|  |  |
| --- | --- |
| 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: Бутковский Данила,  студент гр. 1ПИб-02-1оп-22    подпись, дата |
| Череповец, 2024 | |

**Реферат**

Предметом исследования являются Коды Шенно-Фано и Хаффмана для кодирования сообщений.

Цель: научиться эффективному кодированию по методу Шенно-Фано и Хаффмана

Учитывая статистические свойства источника сообщения, можно минимизировать среднее число двоичных символов, требующихся для выражения одной буквы сообщения, что при отсутствии шума позволяет уменьшить время передачи или объем запоминающего устройства.

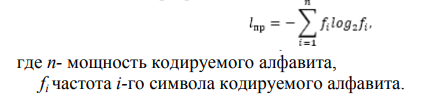
Такое эффективное кодирование базируется на основной теореме Шеннона для каналов без шума. Шеннон доказал, что сообщения, составленные из букв некоторого алфавита, можно закодировать так, что среднее число двоичных символов на букву будет сколь угодно близко к энтропии источника этих сообщений, но не меньше этой величины.

Теорема не указывает конкретного способа кодирования, но из нее следует, что при выборе каждого символа кодовой комбинации необходимо стараться, чтобы он нес максимальную информацию. Следовательно, каждый символ должен принимать значения 0 или 1 по возможности с равными вероятностями, и каждый выбор должен быть независим от значений предыдуіцих символов.

*Эффективное* или экономичное кодирование используется для уменьшения объемов информации на носителе-сигнале таким образом, чтобы устранить избыточность.

Для кодирования символов исходного алфавита используются двоичные коды переменной длины: чем больше частота символа, тем короче его код. Эффективность кода определяется средним числом двоичных разрядов для кодирования одного символа.

При эффективном кодировании существует предел сжатия, ниже которого не «спускается» ни один метод эффективного кодирования — иначе будет потеряна информация. Этот параметр определяется предельным значением двоичных разрядов возможного эффективного кода

****

Для случая отсутствия статистической взаимосвязи между буквами конструктивные методы построения эффективных кодов были даны впервые Шенноном и Фано. Их методики существенно не отличаются, и поэтому соответствующий код получил название кода Шеннона - Фано.

**Введение**

**Метод Шеннона-Фано**

Упорядоченный в порядке невозрастания вероятностей список букв делится на две последовательные части так, чтобы суммы вероятностей входящих в них букв как можно меньше отличались друг от друга. Буквам из первой части приписываем символ 0, а буквам из второй части — символ 1. Далее точно так же поступаем с каждой из полученных частей, если она содержит хотя бы две буквы. Построенный код является префиксным, и ему соответствует насыщенное кодовое дерево.

В алгоритме Фано кодовое дерево строится от корня, а в алгоритме Хаффмана начиная с листьев. Это отличие позволяет в алгоритме Хаффмана полнее использовать специфику данного распределения вероятностей и строить оптимальный код. Алгоритм Фано строит код, близкий к оптимальному.

Метод Шеннона-Фано соответствует требованию оптимального кодирования. Алгоритм построения кода Шеннона-Фано состоит в том, что кодируемые символы (буквы) разделяются на две равновероятные подгруппы: для символов 1-й подгруппы на втором месте ставится 0, а для 2-й подгруппы — 1 и т.д.

**Методика Хаффмана**

Методика Шеннона-Фано не всегда приводит к однозначному построению кода. Ведь при разбиении на подгруппы можно сделать большей по вероятности как верхнюю, так и нижнюю подгруппу.

От указанного недостатка свободна методика Хаффмана. Она гарантирует однозначное построение кода с наименьшим для данного распределения вероятностей средним числом символов на букву.

Для двоичного кода методика сводится к следующему. Буквы алфавита сообщений выписываются в основной столбец в порядке убывания вероятностей. Две последние буквы объединяются в одну вспомогательную букву, которой приписывается суммарная вероятность. Вероятности букв, не участвовавших в объединении, и полученная суммарная вероятность снова располагаются в порядке убывания вероятностей в дополнительном столбце, а две последние объединяются. Процесс продолжается до тех пор, пока не получим единственную вспомогательную букву с вероятностью, равной единице.

Для составления кодовой комбинации, соответствующей данному сообщению, необходимо проследить путь перехода сообщения по строкам и столбцам таблицы.

Эффективность рассмотренных алгоритмов достигается благодаря присвоению более коротких кодовых комбинаций (кодовых символов) символам источника сообщений, вероятность которых более высока, и более длинных кодовых комбинаций - символам источника сообщений с малой вероятностью. Это ведет к различиям в длине кодовых символов и, как следствие, к трудностям при их декодировании. Для разделения отдельных кодовых символов можно использовать специальный разделительный элемент, но при этом существенно снижается эффективность кода, т.к. средняя длина кодового символа фактически увеличивается на один элемент символа кода.

Целесообразнее обеспечить декодирование без введения дополнительных элементов символов. Этого можно добиться, если в эффективном коде ни одна кодовая комбинация не будет совпадать с началом более длинной кодовой комбинации. Коды, удовлетворяющие этому условию, называют префиксными кодами (префиксом или началом называют первый элемент в кодовом символе, а последний элемент — окончанием или постфиксом).

Коды, построенные по алгоритмам Шеннона-Фано или Хаффмана, являются префиксными.

**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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Буква | Вероятности | Символы кода | | | | 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 |

Сообщению “вилка” соответствует выходная последовательность кодов 01101100111100.

**Задача 2**

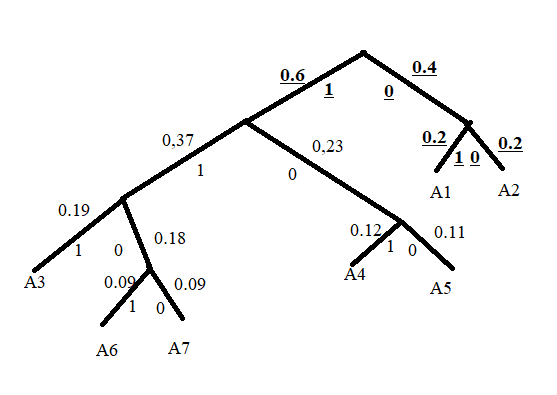
Построить таблицу кодов символов методами Шеннона-Фано и Хаффмана.  
  
Пусть A{a1, a2, а3, a4, a5, a6, a7}, P=(0,20; 0,20; 0,19; 0,12; 0,11; 0,09; 0,09).

**Метод Шеннона-фано**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Буква | Вероятности | Символы кода | | | | Код |
| А1 | 0.2 | 0 | 0 |  |  | 00 |
| A2 | 0.2 | 1 |  |  | 01 |
| A3 | 0.19 | 1 | 0 | 0 |  | 100 |
| A4 | 0.12 | 1 |  | 101 |
| A6 | 0.11 | 1 | 0 |  | 110 |
| A6 | 0.09 | 1 | 0 | 1110 |
| A7 | 0.09 | 1 | 1111 |

**Метод Хаффмана**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| A1 | 0.20 | 0.20 | 0.23 | 0.37 | 0.40 | 0.60 | 1 |
| A2 | 0.20 | 0.20 | 0.20 | 0.23 | 0.37 | 0.40 |  |
| A3 | 0.19 | 0.19 | 0.20 | 0.20 | 0.23 |  |  |
| A4 | 0.12 | 0.18 | 0.19 | 0.20 |  |  |  |
| A5 | 0.11 | 0.12 | 0.18 |  |  |  |  |
| A6 | 0.09 | 0.11 |  |  |  |  |  |
| A7 | 0.09 |  |  |  |  |  |  |

****

**Задача 3**

Построить оптимальный неравномерный код методом Хаффмана.

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

**Метод Хаффмана**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| pA2 | 0.58 | 0.58 | 0.58 | 0.58 | 1 |
| pA1 | 0.22 | 0.22 | 0.22 | 0.24 |  |
| pA5 | 0.16 | 0.16 | 0.2 |  |  |
| pA4 | 0.03 | 0.4 |  |  |  |
| pA3 | 0.01 |  |  |  |  |

**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 |

**Метод Шеннона-Фано**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Буква | Вероятности | Символы кода | | | | | Код |
| А1 | 0.35 | 0 | 0 |  |  |  | 00 |
| A2 | 0.15 | 1 |  |  |  | 01 |
| A3 | 0.13 | 1 | 0 | 0 |  |  | 100 |
| A4 | 0.09 | 1 |  |  | 101 |
| A5 | 0.08 | 1 | 0 | 0 |  | 1100 |
| A6 | 0.08 | 1 |  | 1101 |
| A7 | 0.05 | 1 | 0 |  | 1110 |
| A8 | 0.04 | 1 | 0 | 11110 |
| A9 | 0.02 | 1 | 11111 |

Энтропия = 2.73бит

Средняя длина символа =3.55

**Метод Хаффмана**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| А1 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.37 | 0.63 | 1 |
| А2 | 0.15 | 0.15 | 0.15 | 0.17 | 0.2 | 0.28 | 0.35 | 0.37 |  |
| А3 | 0.13 | 0.13 | 0.13 | 0.15 | 0.17 | 0.2 | 0.28 |  |  |
| А4 | 0.09 | 0.09 | 0.11 | 0.13 | 0.15 | 0.17 |  |  |  |
| А5 | 0.09 | 0.09 | 0.09 | 0.11 | 0.13 |  |  |  |  |
| А6 | 0.08 | 0.08 | 0.09 | 0.09 |  |  |  |  |  |
| А7 | 0.05 | 0.06 | 0.08 |  |  |  |  |  |  |
| А8 | 0.04 | 0.05 |  |  |  |  |  |  |  |
| А9 | 0.02 |  |  |  |  |  |  |  |  |

Энтропия =2.75бит

Средняя длина символа =2.82

Вывод: Код Хаффмана более короткий.

**Вопросы для защиты лабораторной работы**

1. **Что понимают под кодированием сообщения?** Кодирование — это перевод информации в удобную для передачи, обработки или хранения форму с помощью некоторого кода.
2. **Какие коды называются равномерными?**

Код называется равномерным (или кодом постоянной длины), если все его кодовые слова содержат одинаковое число букв (одинаковую длину слов).

**3) Как строится код Шеннона-Фано?**

Код Шеннона — Фано строится с помощью дерева. Построение этого дерева начинается от корня. Всё множество кодируемых элементов соответствует корню дерева (вершине первого уровня). Оно разбивается на два подмножества с примерно одинаковыми суммарными вероятностями. Эти подмножества соответствуют двум вершинам второго уровня, которые соединяются с корнем.

**4) Как определяется число элементарных сигналов, приходящихся на одну букву сообщения?**

Сумма длин каждого символа поделить на кол-во символов

**5) Сформулировать основную теорему о кодировании.**

Основная теорема о кодировании при отсутствии помех К. Шеннона утверждает: минимальное среднее число кодовых символов, приходящихся на один символ сообщения, можно сделать сколь угодно близким к H(A)/log L, где Н(А) – энтропия источника, состоящего из символов сообщения, L – число кодовых символов (объем алфавита).

**6) Что называется декодированием сообщения?**

Декодирование — это восстановление информационного сообщения из последовательности кодов. Сообщения, записанные с помощью равномерного кода, всегда декодируются однозначно.

**7) Что называется блочным кодированием?**

Блочное кодирование является особым случаем кодирования контроля ошибок. Методы блочного кодирования сопоставляют постоянное число символов сообщения к постоянному числу кодовых знаков. Кодер блока обрабатывает каждый блок данных независимо и является устройством без памяти. Communications Toolbox содержит возможности блочного кодирования путем обеспечения блоков Simulink, Системных объектов и функций MATLAB.

**8) Объяснить принцип построения кода Хаффмана.**

Алгоритм Хаффмана является алгоритмом префиксного кодирования: символы произвольного алфавита кодируются двоичными последовательностями различной длины таким образом, что ни один более короткий код не является префиксом более длинного. При этом, для достижения оптимального кодирования, наиболее часто встречающиеся символы кодируются наиболее короткой последовательностью, а наиболее редко встречающиеся – наиболее длинной.

**9) Назначение и цели эффективного кодирования.**

Эффективное кодирование – это такое кодирование, при котором устраняется избыточность. Избыточность возникает в том случае, когда энтропия 𝐻 (𝐴) исходного сообщения меньше максимально возможной величины (𝐻 (𝐴) < 𝐻𝑚𝑎𝑥). В результате такого кодирования скорость передачи информации увеличивается до максимально-возможной величины.