

Вариант 13

Эффективное кодирование. Кодирование для ДИБП. Кодовые слова переменной длины. Теорема кодирования. Алгоритм Хаффмена

Реальные источники информации обладают большой избыточностью. Поэтому для ее уменьшения прибегают к эффективному кодированию.

Эффективное кодирование – это процедуры направленные на устранение избыточности.

Эффективное кодирование для ДИБП (дискретный источник информации без памяти):

Пусть ДИБП выдает буквы или символы каждые τ_s секунд. Каждый символ выбирается из конечного алфавита $A \in \{a_k\}, k=1,2,\dots,L$ с вероятностью $p(a_k)$.

Энтропия такого источника определяется по формуле (2.4) и ограничивается сверху значением, вычисляемым по (4.5), т.е. $H(X) \leq \log_2(L)$. Как говорилось выше, знак « \Rightarrow » выполняется, если вероятности символов на выходе источника

одинаковы и равны $p = \frac{1}{L}$.

Кодовые слова переменной длины:

Если символы источника не равновероятны, то более эффективно использовать кодовые слова переменной длины.

Пример: код Морзе (19 век). Символам, возникающим более часто, ставятся в соответствие более короткие кодовые слова, а символам, возникающим менее часто, сопоставляются более длинные кодовые слова. Такой метод кодирования, который требует знания вероятностей появления символов источника, называется **энтропийным**.

В общем, **префиксное условие** кода требует, чтобы для кодового слова длины K ($b_1 \dots b_M b_{M+1} \dots b_K$) не существовало других кодовых слов длины $M < K$ с элементами ($b_1 \dots b_M$). Это свойство делает кодовые слова однозначно декодируемыми.

Критерий оптимальности однозначно декодируемых кодов переменной длины имеет вид:

$$\bar{K} = \sum_{k=1}^L n_k p(a_k) = \min, \quad (4.13)$$

Теорема Шеннона кодирования ДИБП:

Пусть X - ансамбль символов ДИБП с конечной энтропией $H(X)$ и выходными символами из алфавита $A = \{a_1, \dots, a_L\}$ с вероятностями выхода $p(a_k)$, $k = 1, 2, \dots, L$. Тогда существует возможность создать код, который удовлетворяет префиксному условию и имеет среднюю длину K , удовлетворяющую неравенству: $H(X) \leq K < H(X) + 1$

Алгоритм Хаффмена:

Рассмотрим метод кодирования Хаффмена, применение которого к любому произвольному ансамблю символов ДИБП обеспечивает получение оптимального по критерию префиксного кода. Алгоритм кодирования Хаффмена. Критерий оптимальности кодов Хаффмена – минимум средней длины кодового слова.

Пример: