

4.5. Оценки стоимости кодирования разделенного кода

Теорема 3 (нижн. гр. стоим-ти кодир-я разд. кода):

Для любого разделенного кода \tilde{C} источника S : $C(\tilde{C}, S) \geq H(S)$

Д-во: Пусть $l_i = \tilde{P}(x_i) \quad i=1, N$. $C(\tilde{C}, S) = \sum_{i=1}^N p_i l_i = \sum p_i \log 2^{l_i}$

Рассмотрим:

$$H(S) - C(\tilde{C}, S) = \sum p_i \log \frac{1}{p_i} + \sum p_i \log 2^{-l_i} = \sum p_i \log \frac{2^{-l_i}}{p_i}$$

Всегда первое: $\ln x \leq x-1$. Тогда:

$$H(S) - C(\tilde{C}, S) = \log e \sum p_i \left(\frac{2^{-l_i}}{p_i} - 1 \right) = \log e (\sum 2^{-l_i}) - (\sum p_i) = \log e (\sum 2^{-l_i} - 1) \stackrel{\text{no t. 2}}{\leq} 0$$

Зад $C(\tilde{C}, S) = H(S) \Leftrightarrow p_i = 2^{-l_i} \quad i=1, N$

Опр Оптимальное кодир-е — то, чьё стоимость наименьшая

Теорема 4 (верхняя граница стоим-ти оптималь. кодир-я разд. кода):

Э разд. код такой, что: $C(\tilde{C}, S) < H(S) + 1$

Д-во: Пусть s_i — наим. целое число такое, что: $\log \frac{1}{p_i} \leq s_i < \log \frac{1}{p_i} + 1$

Имеем: $\sum 2^{-s_i} \leq \sum 2^{-\log \frac{1}{p_i}} = \sum p_i = 1 \Rightarrow$ числа s_i это делители код. слов

Для разд. код S (но т. 2)

Оценки стоимость такого кодир-я:

$$C(\tilde{C}, S) = \sum p_i s_i < \sum p_i (\log \frac{1}{p_i} + 1) = H(S) + 1$$

Зад Кодир-е в Т.4 — это кодирование Шеннона

Опр Код. Шеннона:

Занумеруем буквы x_1, \dots, x_N так, что $p_1 \geq p_2 \geq \dots \geq p_N$

Опред Γ_i : $\Gamma_1 = 0, \Gamma_{i+1} = \Gamma_i + p_i$

Положим $\Gamma(x_i)$ равным первым $\lceil \log \frac{1}{p_i} \rceil$ двоичн. знакам числа Γ_i

Ex $A = \{a, b, c, d, e, f, g\} \quad P = \{0.5, 0.2, 0.1, 0.07, 0.05, 0.05, 0.03\}$

$$\Rightarrow \Gamma = \{0, 0.5, 0.7, 0.8, 0.87, 0.92, 0.97\}$$

Ищем двоичн. код. слов:

$$2^{-1} \leq 0.5 < 1 \quad (1) \quad \leftarrow \lceil \log \frac{1}{p_2} \rceil$$

$$2^{-3} \leq 0.2 < 2^{-2} \quad (3)$$

$$2^{-4} \leq 0.1 < 2^{-3} \quad (4)$$

$$2^{-5} \leq 0.07 < 2^{-4} \quad (5)$$

$$2^{-6} \leq 0.05 < 2^{-5} \quad (5)$$

$$2^{-7} \leq 0.03 < 2^{-6} \quad (6)$$

Вспомогат. таблица:

$$2^{-1} = 0.5$$

$$2^{-2} = 0.25$$

$$2^{-3} = 0.125$$

$$2^{-4} = 0.0625$$

$$2^{-5} = 0.03125$$

$$2^{-6} = 0.015625$$

тут можно ещё и сократить коды

0.0	0.5	0.7	u.t.o
0.0	0.0	0.4	
0.0	0.0	0.8	
0.0	0.0	1.6	
0.0	0.0	1.2	

Упр Код Шеннона префиксный.

4.6. Энтропийное кодирование

Уз 4.5 знаем, что $H(S) \leq \overline{C(A, S)} < H(S) + 1$

$x_i \rightarrow \dots$ аффинитное код-е

$(x_1, \dots, x_n) \rightarrow \dots$ блочное код-е (на самом деле адр-е, но с большим аффинитом)

Чтв S' - источник $A' = f(x_1, \dots, x_n)$ $A = \{x_i\}$

$$H(S) \quad H(S') = n \cdot H(S)$$

Тогда из тз, ч S' в код:

$$n \cdot H(S) \leq \overline{C(A, S')} < n \cdot H(S) + 1$$

$$H(S) \leq \frac{\overline{C(A, S')}}{n} < H(S) + \frac{1}{n} \xrightarrow{n \rightarrow \infty} H(S)$$

скорость на букву

Ура Блочному кодированию! Смыслишко стремится к $H(S)$! (лучшаем)

4.7. Код Хаффмана

Опн Запущеруем буквы так, что $p_1 \geq p_2 \geq \dots \geq p_N$
Сканируем буквы с нач. вероятн-ми.

$$S \rightarrow S'$$

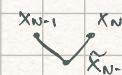
$$p_1 \dots p_N \quad \tilde{p}_1, \dots, \tilde{p}_{N-2}, \tilde{p}_{N-1} = p_{N-1} + p_N$$

$$p_1 \quad p_{N-2}$$

$$A \rightarrow A'$$

Строим дерево от листьев. Листья - буквы x_1, \dots, x_N

Сканика 2х букв соотв-ет такому действию в дереве:



Вновь переупорядочиваем вершины \hat{A} так, что: $\tilde{p}_1 \geq \tilde{p}_2 \geq \dots \geq \tilde{p}_{N-1}$
и продолжаем те же шаги.

За $N-1$ шаг дерево построено. В листьях - буквы, путь к ним - код.

Зад Код Хаффмана префиксный

Теорема 5

Код Хаффмана оптимальн.

Схема д-ва:

исп-е след. утв-я:

Чтв 1 В любом опт. коде код. слово, соотв. наимен. вероятному сообш-ю, имеет наиб. длинну

Чтв 2 В оптимальн. блочн. префиксн. коде 2 наимен. вероятн. сообш-я код-са словами один. длины, одно из к-рд заканч. нулем, а другое - единицей

Чтв 3 Если оптимальн. об. пр. код для \hat{S} , то опт. и получ. из него код для S .

5. Сжатие текстов: сжатие без потерь

5.1. Постановка задачи

В n.4 — коды источника (не знаем, какой текст будет порождаться источником, \Rightarrow готовимся к любому тексту)

В n.5 — коды текста (имеем дело с конкр. текстом и не важно, каким источником были порождены)

$$S: A = \{a, b\} \quad D = \left\{ \frac{253}{256}, \frac{3}{256} \right\} \quad H(S) \approx 0.091976$$

Рассмотрим текст из алфавита источника одинак. 256:

$256 \cdot H(S) \approx 24$ бита \leftarrow при исп. линк. код-а

А если общ. код Хаффмана, то 256 битов!

Если текст $a \underline{bab} \dots \underline{ab} \quad 100 \dots 01$ т.е. всего 10 битов!

Коммогоровская сложность

? Какова max степень сжатия сообщ-я, не зависящ. от источника, а лишь от оно порождалось, и от выч. способностей входных данных

Абсолютный код сообщения

Опр Коммогоровская сложность — макс. длина программы, порожд. сообщ-я

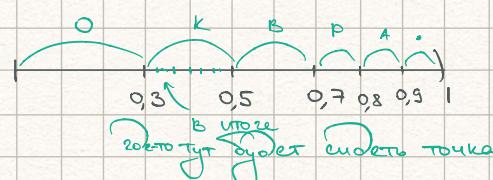
Утв Комм. сложн-ть альг. невычислима

5.2. Арифметическое кодирование

Текст кодируется точкой на интервале $[0, 1]$

Ex КОВ.КОРОВА

Посчитали:	O	3	0.3
	K	2	0.2
	B	2	0.2
	D	1	0.1
	A	1	0.1
	.	1	0.1



Выберем точку из середины после интервала. В двоичн. предст. (за запятой) мы берём ровно столько знаков, чтобы соответствуя коду число достоверно попадало в этот интервал

т.е. делаем, пока есть однозначность. если никто не выб., пока заканчивать

Опр $k(l)$ — миним. число такое, что $l \geq \frac{1}{2^{k(l)-1}}$, где $l = p_1 \dots p_n$

длина арифм. кода сообщ-я

$$0.\underline{1011} \dots$$