

# АДАПТИВНАЯ СИСТЕМА СЖАТИЯ НА ОСНОВЕ КОДА ХАФФМАНА

Ст. преп., к.т.н. Зубань Ю.А., студ. Петров В.В.

Как и всех адаптивных систем, работу устройства сжатия построенного на основе кода Хаффмана можно представить в виде такой структуры (рис. 1), в которую входит моделирование источника информации и кодирование.



Рисунок 1 – Общая структура адаптивной системы сжатия информации

Моделирование источника системой определяет степень адаптивности последней к информационному потоку. Тем самым, определяя степень и скорость кодирования (сжатия). Выбор алгоритма и накладывает свои ограничения на модель источника, – как бы ни стремилась модель источника к абсолютной, построение кода (сжатие) и/или все названные выше параметры могут привести к тому, что часть затрат потраченная на процесс моделирования источника будет не оправдана или попросту не нужна.

В данном случае выбран код Хаффмана, являющийся оптимальным неравномерным, обладающий свойством

префиксности, и адаптивный (динамический) алгоритм его построения. Это существенно влияет на процесс моделирования источника и накладывает такие ограничения:

а) каким бы ни был входной алфавит, вероятности его элементов при кодировании сводятся к  $2^{-c}$ , где  $c = 0, 1, 2, \dots$ ;

б) алгоритм построения кода и соответственно модификации дерева Хаффмана предусматривает сортировку и поиск.

Первое – влияет на степень сжатия, уменьшая ее. Это наиболее проявляется при входном алфавите, состоящем из двух символов, где модель может изменяться в соответствии с источником, но при кодировании коды символов не меняются и сжатие не происходит. Второе – влечет за собой затраты времени на поиск и сортировку элементов.

Исходя из вышесказанного, и с учетом проведенных исследований получены критерии, которые должны лечь в основу построения адаптивной системы сжатия на основе кода Хаффмана:

а) условием для изменения модели источника и перестройки кодового дерева Хаффмана в соответствии с полученной статистикой является выражение:

$$N(a_{\max L}) > N(a_{\min L-1}), L=1, 2, 3, \dots, \quad (1)$$

где  $N(a_i)$  – вес (частота повторений или вероятность) символа;

$L$  – номер уровня дерева Хаффмана, поступившего символа,  $L = 0$  – корень дерева.

Смысл формулы (1) заключается в том, что перестраивать дерево (модифицировать код) нужно, если



вес символа  $a_{\max}$  максимального на уровне  $L$ , станет больше веса символа  $a_{\min}$ , минимального на  $L-1$  уровне. По сути это учитывает ограничения кода на модели источника;

б) процесс кодирования с учетом служебной информации, как сжатие. Длина выходной последовательности ограничена выражением:

$$\sum_{i=1}^{k_1} (N(a_i) \cdot n_{\text{вх.}}(a_i)) \Big|_{k_1=1,2,3,\dots}^{\text{вх.}} > \sum_{i=1}^{k_2} (N(a_i) \cdot n_{\text{вых.}}(a_i)) \Big|_{k_2=1,2,3,\dots}^{\text{вых.}} + I_{\text{служ.}}, k_1 = k_2, \quad (2)$$

где  $n_{\text{вх.}}(a_i)$  – разрядность входной кодовой комбинации,  $n_{\text{вх.}}(a_1) = n_{\text{вх.}}(a_2) = \dots = n_{\text{вх.}}(a_k) = n$  – для равномерного входного кода;

$n_{\text{вых.}}(a_i)$  – разрядность выходной кодовой комбинации Хаффмана для символа  $a_i$ ;

$I_{\text{служ.}}$  – длина служебной части, необходимой для декодирования.

Условие (2) говорит о том, что при сжатии информации между идущими последовательно двумя модификациями (моделированиями источника), суммарное количество бит пришедшее на вход системы не должно быть больше закодированных плюс информация об изменении модели.

Таким образом, применение рассмотренных критериев построения системы сжатия данных на основе адаптивного алгоритма построения кода Хаффмана позволяет оптимизировать и упростить реализацию данного устройства на практике, существенно улучшив скорость и степень сжатия.