|  |  |
| --- | --- |
| Череповецкий государственный университет  Кафедра «Математического и программного обеспечения ЭВМ» | |
| ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ  по дисциплине «Теория информации»  ИССЛЕДОВАНИЕ СВОЙСТВ ЭНТРОПИИ ДИСКРЕТНОГО ИСТОЧНИКА | |
|  | Принял:  преподаватель Е.Н. Руденко    подпись, дата  Выполнил: Овчинников Максим Владимирович  студент гр. 1ПИб-02-2оп-22    подпись, дата |
| Череповец, 2022 | |

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 найдены условия минимального и максимального значения энтропии. Численно были получены значения энтропии стационарных источников двух и нескольких видов сообщений. Произведено моделирование и исследована динамика изменения энтропии нестационарного источника.

Количественная оценка значения энтропии применяется при моделировании и кодировании источников.

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.

Цена кодирования (средняя длина кодового слова $l)$ является критерием степени оптимальности кодирования.

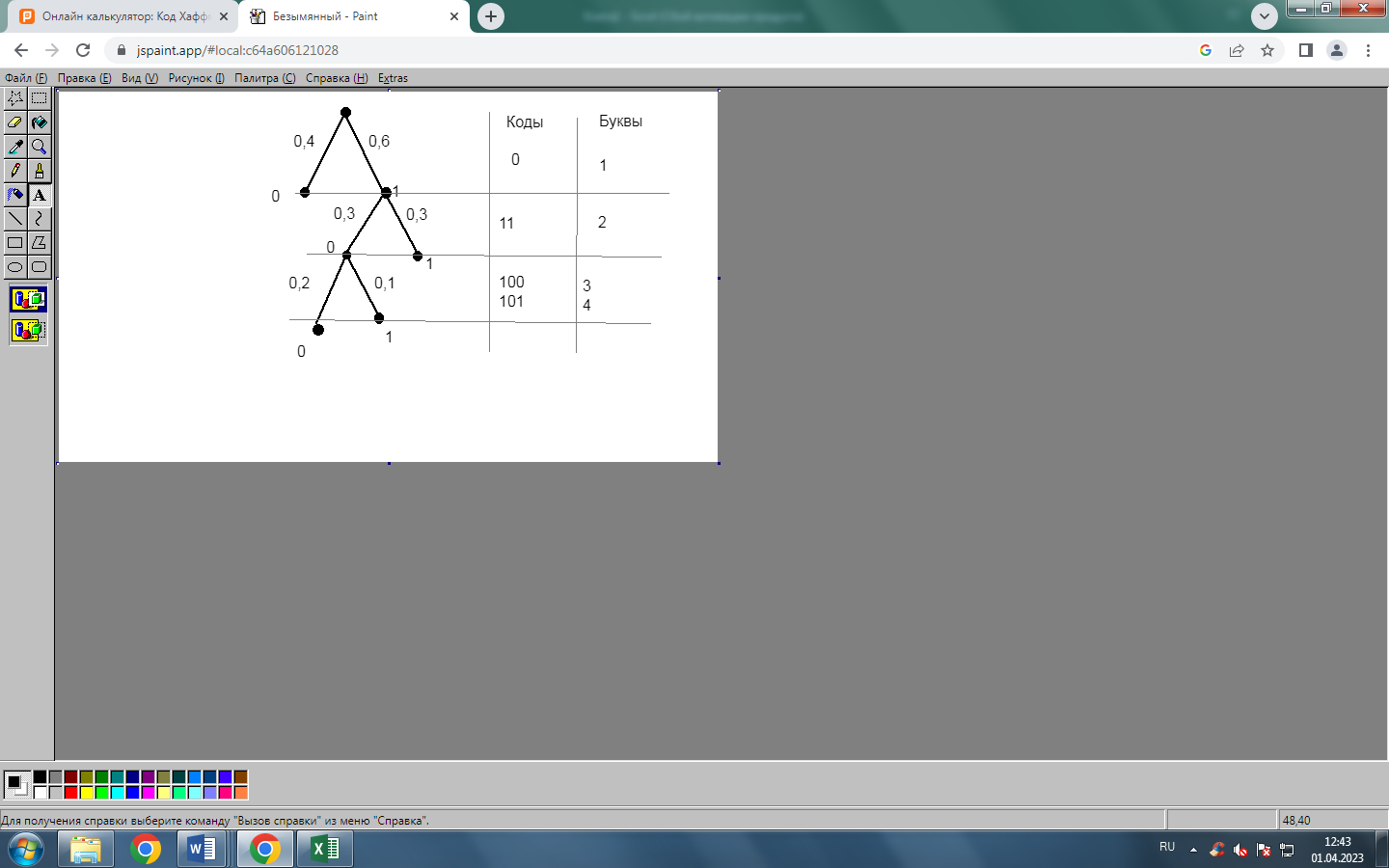
Ход работы

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Буква | 1 | 2 | 3 | 4 |
| Вероятность | 0,4 | 0,3 | 0,2 | 0,1 |

Таблица 1. Таблица данных задания 1.

Решение:



Кодовое дерево.

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.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Буква | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Вероятность | 0,4 | 0,2 | 0,1 | 0,1 | 0,1 | 0,05 | 0,05 |

Таблица 2. Таблица данных задания 2.

Решение:



Таблица 3. Табличный способ кодирования Фано

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.

|  |  |  |  |
| --- | --- | --- | --- |
| Комбинации букв | Вероятность | разбиение | код |
| АА АБ БА ББ | 0.64  0.16  0.16  0.01 |  | 0  10  110  111 |

Таблица 4. Таблица для двухбуквенных 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 |

Таблица 5. Таблица для трехбуквенных комбинаций.

4. Проведите кодирование по методу Хаффмана трехбуквенных слов из предыдущей задачи.

Решение:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Сообщение | p | p1 | p2 | p3 | p4 | p5 | p6 |
| ААА | 0,512 | 0,512 | 0,512 | 0,512 | 0,512 | 0,512 | 0,512 |
| ААБ | 0,128 | 0,128 | 0,128 | 0,128 | 0,232 | 0,256 | 0,488 |
| АБА | 0,128 | 0,128 | 0,128 | 0,128 | 0,128 | 0,232 |  |
| АББ | 0,128 | 0,128 | 0,128 | 0,128 | 0,128 |  |  |
| БАА | 0,032 | 0,04 | 0,064 | 0,104 |  |  |  |
| БАБ | 0,032 | 0,032 | 0,04 |  |  |  |  |
| ББА | 0,032 | 0,032 |  |  |  |  |  |
| БББ | 0,008 |  |  |  |  |  |  |

Таблица 6. Таблица кодирование по Хаффману трехбуквенных комбинаций

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

Таблица 7. Таблица Полученного кода методом Хаффмана

5. Проведите кодирование 7 букв из задачи 2 по методу Хаффмана.

Решение:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Буква | p | p1 | p2 | p3 | p4 | p5 | Код |
| 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 | 0 1 |
| 3 | 0,1 | 0,1 | 0,2 | 0,2|000 | 0,2|01 |  | 0 0 10 |
| 4 | 0,1 | 0,1 | 0,1|0010 | 0,2|001 |  |  | 0 0 11 |
| 5 | 0,1 | 0,1||0000 | 0,1|0011 |  |  |  | 0 0 0 0 |
| 6 | 0,05||00010 | 0,1||0001 |  |  |  |  | 0 0 0 10 |
| 7 | 0,05||00011 |  |  |  |  |  | 0 0 0 11 |

Таблица 8. Таблица кодирование по Хаффману трехбуквенных комбинаций

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

Решение:

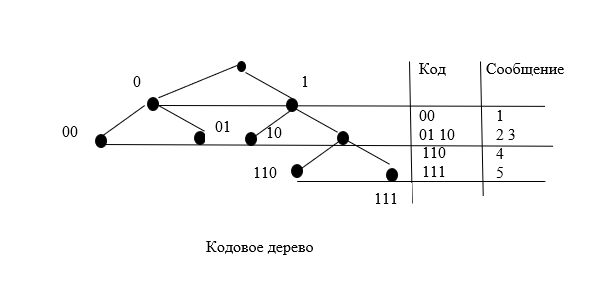
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Сообщение | 1 | 2 | 3 | 4 | 5 |
| Вероятность | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Буква | p | p1 | p2 | p3 | Код |
| 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 |  | 0 1 |
| 4 | 0,2|000 | 0,2||001 |  |  | 0 0 1 |
| 5 | 0,2|001 |  |  |  | 0 0 0 |

Таблица 9. Кодирование методом Хаффмана



Таблица 10. Кодирование методом Фано.



**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.

1. 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?

Самым выгодным будет код, средняя длина которого минимальна.

7. Основная теорема о кодировании.

Неравномерные коды строятся на основе соблюдения условия Фано.

8. Энтропия конкретных типов сообщений.

Заключение

Предметом исследования лабораторной работы являются коды Фано и Хаффмана. В ходе заданий произведено кодирование алфавита из 4, из 7 букв методом Фано. Также произведено кодирование двухбуквенных и трехбуквенных комбинаций с использованием методов Фано и Хаффмана. В ходе работы произведены теоретические исследования кодов Фано и Хаффмана, применялись как традиционный табличный способ кодирования, так и "кодовые деревья". Задачи, поставленные в ходе выполнения лабораторной работы, выполнены в полной мере.