|  |  |
| --- | --- |
| Череповецкий государственный университет  Кафедра «Математического и программного обеспечения ЭВМ» | |
| ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ  по дисциплине «Теория информации»  КОДЫ ФАНО И ХАФМАНА | |
|  | 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 Фано и Хаффмана для кодирования сообщений.

Цель работы – научиться кодировать и декодировать по методу Фано и Хаффмана.

В ходе работы проводились теоретические изучения методов кодирования Фано и Хаффмана, а также работа с таблицами и вычисление вероятностей сообщений для работы с методом Хаффмана.

В результате аналитических исследований были найдено кодирование по методу Фано двухбуквенных комбинаций, кодирование по Хаффману трехбуквенных комбинаций, построено тринарное дерево Хаффмана для кодирования трехбуквенных комбинаций.

Коды Фано и Хаффмана используются для кодирования приведенной последовательности применяется неравномерный двоичный код, при помощи которого можно осуществить однозначное декодирование.

**Введение**

Коды Фано и Хаффмана являются оптимальными и префиксными. При построении искомых кодов будем применять как традиционный табличный способ кодирования, так и использовать "кодовые деревья".

При кодировании по Фано все сообщения записываются в таблицу по степени убывания вероятности и разбиваются на две группы примерно (насколько это возможно) равной вероятности. Соответственно этой процедуре из корня кодового дерева исходят два ребра, которым в качестве весов присваиваются полученные вероятности. Двум образовавшимся вершинам приписывают кодовые символы 0 и 1. Затем каждая из групп вероятностей вновь делится на две подгруппы примерно равной вероятности. В соответствии с этим из каждой вершины 0 и 1 исходят по два ребра с весами, равными вероятностям подгрупп, а вновь образованным вершинам приписывают символы 00 и 01, 10 и 11. В результате многократного повторения процедуры разделения вероятностей и образования вершин приходим к ситуации, когда в качестве веса, приписанного ребру бинарного дерева, выступает вероятность одного из данных сообщений. В этом случае вновь образованная вершина оказывается листом дерева, т.к. процесс деления вероятностей для нее завершен. Задача кодирования считается решенной, когда на всех ветвях кодового бинарного дерева образуются листья.

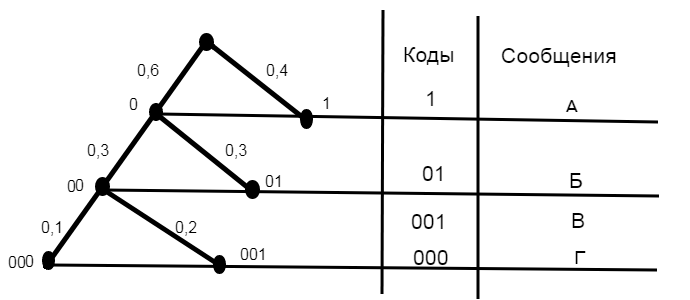
Для того, чтобы закодировать сообщения по Хаффману, предварительно преобразуется таблица, задающая вероятности сообщений. Исходные данные записываются в столбец, две последние (наименьшие) вероятности в котором складываются, а полученная сумма становится новым элементом таблицы, занимающим соответствующее место в списке убывающих по величине вероятностей. Эта процедура продолжается до тех пор, пока в столбце не останутся всего два элемента.

**Ход работы**

**Задание 1.**

Проведите кодирование по методу Фано алфавита из четырех букв, вероятности которых равны 0,4; 0,3; 0,2 и 0,1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Сообщение | А | Б | В | Г |
| P | 0,4 | 0,3 | 0,2 | 0,1 |



В итоге получится следующий алфавит:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Сообщение | А | Б | В | Г |
| P | 1 | 01 | 001 | 000 |

**Задание 2.**

Алфавит содержит 7 букв, которые встречаются с вероятностями 0,4; 0,2; 0,1; 0,1; 0,1; 0,05; 0,05. Осуществите кодирование по методу Фано.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Сообщение | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| P | 0,4 | 0,2 | 0,1 | 0,1 | 0,1 | 0,05 | 0,05 |



**Задание 3.**

Алфавит состоит из двух букв, $A$ и $Б$, встречающихся с вероятностями $P(A)$ = 0,8 и $P(Б)$ = 0,2. Примените метод Фано к кодированию всевозможных двухбуквенных и трехбуквенных комбинаций.

Таблица для двухбуквенных комбинаций:

|  |  |  |  |
| --- | --- | --- | --- |
| Комбинации | Р | Разбиение | Код |
| АА АБ БА ББ | 0.64  0.16  0.16  0.01 |  | 0  10  110  111 |

Таблица для трёхбуквенных комбинаций:

|  |  |  |  |
| --- | --- | --- | --- |
| Комбинации | Вероятность | Разбиение | Код |
| ААА ААБ АБА БАА АББ БАБ ББА БББ | 0.512 0.128 0.128 0.128 0.032 0.032 0.032 0.008 |  | 0 100 101 110 11100 11101 11110 11111 |

**Задание 4.**

Проведите кодирование по методу Хаффмана трехбуквенных слов из предыдущей задачи.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Сообщ. | P | P1 | P2 | P3 | P4 | P5 | P6 |
| ААА | 0,512 | 0,512 | 0,512 | 0,512 | 0,512 | 0,512 | 0,512 |
| БАА | 0,128 | 0,128 | 0,128 | 0,128 | 0,232 (11) | 0,256 (10) | 0,488 (1) |
| АБА | 0,128 | 0,128 | 0,128 | 0,128 | 0,128 (100) | 0,232 (11) |  |
| ААБ | 0,128 | 0,128 | 0,128 | 0,128 (110) | 0,128 (101) |  |  |
| ББА | 0,128 | 0,04 (1111) | 0,064 (1110) | 0,104 (111) |  |  |  |
| БАБ | 0,128 | 0,032 (11100) | 0,04 (1111) |  |  |  |  |
| АББ | 0,032 (11110) | 0,032 (11101) |  |  |  |  |  |
| БББ | 0,008 (11111) |  |  |  |  |  |  |

Полученные коды:

|  |  |
| --- | --- |
| Сообщение | Код |
| ААА ААБ АБА БАА АББ БАБ ББА БББ | 0  100  101  110  11100  11101  11110  11111 |

**Задание 5.**

Проведите кодирование 7 букв из задачи 3 по методу Хаффмана.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Сооб. | P | P1 | P2 | P3 | P4 | P5 | Коды |
| 1 | 0,4 | 0,4 | 0,4 | 0,4 | 0,4 | 0,6 (0) | 1 |
| 2 | 0,2 | 0,2 | 0,2 | 0,2 | 0,4 (00) | 0,4 (1) | 01 |
| 3 | 0,1 | 0,1 | 0,2 (000) | 0,2 (000) | 0,2 (01) |  | 0010 |
| 4 | 0,1 | 0,1 | 0,1 (0010) | 0,2 (001) |  |  | 0011 |
| 5 | 0,1 | 0,1 (0000) | 0,1 (0011) |  |  |  | 0000 |
| 6 | 0,05 (00010) | 0,1 (0001) |  |  |  |  | 00010 |
| 7 | 0,05 (00011) |  |  |  |  |  | 00011 |

**Задание 6.**

Проведите кодирование по методам Фано и Хаффмана пяти букв, равновероятно встречающихся.

Если мы имеем 5 букв и их их вероятности равны, то вероятность одной буквы будет равна 1 : 5= **0.2**.

Метод Фано (табличный способ):



Метод Хафмана:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Сообщение | P | P1 | P2 | P3 | Кодовое обозначение |
| 1 | 0,2 | 0,4 (00) | 0,4 | 0,6 (0) | 11 |
| 2 | 0,2 | 0,2 | 0,4 (00) | 0,4 (1) | 10 |
| 3 | 0,2 | 0,2 (000) | 0,2 (01) |  | 01 |
| 4 | 0,2 (000) | 0,2 (001) |  |  | 001 |
| 5 | 0,2 (001) |  |  |  | 000 |

**Вопросы для самоконтроля**

**1. Как определяется среднее число элементарных сигналов, приходящихся на одну букву алфавита?**

**2. Префиксные коды**.  
Префиксные коды – это коды, которые удовлетворяют условию Фано.  
  
Для того, чтобы неравномерный код был однозначно и правильно декодируем, достаточно при его построении обеспечить выполнение следующее условие: никакое кодовое слово не должно быть началом никакого другого кодового слова.

**3. Сколько требуется двоичных знаков для записи кодированного сообщения?**

Количество двоичных знаков для записи кодированного сообщения зависит от выбранного способа кодирования и вероятностей появления каждой буквы.

**4. На чем основано построение кода Фано?**Построение кода Фано основано на принципе оптимального кодирования, при котором избыточность кода минимальна. Код Фано является префиксным кодом.

**5. Что такое сжатие алфавита?**Сжатие алфавита – метод Хаффмана.Для того, чтобы закодировать сообщения по Хаффману, предварительно преобразуется таблица, задающая вероятности сообщений. Исходные данные записываются в столбец, две последние (наименьшие) вероятности в котором складываются, а полученная сумма становится новым элементом таблицы, занимающим соответствующее место в списке убывающих по величине вероятностей. Эта процедура продолжается до тех пор, пока в столбце не останутся всего два элемента

**6. Какой код самый выгодный?**Код, средняя длина которого минимальна.

**7. Основная теорема о кодировании.**Неравномерные коды строятся на основе соблюдения условия Фано.

**8. Энтропия конкретных типов сообщений.**

**Заключение**

В ходе лабораторной работы были закодированы и декодированы различные сообщения по методу Фано и Хаффмана. Было произведено кодирование по методу Фано двухбуквенных комбинаций, кодирование по Хаффману трехбуквенных комбинаций, а также построено тринарное дерево Хаффмана для кодирования трехбуквенных комбинаций. Задачи, поставленные в ходе выполнения лабораторной работы, выполнены в полной мере.