МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ   
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ

ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ им. В. Г. ШУХОВА»

(БГТУ им. В.Г. Шухова)

Кафедра программного обеспечения вычислительной техники и автоматизированных систем

Лабораторная работа 1

по дисциплине: Теория информации

тема: «Исследование кодирования по методу Хаффмана. Оценка эффективности кода»

Выполнил: ст. группы ПВ-211

Шамраев Александр Анатольевич

Проверил:

Твердохлеб Виталий Викторович

Белгород 2023 г.

содержание отчета

[Задание 3](#_Toc127141206)

[1 Построить кодовое представление сообщения, вероятности появления символов в пределах алфавита которого приведены в табл.1. 4](#_Toc127141207)

[2 Построить кодовое представление сообщения, вероятности появления символов в пределах алфавита которого приведены в табл.2. 7](#_Toc127141208)

Задание

1. Построить кодовое представление сообщения, вероятности появления символов в пределах алфавита которого приведены в табл.1.   
   Таблица 1 – Вероятности появления символов в пределах алфавита исходного сообщения

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Символ | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 |
| Вероятность | 0.23 | 0.19 | 0.16 | 0.16 | 0.10 | 0.10 | 0.05 | 0.01 |

1. Построить кодовое представление сообщения, вероятности появления символов в пределах алфавита которого приведены в табл.2.  
   Таблица 2 – Вероятности появления символов в пределах алфавита исходного сообщения

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Символ | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 |
| Вероятность | 0.25 | 0.22 | 0.13 | 0.11 | 0.10 | 0.09 | 0.07 | 0.03 |

1. Для условий, приведенных в заданиях 1 и 2, выявить возможность построения альтернативных кодовых моделей сообщения. В случае обнаружения таковых, выявить наиболее эффективные из них по критериям и .

# Построить кодовое представление сообщения, вероятности появления символов в пределах алфавита которого приведены в табл.1.

Алгоритм построения кода Хаффмана по заданным вероятностям появления символов:

1. Отсортировать символы в порядке невозрастания вероятности их появления.
2. Выбрать 2 последних символа (в силу упорядоченности они имеют самую низкую вероятность появления) и объединить их в группу.
3. Исключить их из таблицы и вместо них добавить в таблицу полученную группу, не нарушая упорядоченность таблицы (вероятность группы символов равна сумме вероятностей символов ее составляющих). Вставку группы в таблицу можно осуществлять методом поиска позиции элемента, используемом в алгоритме сортировки вставками (таким образом нет необходимости заново сортировать таблицу).
4. Попутно строим Дерево Хаффмана. Каждый элемент группы будем воспринимать как узел искомого дерева (правую ветвь будем обозначать 1, левую - 0).
5. Повторяем пункт 2-4, раз, где . Таким образом, в конечном итоге в таблице останется всего 1 элемент.
6. Начинаем обход дерева в глубину. Записываем последовательность ребер (0 или 1), по которым мы прошли, по пути к листу, таким образом, получая конечные коды символов.

Поскольку сначала мы добавляли элементы с самой низкой вероятностью – то, чтобы дойти до них, необходимо будет пройти наибольшее число ребер дерева, а, соответственно, и их код будет самым длинным (а у символов с наибольшей вероятностью, наоборот – самым коротким).



Полученная таблица с кодами

|  |  |
| --- | --- |
| S1 | 10 |
| S2 | 11 |
| S3 | 011 |
| S4 | 010 |
| S5 | 001 |
| S6 | 0001 |
| S7 | 00001 |
| S8 | 00000 |

Так как символы изначально были упорядочены по невозрастанию частоты появления, то их коды также получились упорядоченными по неубыванию длины.

Из полученного дерева видно, что оно имеет несколько альтернативных, так как некоторые символы и группы символов имеют одинаковую вероятность. Однако обмен местами символов с одинаковой вероятностью не влияет на эффективность кодирования, поэтому такие альтернативные деревья не имеет особого смысла рассматривать. Рассмотрим случай с обменом группы и одного из символов, имеющих вероятность 0.16. Так они все находятся на одинаковом удалении от корня дерева, то их обмен также не приведет к качественному изменению кода и необходимости перестроения дерева, а потому бессмысленный. Поэтому можно сделать вывод, что с точки зрения качественных характеристик полученного кода в данном примере есть только один вариант дерева (остальные будут иметь идентичные характеристики).

# Построить кодовое представление сообщения, вероятности появления символов в пределах алфавита которого приведены в табл.2.



Полученная таблица с кодами

|  |  |
| --- | --- |
| S1 | 00 |
| S2 | 11 |
| S3 | 011 |
| S4 | 100 |
| S5 | 101 |
| S6 | 0101 |
| S7 | 01001 |
| S8 | 01000 |

Однако здесь возможен еще один вариант построения дерева. Если на втором шаге объединить S5 и S6, то дерево кардинально изменится.



Полученная таблица с кодами

|  |  |
| --- | --- |
| S1 | 10 |
| S2 | 00 |
| S3 | 111 |
| S4 | 011 |
| S5 | 1101 |
| S6 | 1100 |
| S7 | 0100 |
| S8 | 0101 |

Как видим, в обоих случаях коды получились разной длины.

# Для условий, приведенных в заданиях 1 и 2, выявить возможность построения альтернативных кодовых моделей сообщения. В случае обнаружения таковых, выявить наиболее эффективные из них по критериям и .

Альтернативны кодовые модели были рассмотрены в прошлых заданиях. Для задания 1 имеет смысл рассматривать только 1 модель.

Вычислим :

– число символов в сообщении (т.к. все вероятности округлены до сотых, то возьмем )

равен сумме всех произведений кол-ва символов в сообщении на их информационный вес.

Вычислим :

Во 2-м задании мы получили 2 альтернативных дерева, имеющих различные по длине коды (остальные будут идентичны одному из этих деревьев).

Для первого дерева:

Как видим, данный код Хаффмана вместо сжатия, наоборот, расширил исходное сообщение, поэтому очевидно, что применять код не имеет смысла – лучше воспользоваться исходной 3-битной кодировкой.

Для второго:

Вторая кодовая модель эффективнее первой по обоим параметрам (и в отличие от нее она способна сжать исходные данные).

Вывод: были получены навыки построения кода Хаффмана и определения основных метрик кодовых моделей сжатия информации.