|  |  |
| --- | --- |
| 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-02-01оп-22 Маслов Владислав Андреевич    подпись, дата |
| Череповец, 2023 | |

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.

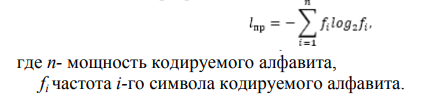
Такое эффективное кодирование базируется на основной теореме Шеннона для каналов без шума. Шеннон доказал, что сообщения, составленные из букв некоторого алфавита, можно закодировать так, что среднее число двоичных символов на букву будет сколь угодно близко к энтропии источника этих сообщений, но не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

Эффективноеили 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

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 | P | 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 |

Табл.1.

Сообщению “вилка” соответствует выходная последовательность кодов 01101100111100.

1. 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 | P | 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 |

Табл.2.

Метод Хаффмана:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Символ | Код | P | | | | | | |
| A1 | 10 | 0.2 | 0.2 | 0.23 | 0.37 | 0.4 | 0.6 | 1 |
| A2 | 11 | 0.2 | 0.2 | 0.2 | 0.23 | 0.37 | 0.4 |
| A3 | 000 | 0.19 | 0.19 | 0.2 | 0.2 | 0.23 |
| A4 | 010 | 0.12 | 0.18 | 0.19 | 0.2 |
| A5 | 011 | 0.11 | 0.12 | 0.18 |
| A6 | 0010 | 0.09 | 0.11 |
| A7 | 0011 | 0.09 |

Табл.3.

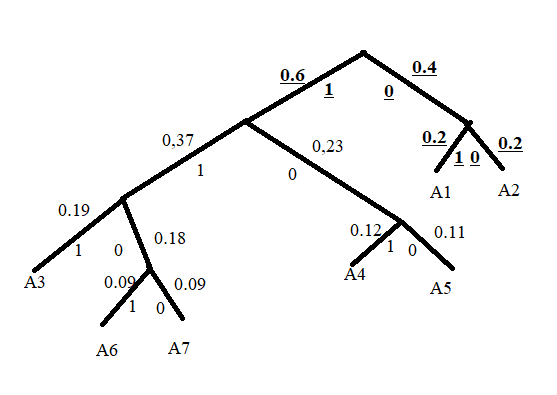
****

Рис.1.

1. Построить оптимальный неравномерный код методом Хаффмана.

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

Метод Хаффмана:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Символ | Код | P | | | | |
| A2 | 0 | 0.58 | 0.58 | 0.58 | 0.58 | 1 |
| A1 | 10 | 0.22 | 0.22 | 0.22 | 0.42 |
| A5 | 110 | 0.16 | 0.16 | 0.2 |
| A4 | 1110 | 0.03 | 0.4 |
| A3 | 1111 | 0.01 |

Табл.4.

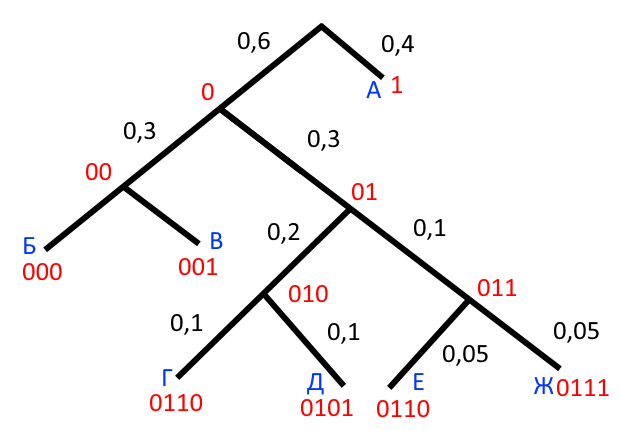


Рис.2.

Цена кодирования: l=1\*0,4+3\*0,2+3\*0,1+4\*(0,1\*2+0,05\*2) = 2,5

1. 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 |

Табл.5.

Метод Шеннона-Фано:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Буква | P | 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 |

Табл.6.

Энтропия сообщения H = = 2,75483

Средняя длина слова nср = = 2,84

Метод Хаффмана:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Символ | Код | P | | | | | | | | |
| А1 | 00 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.37 | 0.63 | 1 |
| А2 | 010 | 0.15 | 0.15 | 0.15 | 0.17 | 0.2 | 0.28 | 0.35 | 0.37 |  |
| А3 | 011 | 0.13 | 0.13 | 0.13 | 0.15 | 0.17 | 0.2 | 0.28 |  |  |
| А4 | 101 | 0.09 | 0.09 | 0.11 | 0.13 | 0.15 | 0.17 |  |  |  |
| А5 | 110 | 0.09 | 0.09 | 0.09 | 0.11 | 0.13 |  |  |  |  |
| А6 | 111 | 0.08 | 0.08 | 0.09 | 0.09 |  |  |  |  |  |
| А7 | 1001 | 0.05 | 0.06 | 0.08 |  |  |  |  |  |  |
| А8 | 10000 | 0.04 | 0.05 |  |  |  |  |  |  |  |
| А9 | 10001 | 0.02 |  |  |  |  |  |  |  |  |

Табл.7.

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

В ходе лабораторной работы были исследованы методы Шеннона-Фано и Хаффмана для эффективного кодирования информации и применены 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.