|  |  |
| --- | --- |
| 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  Митин Юрий Олегович  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.

Решение:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| а | в | л | и | е | с | к |
| 0,3 | 0,2 | 0,15 | 0,1 | 0,1 | 0,08 | 0,07 |

Таблица 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 |

|  |  |
| --- | --- |
| Символ | Код |
| а | 00 |
| в | 01 |
| л | 100 |
| и | 101 |
| е | 110 |
| с | 1110 |
| к | 1111 |

При кодировании «вилка» получается: 01101100111100

Раскодируем сообщение «01101100111100»

01 = в

101 = и

100 = л

1111 = к

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:

Вероятности:

|  |  |
| --- | --- |
| 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 |

Таблица кодов методом Фано:

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

Воспользуемся методом Хаффмана:



Таблица кодов методом 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:

Вероятности:

|  |  |
| --- | --- |
| Символ | Вероятность |
| а1 | 0,22 |
| а2 | 0,58 |
| а3 | 0,01 |
| а4 | 0,03 |
| а5 | 0,16 |

Воспользуемся методом Хаффмана:

Складываем синее, получаем оранжевое



Таблица кодов:

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

Оптимальный неравномерный код построен.

**Задача 4**

Построить оптимальный код по методам Шеннона-Фано и Хаффмана.

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

Решение:

Вероятности:

|  |  |
| --- | --- |
| Символ | Вероятность |
| X1 | 0,35 |
| X2 | 0,15 |
| X3 | 0,13 |
| X4 | 0,09 |
| X5 | 0,09 |
| X6 | 0,08 |
| X7 | 0,05 |
| X8 | 0,04 |
| X9 | 0,02 |

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: Кодирование Фано эффективнее закодировало сообщение, чем кодирование Хаффмана, на 0,02.

Вопросы для защиты лабораторной работы

1. **Что понимают под кодированием сообщения?**

Кодирование сообщения - это процесс преобразования информации в определенный набор символов или кодов, который может быть передан или хранен. При кодировании каждый символ или элемент сообщения заменяется соответствующим кодом или представлением. Кодирование позволяет эффективно представить информацию, уменьшить объем передаваемых данных или улучшить их защиту.

Примеры кодирования сообщений включают использование двоичного кодирования, где каждый символ представляется последовательностью из двух символов (0 и 1), или использование кодовых таблиц, где каждому символу сопоставляется уникальный код. Кодирование также может включать сжатие данных, где информация сокращается до более короткой формы с помощью определенных алгоритмов.

Цель кодирования сообщения состоит в том, чтобы представить информацию эффективно, минимизировать объем передаваемых данных и обеспечить правильную интерпретацию сообщения при его декодировании на стороне получателя.

1. **Какие коды называются равномерными?**

Равномерные коды, также известные как блочные коды, это коды, в которых каждое кодовое слово имеет одинаковую длину. В равномерных кодах каждый символ или элемент исходного сообщения кодируется с помощью одинакового числа бит.

Преимущество равномерных кодов заключается в их простоте и эффективности в реализации, так как декодирование равномерных кодов требует фиксированного количества бит на символ без необходимости перехода через дополнительные проверки или разбор специальных инструкций. Это делает их особенно полезными в контексте некоторых приложений, где требуется быстрое декодирование, таких как аудио- и видеокодирование.

Однако, равномерные коды не всегда являются наилучшим выбором для сжатия данных, особенно когда вероятности появления символов в исходном сообщении неодинаковы. В таких случаях применение других типов кодирования, таких как коды Хаффмана или коды Фано, может привести к более эффективному сжатию данных.

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 (теорема Шеннона) утверждает, что для источника сообщений с энтропией H, существует код, который может быть построен средней длиной кодовых слов, не превышающей H бит на символ, при условии, что размер кодирующего алфавита достаточно большой.

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, при котором исходное сообщение разбивается на блоки фиксированной длины, а каждый блок кодируется отдельно и независимо от остальных блоков. Кодирование и декодирование происходят блок за блоком, без взаимного влияния между блоками.

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 несколько назначений и целей:

1. Сжатие данных: Одной из основных целей эффективного кодирования является сжатие данных. Цель состоит в том, чтобы представить информацию с наименьшим возможным объемом данных, минимизируя количество бит, необходимых для представления каждого символа или элемента. Сжатие данных позволяет экономить место на диске или в памяти, ускоряет передачу данных и уменьшает потребление пропускной способности сети.

2. Повышение скорости передачи данных: Эффективное кодирование может улучшить скорость передачи данных, особенно при использовании сжатия данных. Уменьшение объема данных сокращает время передачи и улучшает производительность системы.

3. Улучшение защиты данных: Эффективное кодирование может быть использовано для обеспечения защиты данных путем добавления дополнительных проверочных сумм, контрольных битов или кодов ошибок. Такие коды помогают обнаруживать и исправлять ошибки, возникающие при передаче или хранении данных.

4. Оптимизация использования ресурсов: Эффективное кодирование позволяет оптимизировать использование ресурсов, таких как пропускная способность сети, память и вычислительная мощность. Уменьшение объема данных снижает нагрузку на ресурсы и позволяет эффективнее использовать доступные ресурсы.

5. Сохранение качества информации: При кодировании данных важно сохранить качество информации. Цель состоит в том, чтобы минимизировать потери информации при сжатии или кодировании данных, особенно в случае сжатия потерь или аудио/видеокодирования.

Общая цель эффективного кодирования заключается в оптимизации представления, передачи и хранения информации, уменьшении объема данных и повышении производительности системы при сохранении качества информации.

**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.