

Практика №2. Построение префиксного кода с помощью алгоритма Фано. Кодирование и декодирование сообщения.

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ.

Определение 2.1. m -значным кодированием сообщений α алфавита A , в кодовом алфавите B , называется отображение $F: S \rightarrow B^*$, где S – множество сообщений, B^* – множество всех слов в алфавите B , содержащем m символов. $F(\alpha)$ называется кодом сообщения α .

Определение 2.2. Кодирование называется алфавитным, если оно сохраняет произведения слов. Для алфавитного кодирования коды однобуквенных сообщений называются элементарными.

Определение 2.3. Соответствие между буквами алфавита A и их элементарными кодами при алфавитном кодировании называется схемой кодирования.

Определение 2.4. Схема кодирования называется префиксной, если никакой элементарный код не является началом другого элементарного кода.

Определение 2.5. Средней длиной элементарного кода называется $\bar{l} = \sum_{i=1}^n p_i \cdot l_i$, где $l_i = l(\beta_i)$ – длина элементарного кода β_i .

Определение 2.6. Коэффициентом относительной эффективности кодирования называется величина

$$\eta = \frac{H(X)}{\bar{l}}$$

Определение 2.7. Оптимальным для данного стохастического источника сообщений называется такое алфавитное кодирование, для которого достигается минимальная средняя длина элементарного кода.

Теорема 2.1. Для любого дискретного источника, характеризующегося вероятностной схемой X с конечным алфавитом и энтропией $H(X)$, существует m -ичный префиксный код, в котором средняя длина кодового слова удовлетворяет неравенству

$$\frac{H(X)}{\log m} \leq \bar{l} < \frac{H(X)}{\log m} + 1.$$

При построении оптимальных кодов можно использовать алгоритмы Шеннона-Фано или Хаффмана.

Алгоритм Шеннона-Фано.

1. Множество сообщений данной вероятностной схемы располагается в порядке убывания вероятностей.
2. Множество сообщений разбивается на части, приблизительно равные по суммарной вероятности. Первой части присваивается ноль, второй единица.
3. К каждой из частей применяются действия пункта 2.

Условием окончания работы алгоритма является наличие одного символа в каждой из подгрупп.

ПРИМЕР

Задание. Произвести статистическую обработку данного сообщения, считая, что источник сообщений периодически, достаточно долго выдаёт следующую последовательность символов 12342334551233. Определить энтропию, приходящуюся в среднем на одну букву, длину кода при равномерном кодировании и избыточность. Построить схемы алфавитного кодирования методами Фано. Найти среднюю длину элементарного кода, эффективность сжатия.

Статистическая обработка приведённого сообщения, была выполнена в предыдущем примере, где и была получена вероятностная схема

X	1	2	3	4	5	Σ
n	2	3	5	2	2	14
w	$\frac{2}{14}$	$\frac{3}{14}$	$\frac{5}{14}$	$\frac{2}{14}$	$\frac{2}{14}$	1

Построим схему кодирования по алгоритму Шеннона-Фано.

СИМВОЛ	P				код
3	$\frac{5}{14}$	0	0		00
2	$\frac{3}{14}$		1		01
1	$\frac{2}{14}$	1	0		10
4	$\frac{2}{14}$		1	0	110
5	$\frac{2}{14}$			1	111

Средняя длина кодового слова равна

$$\bar{l} = \frac{5}{14} \cdot 2 + \frac{3}{14} \cdot 2 + \frac{2}{14} \cdot 3 \cdot 2 = 2.29$$

Коэффициент эффективности равен

$$\eta = \frac{2.21}{2.29} = 0.97$$

ЗАДАНИЕ

Провести статистическую обработку текста.

Определить энтропию, приходящуюся в среднем на одну букву, длину кода при равномерном кодировании и избыточность.

Построить схему алфавитного кодирования для однобуквенных сочетаний методом Шеннона-Фано. Найти среднюю длину элементарного кода, эффективность сжатия. Закодировать текст. Декодировать текст.

Построить схему алфавитного кодирования для двухбуквенных сочетаний методом Шеннона-Фано. Найти среднюю длину элементарного кода, эффективность сжатия, сравнить с результатами для однобуквенных сочетаний. Закодировать текст. Декодировать текст.