|  |  |
| --- | --- |
| Череповецкий государственный университет  Кафедра «Математического и программного обеспечения ЭВМ» | |
| ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ  по дисциплине «Теория информации»  КОДЫ ФАНО И ХАФФМАНА | |
|  | Принял:  преподаватель Е.Н. Руденко    подпись, дата  Выполнил:  студент гр. 1ПИб-02-1оп-22  Остапенко Степан Ярославович  подпись, дата |
| Череповец, 2023 | |

**Реферат**

Предметом исследования являются методы кодирования сообщений по Фано и Хаффману.

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

В ходе работы происходило знакомство с методами кодирования по Фано и Хаффману, а также применение этих методов для решения некоторых задач.

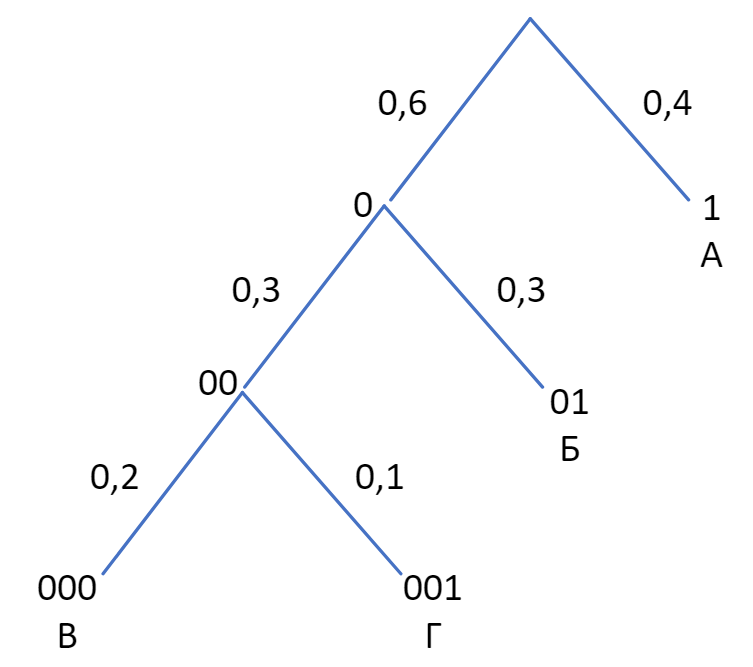
Коды Фано и Хаффмана являются оптимальными и префиксными. При построении искомых кодов будем применять как традиционный табличный способ кодирования, так и использовать "кодовые деревья".

**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.

Решение:

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



|  |  |
| --- | --- |
| Символ | Код |
| А | 1 |
| Б | 01 |
| В | 000 |
| Г | 001 |

Цена кодирования l = ΣLi\*Pi = 1\*0,4+2\*0,3+3\*0,2+3\*0,1 = 1,9

**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.

Решение:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| А | Б | В | Г | Д | Е | Ж |
| 0,4 | 0,2 | 0,1 | 0,1 | 0,1 | 0,05 | 0,05 |



Цена кодирования l = ΣLi\*Pi = 1\*0,4+3\*0,2+3\*0,1+4\*0,1+4\*0,1+4\*0,05+4\*0,05 = 2,5

**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.

Решение:

Рассмотрим двухбуквенные комбинации:



|  |  |
| --- | --- |
| Слово | Код |
| АА | 0 |
| АБ | 10 |
| БА | 110 |
| ББ | 111 |

Рассмотрим трёхбуквенные комбинации:



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

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

|  |  |
| --- | --- |
| Слово | P |
| ААА | 0,512 |
| ААБ | 0,128 |
| АБА | 0,128 |
| АББ | 0,032 |
| БАА | 0,128 |
| БАБ | 0,032 |
| ББА | 0,032 |
| БББ | 0,008 |



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

**Задание 5**. Проведите кодирование 7 букв из задачи 302 по методу Хаффмана.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| А | Б | В | Г | Д | Е | Ж |
| 0,4 | 0,2 | 0,1 | 0,1 | 0,1 | 0,05 | 0,05 |



|  |  |
| --- | --- |
| Символ | Код |
| А | 1 |
| Б | 01 |
| В | 0010 |
| Г | 0011 |
| Д | 0000 |
| Е | 00010 |
| Ж | 00011 |

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A | B | C | D | E |
| 0,2 | 0,2 | 0,2 | 0,2 | 0,2 |



Код по Фано:

|  |  |
| --- | --- |
| Буква | Код |
| A | 10 |
| B | 11 |
| C | 00 |
| D | 010 |
| E | 011 |

Код по Хаффману:

|  |  |
| --- | --- |
| Буква | Код |
| A | 01 |
| B | 000 |
| C | 001 |
| D | 10 |
| E | 11 |

**Задание 7**. Осуществите кодирование двухбуквенных комбинаций четырех букв из задачи 1.





|  |  |
| --- | --- |
| Слово | Код |
| АА | 000 |
| АБ | 001 |
| АВ | 1000 |
| АГ | 1100 |
| БА | 010 |
| ББ | 011 |
| БВ | 1010 |
| БГ | 11100 |
| ВА | 1001 |
| ВБ | 1011 |
| ВВ | 11010 |
| ВГ | 11110 |
| ГА | 11011 |
| ГБ | 11101 |
| ГВ | 111110 |
| ГГ | 111111 |

**Задание 8**. Проведите кодирование всевозможных четырехбуквенных слов из задачи 3.

|  |  |
| --- | --- |
| Символ | P |
| А | 0,8 |
| Б | 0,2 |



|  |  |
| --- | --- |
| Слово | Код |
| АААА | 00 |
| АААБ | 01 |
| ААБА | 100 |
| ААББ | 1101 |
| АБАА | 101 |
| АБАБ | 11100 |
| АББА | 111010 |
| АБББ | 1111100 |
| БААА | 1100 |
| БААБ | 111011 |
| БАБА | 111100 |
| БАББ | 1111101 |
| ББАА | 111101 |
| ББАБ | 1111110 |
| БББА | 11111110 |
| ББББ | 11111111 |

**Задание 9**. Сравните эффективность кодов Фано и Хаффмана при кодировании алфавита из десяти букв, которые встречаются с вероятностями 0,3; 0,2; 0,1; 0,1; 0,1; 0,05; 0,05; 0,04; 0,03; 0,03.

|  |  |
| --- | --- |
| Буква | P |
| А | 0,3 |
| Б | 0,2 |
| В | 0,1 |
| Г | 0,1 |
| Д | 0,1 |
| Е | 0,05 |
| Ж | 0,05 |
| З | 0,04 |
| И | 0,03 |
| К | 0,03 |

Кодирование по Фано: 

Кодирование по Хаффману: 

Фано:



Цена кодирования l = 2\*0,3 + 2\*0,2 + 3\*0,1 + 3\*0,1 + 4\*0,1 + 4\*0,05 + 5\*0,05 + 5\*0,04 + 5\*0,03 + 5\*0,03 = 2,95

Хаффман:



Цена кодирования l = 2\*0,3 + 2\*0,2 + 3\*0,1 + 3\*0,1 + 4\*0,1 + 4\*0,05 + 5\*0,05 + 5\*0,04 + 5\*0,03 + 5\*0,03 = 2,95

В данном случае цены кодирования по Хаффману и по Фано равны, следовательно, для данного алфавита оба метода кодирования в равной степени эффективны.

**Задание 10**. Сравните эффективность двоичного кода Фано и кода Хаффмана при кодировании алфавита из 16 букв, которые встречаются с вероятностями 0,25; 0,2; 0,1; 0,1; 0,05; 0,04; 0,04; 0,04; 0,03; 0,03; 0,03; 0,03; 0,02; 0,02; 0,01; 0,01.



Кодирование по Фано: 

Цена кодирования l = 2\*0,25 + 3\*0,2 + 3\*0,1 + 4\*0,1 + 4\*0,05 + 4\*0,04 + 4\*0,04 + 4\*0,04 + 5\*0,03 + 5\*0,03 + 5\*0,03 + 5\*0,03 + 6\*0,02 + 6\*0,02 + 6\*0,01 + 6\*0,01 = 3,44

Кодирование по Хаффману: 

Цена кодирования l = 2\*0,25 + 2\*0,2 + 3\*0,1 + 4\*0,1 + 4\*0,05 + 5\*0,04 + 5\*0,04 + 5\*0,04 + 5\*0,03 + 5\*0,03 + 5\*0,03 + 6\*0,03 + 6\*0,02 + 6\*0,02 + 7\*0,01 + 7\*0,01 = 3,41

Цена кодирования алфавита по Хаффману меньше, чем по Фано, следовательно, для этого алфавита кодировать по Хаффману эффективнее.

**Вопросы для самоконтроля**

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

Среднее число элементарных сигналов, приходящихся на одну букву алфавита, определяется с помощью средней длины кодового слова. Для каждой буквы алфавита умножаем длину соответствующего кодового слова на вероятность появления этой буквы, а затем суммируем результаты для всех букв. Формула выглядит следующим образом:

Среднее число элементарных сигналов = Σ (Длина кодового слова \* Вероятность буквы)

**2. Префиксные коды.**

Префиксные коды - это коды, в которых ни одно кодовое слово не является префиксом другого кодового слова. Это означает, что ни одно кодовое слово не может быть началом другого кодового слова. Префиксные коды обеспечивают однозначность декодирования без необходимости использования разделителей между кодовыми словами.

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

Для записи кодированного сообщения требуется количество двоичных знаков, равное сумме длин кодовых слов для всех символов в сообщении. Каждый символ заменяется соответствующим кодовым словом, и эти кодовые слова объединяются для записи сообщения.

**4. На чем основано построение кода Фано?**

Построение кода Фано основано на разделении алфавита на две примерно равные по вероятности половины. Затем рекурсивно процесс разделения повторяется для каждой половины до достижения отдельных символов. В результате получаются префиксные коды, где кодовые слова для более вероятных символов более короткие, чем для менее вероятных символов.

**5. Что такое сжатие алфавита?**

Сжатие алфавита - это процесс уменьшения количества символов в алфавите или использования более компактного представления символов, чтобы уменьшить объем памяти или количество бит, требуемых для представления информации. Целью сжатия алфавита является экономия ресурсов и повышение эффективности передачи или хранения данных.

**6. Какой код самый выгодный?**

Самым выгодным кодом считается код, который обеспечивает минимальную среднюю длину кодового слова. Это позволяет достичь наилучшего сжатия и оптимального использования ресурсов.

**7. Основная теорема о кодировании.**

Основная теорема о кодировании (теорема Шеннона) утверждает, что средняя длина кодового слова в оптимальном префиксном коде для источника информации с заданным алфавитом и вероятностями символов близка к энтропии этого источника. Энтропия является мерой неопределенности и определяет нижнюю границу для средней длины кодового слова.

**8. Энтропия конкретных типов сообщений.**

Энтропия конкретных типов сообщений является мерой средней неопределенности или информационной содержательности сообщений. Она выражает количество информации, содержащейся в каждом символе или сообщении. Энтропия зависит от вероятностей появления различных символов или сообщений и используется для определения оптимальной длины кодовых слов в кодировании информации.

**Заключение**

Научился кодировать и декодировать по методу Фано и Хаффмана.