Неравномерное побуквенное кодирование. Префиксные коды. Неравенство Крафта

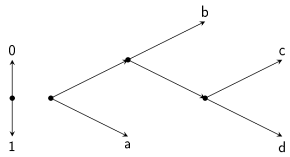
4 января 2023 г. 21:20

Неравномерное побуквенное кодирование (variable-length coding, VLC)

Рассмотрим дискретный источник без памяти
То есть имеем алфавит от 1 до M $X = \{1, \dots, M\}$, вероятности букв $\{p_1, \ \dots, \ p_M\}$, сумма которых равна 1. Каждой букве из X сопоставляем кодовые слова длины $l_1, \dots, l_M$: $C = \{c_1, \dots, c_M\}$.
Средняя длина кодового слова - матожидание от длины кодового слова:
$\overline{l} \ = E[l_i] = \sum_{i=1}^{M} p_i l_i$
H(X) - нижняя граница для средней длины кодового слова

Наилучший неравномерный побуквенный код - код Хаффмана

Для корректности алгоритма код должен быть однозначно декодируемым, т.е. префиксным. Его можно представить в виде двоичного кодового дерева



Свойства префиксного кода:
1. Любое кодовое слово не должно быть началом другого
2. Декодирование однозначно (причем однозначно декодируемый код не обязательно префиксный)
3. Кодовым словам соответствуют только листья двоичного кодового дерева
4. Древовидный код является префиксным

Неравенство Крафта

Theorem
Для любого однозначно декодируемого двоичного кода объёмом M с длинами кодовых слов $l_1, \dots, l_M$ справедливо неравенство:

$$\sum_{i=1}^{M} 2^{-l_i} \leq 1.$$

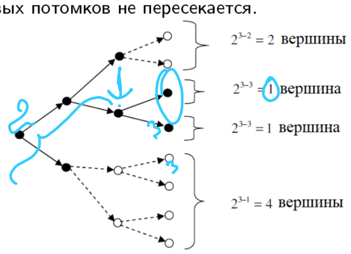
Необходимым и достаточным условием существования префиксного кода объемом M с длинами кодовых слов $l_1, \dots, l_M$ является выполнение неравенства:
$\sum_{I=1}^{M} 2^{-l_i} \ \leq 1$

Свойства неравенства Крафта:
1. Необходимость: Неравенство Крафта верно для любого префиксного кода
2. Достаточность: Если неравенство Крафта выполняется, то существует код с заданным набором длин кодовых символов

Необходимость

Неравенство верно для любого префиксного кода.

Выберем L, такое что $L \geq \max_i l_i$. Концевая вершина исходного дерева, расположенная на глубине l_i, имеет 2^{L-l_i} потомков на глубине L. Причем, множества концевых потомков не пересекается.



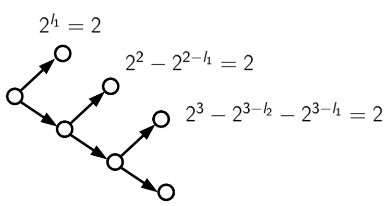
$$\sum_{i=1}^{M} 2^{L-l_i} \leq 2^L.$$

Если формулу с рисунка разделить на 2^L, получится неравенство Крафта

Достаточность

Если неравенство Крафта выполняется, то существует код с заданным набором длин кодовых слов.

Пример: $l_1 = 1$, $l_2 = 2$, $l_3 = l_4 = 3$.



На каждом шаге количество "свободных" вершин равно 2, т.е., можно поместить следующее кодовое слово.

- Сортируем $\{l_i\}$ по убыванию.
- Из всего числа вершин на ярусе l_1 выберем любую и закрепим её за первым кодовым словом. Продолжим оставшиеся вершины до яруса l_2. На этом ярусе будет свободно $2^{l_2} - 2^{l_2 - l_1}$ вершин.
  Умножив $\sum_{i=1}^{M} 2^{-l_i} \leq 1$ для $M = 2$ на 2^{l_2}, получим $2^{l_2} - 2^{l_2 - l_1} \geq 1$,
  то есть как минимум одна свободная вершина есть.
- $2^{l_3} - 2^{l_3 - l_2} - 2^{l_3 - l_1} \geq 1$.
- ...
- $2^{l_M} - 2^{l_M - l_{M-1}} - 2^{l_M - l_{M-2}} - \ldots - 2^{l_M - l_1} \geq 1$

Избыточность неравномерного кода - разность средней длины кодового слова и энтропии