|  |  |
| --- | --- |
| 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  Митин Юрий Олегович  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.

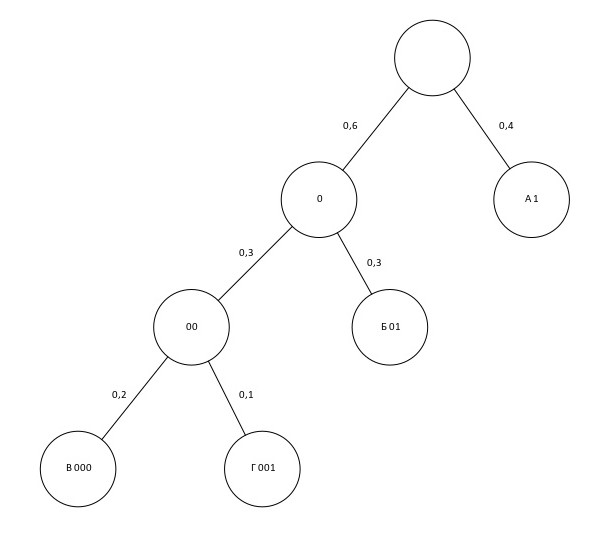
Коды Фано и Хаффмана являются оптимальными и префиксными. При построении искомых кодов будем применять как традиционный табличный способ кодирования, так и использовать "кодовые деревья".

**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:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Символ | А | Б | В | Г |
| P | 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:



|  |  |  |
| --- | --- | --- |
|  | Трехбуквенные комбинации: |  |
| Слово | P | Код |
| ААА | 0,512 | 0 |
| ААБ | 0,128 | 100 |
| АБА | 0,128 | 101 |
| АББ | 0,032 | 110 |
| БАА | 0,128 | 11100 |
| БАБ | 0,032 | 11101 |
| ББА | 0,032 | 11110 |
| БББ | 0,008 | 11111 |

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



|  |  |  |
| --- | --- | --- |
| Слово | P | Код |
| ААА | 0,512 | 0 |
| ААБ | 0,128 | 100 |
| АБА | 0,128 | 101 |
| АББ | 0,032 | 11100 |
| БАА | 0,128 | 110 |
| БАБ | 0,032 | 11101 |
| ББА | 0,032 | 11110 |
| БББ | 0,008 | 11111 |

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

Складываем синее, получаем оранжевое



|  |  |  |
| --- | --- | --- |
| Символ | P | Код |
| А | 0,4 | 1 |
| Б | 0,2 | 01 |
| В | 0,1 | 0010 |
| Г | 0,1 | 0011 |
| Д | 0,1 | 0000 |
| Е | 0,05 | 00010 |
| Ж | 0,05 | 00011 |

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

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



|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ХАФФМАН | | | | | | | | | |
| Буква | Код | P | | P | | P | | P | |
| A | 01 | 0,2 | 01 | 0,4 | 1 | 0,4 | 1 | 0,6 | 0 |
| B | 000 | 0,2 | 000 | 0,2 | 01 | 0,4 | 00 | 0,4 | 1 |
| C | 001 | 0,2 | 001 | 0,2 | 000 | 0,2 | 01 |  |  |
| D | 10 | 0,2 | 10 | 0,2 | 001 |  |  |  |  |
| E | 11 | 0,2 | 11 |  |  |  |  |  |  |

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

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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Слово | P | разбиение на подгруппы | | | | | | Код |
| АА | 0,16 | I | I | I |  |  |  | 000 |
| АБ | 0,12 | II | 001 |
| БА | 0,12 | II | I | 010 |
| ББ | 0,09 | II | 011 |
| АВ | 0,08 | II | I | I | I | 1000 |
| ВА | 0,08 | II | 1001 |
| БВ | 0,06 | II | I | 1010 |
| ВБ | 0,06 | II | 1011 |
| АГ | 0,04 | II | I | I | 1100 |
| ВВ | 0,04 | II | I | 11010 |
| ГА | 0,04 | II | 11011 |
| БГ | 0,03 | II | I | I | 11100 |
| ГБ | 0,03 | II | 11101 |
| ВГ | 0,02 | II | I | 11110 |
| ГВ | 0,02 | II | I | 111110 |
| ГГ | 0,01 | II | 111111 |

|  |  |  |
| --- | --- | --- |
| Слово | P | Код |
| АА | 0,16 | 000 |
| АБ | 0,12 | 001 |
| АВ | 0,08 | 1000 |
| АГ | 0,04 | 1100 |
| БА | 0,12 | 010 |
| ББ | 0,09 | 011 |
| БВ | 0,06 | 1010 |
| БГ | 0,03 | 11100 |
| ВА | 0,08 | 1001 |
| ВБ | 0,06 | 1011 |
| ВВ | 0,04 | 11010 |
| ВГ | 0,02 | 11110 |
| ГА | 0,04 | 11011 |
| ГБ | 0,03 | 11101 |
| ГВ | 0,02 | 111110 |
| ГГ | 0,01 | 111111 |

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

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



|  |  |  |
| --- | --- | --- |
| Слово | P | Код |
| АААА | 0,4096 | 00 |
| АААБ | 0,1024 | 01 |
| ААБА | 0,1024 | 100 |
| ААББ | 0,0256 | 1101 |
| АБАА | 0,1024 | 101 |
| АБАБ | 0,0256 | 11100 |
| АББА | 0,0256 | 111010 |
| АБББ | 0,0064 | 1111100 |
| БААА | 0,1024 | 1100 |
| БААБ | 0,0256 | 111011 |
| БАБА | 0,0256 | 111100 |
| БАББ | 0,0064 | 1111101 |
| ББАА | 0,0256 | 111101 |
| ББАБ | 0,0064 | 1111110 |
| БББА | 0,0064 | 11111110 |
| ББББ | 0,0016 | 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.

Кодирование по Фано:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ФАНО | | | | | | | |
| Символ | P | Разбиение на группы | | | | | Код |
| А | 0,3 | I | I |  |  |  | 00 |
| Б | 0,2 | II | 01 |
| В | 0,1 | II | I | I | 100 |
| Г | 0,1 | II | 101 |
| Д | 0,1 | II | I | I | 1100 |
| Е | 0,05 | II | 1101 |
| Ж | 0,05 | II | I | I | 11100 |
| З | 0,04 | II | 11101 |
| И | 0,03 | II | I | 11110 |
| К | 0,03 | II | 11111 |

Кодирование по Хаффману:



Фано:

|  |  |
| --- | --- |
| Символ | Код |
| А | 00 |
| Б | 01 |
| В | 100 |
| Г | 101 |
| Д | 1100 |
| Е | 1101 |
| Ж | 11100 |
| З | 11101 |
| И | 11110 |
| К | 11111 |

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:

|  |  |
| --- | --- |
| Символ | Код |
| А | 00 |
| Б | 10 |
| В | 110 |
| Г | 111 |
| Д | 0100 |
| Е | 0111 |
| Ж | 01010 |
| З | 01011 |
| И | 01100 |
| К | 01101 |

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 кодирования равно 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:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ФАНО | | | | | | | | |
| Символ | P | Разбиение на группы | | | | | | Код |
| X1 | 0,25 | I | I |  |  |  |  | 00 |
| X2 | 0,2 | II | I | 010 |
| X3 | 0,1 | II | 011 |
| X4 | 0,1 | II | I | I | I | 1000 |
| X5 | 0,05 | II | 1001 |
| X6 | 0,04 | II | I | 1010 |
| X7 | 0,04 | II | 1011 |
| X8 | 0,04 | II | I | I | 1100 |
| X9 | 0,03 | II | I | 11010 |
| X10 | 0,03 | II | 11011 |
| X11 | 0,03 | II | I | I | 11100 |
| X12 | 0,03 | II | 11101 |
| X13 | 0,02 | II | I | I | 111100 |
| X14 | 0,02 | II | 111101 |
| X15 | 0,01 | II | I | 111110 |
| X16 | 0,01 | II | 111111 |

|  |  |
| --- | --- |
| Символ | Код |
| X1 | 00 |
| X2 | 010 |
| X3 | 011 |
| X4 | 1000 |
| X5 | 1001 |
| X6 | 1010 |
| X7 | 1011 |
| X8 | 1100 |
| X9 | 11010 |
| X10 | 11011 |
| X11 | 11100 |
| X12 | 11101 |
| X13 | 111100 |
| X14 | 111101 |
| X15 | 111110 |
| X16 | 111111 |

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: 

|  |  |
| --- | --- |
| Символ | Код |
| X1 | 01 |
| X2 | 11 |
| X3 | 101 |
| X4 | 0000 |
| X5 | 1001 |
| X6 | 00011 |
| X7 | 00100 |
| X8 | 00101 |
| X9 | 00111 |
| X10 | 10000 |
| X11 | 10001 |
| X12 | 000100 |
| X13 | 100101 |
| X14 | 001100 |
| X15 | 0011010 |
| X16 | 0011011 |

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, то есть 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, может быть определено с использованием понятия энтропии информации. Энтропия информации представляет собой меру неопределенности или информационного содержания и выражается в единицах бит.

Для определения среднего числа элементарных сигналов на одну букву используется следующая формула:

N = 1 / H

где N - среднее число элементарных сигналов, приходящихся на одну букву, H - энтропия информации.

Энтропия информации для заданного алфавита может быть вычислена по формуле:

H = - Σ (P(i) \* log2(P(i)))

где P(i) - вероятность появления i-й буквы в тексте или сообщении.

**2. Префиксные коды.**

Префиксные коды - это коды, в которых ни одно кодовое слово не является префиксом другого кодового слова. Это означает, что никакая последовательность символов, соответствующая кодовому слову, не является началом другой последовательности символов, соответствующей другому кодовому слову.

Префиксные коды имеют важное значение при сжатии данных, так как они позволяют однозначно декодировать закодированную информацию без неоднозначности или возможности ошибки. Каждое кодовое слово в префиксных кодах является уникальным и не может быть представлено как комбинация других кодовых слов.

Примером префиксного кодирования является код Хаффмана, который строит оптимальные префиксные коды на основе частоты встречаемости символов в сообщении. Другой пример - код Фано, который также строит префиксные коды на основе разделения и объединения символов с учетом их вероятностей.

Префиксные коды позволяют сократить длину представления информации, используя более короткие коды для более часто встречающихся символов и более длинные коды для менее часто встречающихся символов, что приводит к сжатию данных.

**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.

Для префиксных кодов, таких как коды Хаффмана или коды Фано, длины кодовых слов могут быть разными, в зависимости от частоты появления символов. В этом случае, для определения общего количества бит требуется учесть длины всех кодовых слов, умножив каждую длину на соответствующую частоту появления символа в сообщении, а затем сложить полученные значения.

**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 или букв, используемых для представления информации, с целью уменьшить объем данных или сообщения. Вместо использования полного алфавита, сжатие алфавита заменяет некоторые символы или комбинации символов на более короткие представления или коды, что позволяет сократить количество бит, необходимых для хранения или передачи информации. Это позволяет экономить место на диске, ускорять передачу данных и уменьшать потребление пропускной способности сети.

**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 является мерой неопределенности или информационного содержания сообщения. Она позволяет оценить среднее количество информации, необходимое для представления каждого символа или элемента в сообщении. Энтропия измеряется в битах.

Энтропия конкретного типа сообщений зависит от вероятностей появления различных символов или элементов в этом сообщении. Чем более равномерное распределение вероятностей символов, тем выше энтропия и, соответственно, больше информации требуется для представления каждого символа.

**Заключение**

Научился кодировать и декодировать по методу Фано и Хаффмана.