[proof]topsep=30pt, itemsep=5pt, font=,

Критерий Маркова взаимной однозначности алфавитного кодирования

Основные определения

Определение 1. *Пусть задан некий конечный алфавит A. Слово в алфавите A - это конечная последовательность символов из A.*

Определение 2. *Рассмотрим соответствие между буквами алфавита A и некоторыми словами в алфавите B, где каждой букве a_i соответствует непустое слово B_i. Такое соответствие называется схемой и обозначается через $\sum$.*

Определение 3. *Схема определяет алфавитное кодирование следующим образом: каждому непустому слову $a = \alpha_{i_1} \dots \alpha_{i_n}$ ставится в соответствие слово $b = \beta_{i_1} \dots \beta_{i_n}$, называемое кодом слова. Слова $\beta_1 \dots \beta r$ назваются элементарными кодами.*

Определение 4. *Обозначим за l(B) длину слова, то есть количество букв в этом слове.*

Определение 5. *За $L = l(B_1 \dots B_r)$ бозначим длину схемы, где $l(B_i) = l_i$, а L - их сумма.*

Определение 6. *Нетривиальное разложение - это разложение вида $B_i = \beta' B_{i_1} \dots B_{i_w} \beta''$, то есть разложение, отличное от разложения $B_i = B_i$. Более того, β', β'' отличны от элементарных кодов.*

Определение 7. *Параметр $w \in Z \cup \{0\}$. Обозначим через W максимум чисел w взятый по всем разложениям B_i и по всем i, то есть $W = max(w)$.*

Определение 8. *За $Z^N(A)$ обозначим множество всех непустых слов из алфавита A длины не более N.*

Определение 9. *Слово B, допускающее не менее двух расшифровок, называется неприводимым, если каждое слово B', получающееся из B путем вырбасывания непустого куска, допускает не более одной расшифровки.*

Критерий Маркова

Теорема 1. *Для любой схемы алфавитного кодирования $\sum$ с параметрами W, L, r существует такое N_0,*

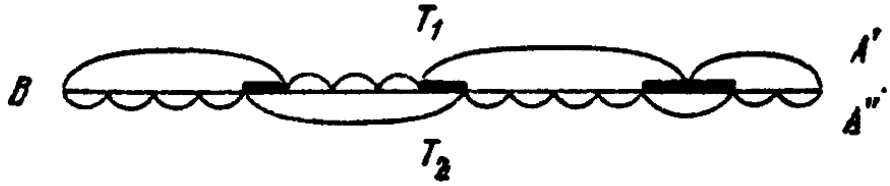
$$N_0 \leq [\frac{(W+1)(L-r+2)}{2}]$$

, что проблема взаимной однозначности алфавитного кодирования сводится к аналогичной проблеме для кодирования конеченого множества $S^{N_0}(A)$.

Доказательство 1. *Если в алфвитном кодировании нарушена взаимная однозначность, то найдется такое слово B, которое допускает по крайней мере две различные расшифровки A' и A". Для доказательства теоремы достаточно показать, что можно найти также такое слово B, что для его расшифровок A' и A" имеют места неравенства*

$$l(A'), l(A'') \leq [\frac{(W+1)(L-r+2)}{2}]$$

. С самого начала можно предполагать, что слово B неприводимо так как из любого слова допускающего несколько расшифровок мы можем сделать неприводимое. Рассмотрим две его расшифровки A' и A".



Видно, что с ним связаны два разбиения слова B на элементарные коды - верхнее T_1 и нижнее - T_2.

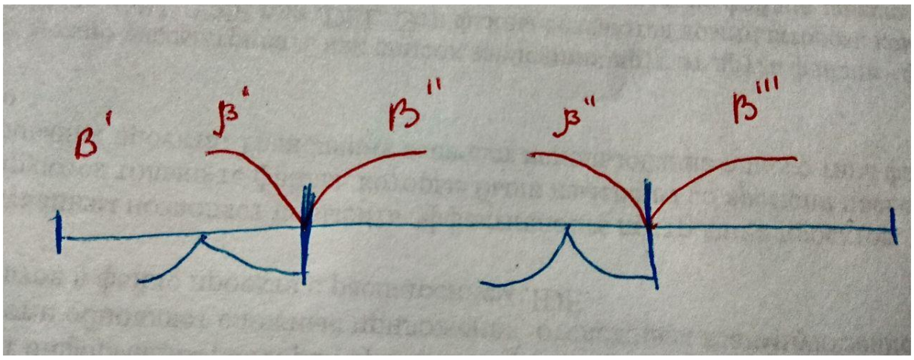
Возьмем произведение этих разбиений T, полученное путем одновременного разбиения T_1 и T_2.

Слова разбиения T разделим на два класса - к первому будут относиться элементарные коды, ко второму - все остальные.

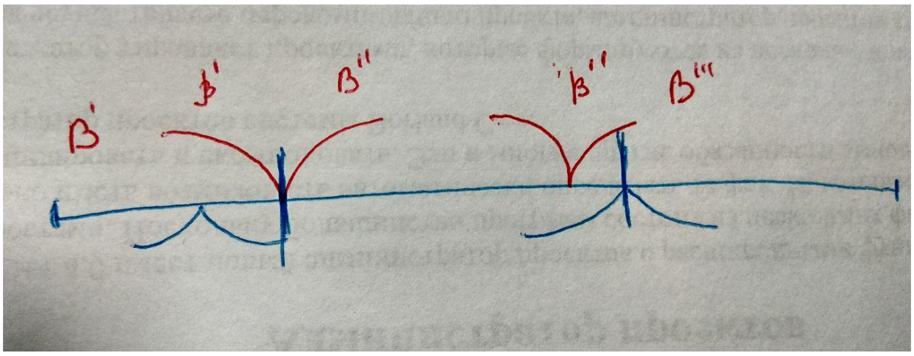
Обозначим слово на рисунке выше как

$$B = B'\beta' B''\beta'' B'''$$

Покажем, что любые два слова из второго класса различны. Предположим обратное. Пусть $\beta' = \beta''$ из второго класса (отмечены на рисунке жирной линией) равны. Выбросим B'', β'' и склеим обе части (так как мы предположили, что $\beta' = \beta''$). Мы покажем, что для слова $B'\beta'B'''$ полученного в результате склейки во всех случаях мы получим 2 расшифровки (на картинке мы сможем пройти и сверху и снизу). Для расположения слов β' и β'' возможны следующие случаи:



В данному случае мы можем пройти как сверху (отмечено красным), так и снизу. Поменяв местами красные и синие линии (пути) мы получим еще один вариант.



В данному случае мы тоже можем пройти как сверху (отмечено красным), так и снизу. Поменяв местами красные и синие линии (пути) мы получим еще один вариант.

Мы только что показали, что для любого слова $B'\beta'B'''$ возможны как минимум 2 расшифровки.

Теперь оценим количество слов из второго класса p. Видно, что их число не может быть больше чем их непустых начал.

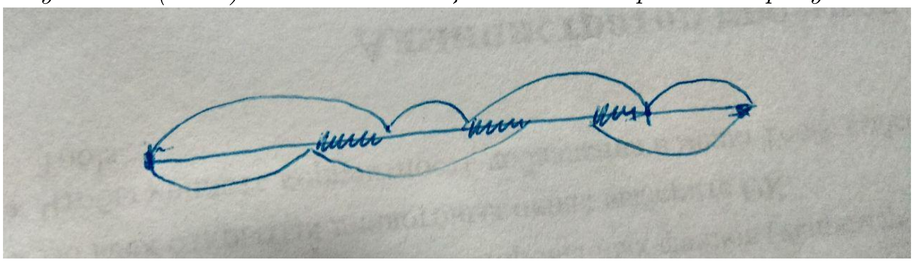
$$p \leq (l(B_1) - 1) + \dots (l(B_r) - 1) = L - r$$

Видно, что так как таких слов не более L - r, то и разбивается наше слово на не более чем L-r + 1 кусков.

Теперь, оценим их длины.

Возможны 2 случая.

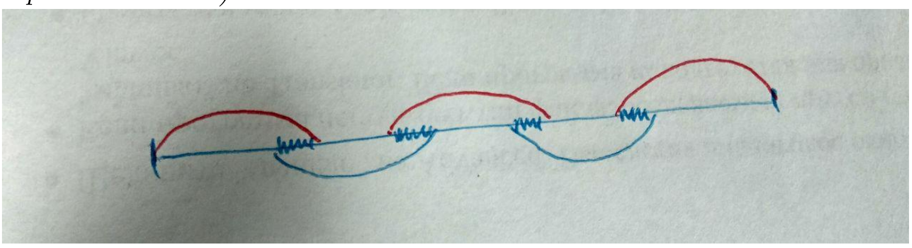
Случай 1 - (L - r) не делится нацело на 2.Изобразим на рисунке.



Элементарный код, например, сверху сначала берет один кусок второго типа, а потом берет их по два. $\frac{L-r-1}{2}+1$ *кусков второго типа и* $(\frac{L-r-1}{2}+1)*W$ *- элементарных кодов между ними.*

Случай 2 - (L - r) делится нацело на 2. Тут возможны 2 подслучая. Подслучай 1 - элементарный код берет сначала 1 кусок второго типа, а потом берет их по 2 (изображено красным).

Подслучай 2 - элементарный код берет по 2 куска второго типа (изображено синим).



Оценка для первого случая

$$\frac{L-r-2}{2}+2+W(\frac{L-r-2}{2}+1)$$

Оценка для второго случая

$$\frac{L-r-2}{2}+W(\frac{L-r}{2}+1)$$