информации. Умение эффективно кодировать информацию пригодится в тех случаях, когда нужно сократить объем передаваемых данных с целью экономии пропускной способности каналов связи и ресурсов хранения.