5.3. Словарные методы сжатия

- LZ'77

окно = словарь + буфер

1) словарь = ∅, буфер содерж. начало сообщ-я
2) ищем в словаре максим. фрагмент, совп. с началом содержимого буфера
3) записываем код (i, j, s), где
   - i – номер позиции в словаре, с ктр нач. совп. фрагмент
   - j – длина совп. фрагм.
   - s – первый симв. буфера, след. за фраг.

<u>Х-ки:</u>
- степень сжатия: 2-4
- тип данных: любые, но лучше разнородн.
- сим-ть: 10:1

- LZ'78

нет скольз. окна

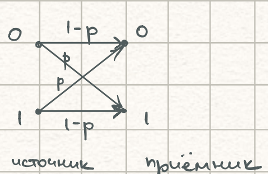
1) словарь = ∅, в буфер помещ. всё сообщ-е
2) ищем в словаре макс. фрагмент, совп. с началом буфера
3) записываем код (i, s):
   - i – номер слова в словаре, с ктр совп. нач. буфера
   - s – перв. символ, след. за фрагментом

<u>Х-ки:</u> степень сжатия: 2-3
тип данных: любые, но лучше однородные
сим-ть: 3:2

- LZW – модиф. LZ'78

6. Введение в теорию кодирования

6.1. Дв. симметр. канал связи



$0 < p \ll \frac{1}{2}$

$u = (u_1 \dots u_k)$ $\xrightarrow[n>k]{}$ $x = (x_1 \dots x_n)$ $\xrightarrow{\text{канал связи}}$ $x' = (x_1' \dots x_n')$ $\xrightarrow{\text{дек-е}}$ $u' = (u_1' \dots u_k')$

инф. блок — кодов. слово

$n \cdot p \cdot (1-p)^{n-1}$ – вер-ть того, что при передаче по каналу связи произв. дв. вект. длины n в нём произ. ровно 1 ошибка.

$C_n^2 p^2 (1-p)^{n-2}$ – —//— 2 ошибки

Пусть мы используем такое код-е, что ∀ число ошибок $\leq t$ м/б скорректировано при декодир-ии

Тогда опр-м вер-ть ошибки при дек-ии:

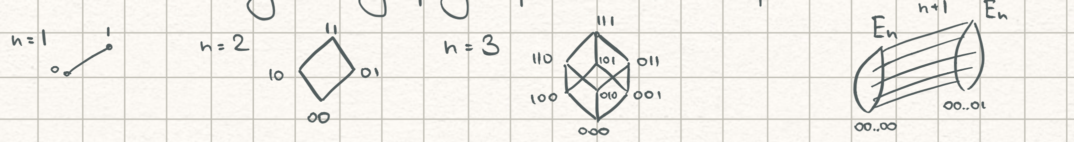
$$\boxed{P_{err} = 1 - \sum_{i=0}^{t} C_n^i p^i (1-p)^{n-i}}$$ кода, исправл. t ошибок

6.2. Булев куб. Метрика Хэмминга.

Опр E^n - булев куб разм-ти n — граф спец. вида:

$V = \{x = (x_1, \dots x_n) \mid x_i \in \mathbb{Z}_2\}$ $\quad |V| = 2^n$

$E = \{(x,y) \mid x$ и y разл. ровно в 1 коор-те$\}$ $\quad |E| = n \cdot 2^{n-1}$



Опр Носитель вектора — $supp(x) = \{i \mid x_i = 1\}$

Вес — $wt(x)$ - число единиц в векторе $= |supp(x)|$

kй слой булевого куба — все векторы веса k

Опр $x \oplus y = (x_1 \oplus y_1, \dots, x_n \oplus y_n)$, где

$\oplus$	0	1
0	0	1
1	1	0

(xor)

Опр Расстояние Хэмминга: $d(x,y)$ — число коор-т, в ктр они различ-ся

$$\boxed{d(x,y) = wt(x \oplus y)}$$

Ex $\begin{array}{c c} x & 01101 \\ y & 11001 \end{array}$ $\quad x \oplus y = 10100 \quad d(x,y) = 2$

Опр Сфера — $S_r(x) = \{y \mid d(x,y) = r\}$ $\quad |S_r(x)| = C_n^r$

Шар — $B_r(x) = \{y \mid d(x,y) \leq r\}$ $\quad |B_r(x)| = \sum_{i=0}^{r} C_n^i$

Опр Скал. произв-е вект. x, y — это число $\langle x,y \rangle = x_1 y_1 \oplus \dots \oplus x_n y_n$ (аналог только)

Опр Линейное подпр-во в булевом кубе — подмн-во булева куба, замкн. отн-но слож-я векторов

Любое лин. подпр-во в E^n имеет мощн-ть 2^k для подход. k, где k — размерность лин. подпр-ва

$k=0 \quad L = \{0\} \qquad k=1 \quad L_x = \{0, x\} \qquad k=2 \quad L = \{0, x, y, x \oplus y\}$

Опр Ортогон. подпр-во $L^{\perp} = \{y \in E^n \mid \langle x,y \rangle = 0 \quad \forall x \in L\}$

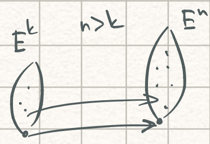
Св-во: $\dim L + \dim L^{\perp} = n$

6.3. Линейные коды

Опр Код — любое подмн-во бул. куба E^n

Опр Лин. код — любое лин. подпр-во бул. куба E^n

$\underset{\text{инф. блок}}{u = (u_1 \dots u_k)} \longrightarrow \underset{\text{код слово}}{x = (x_1 \dots x_n)}$



$C \subseteq E^n \quad |C| = 2^k$

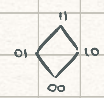
Если код линейный, то каждый бит x_i линейным образом зависит от $u_1 \dots u_k$

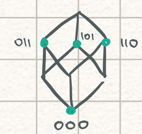
Ex $k=2 \quad n=3$

u_1	u_2		x_1	x_2	x_3
0	0	$\longrightarrow$	0	0	0
0	1		0	1	1
1	0		1	0	1
1	1		1	1	0

бит чётности

$x_1 = u_1$
$x_2 = u_2$
$x_3 = u_1 \oplus u_2$





(01)(11)(00)
↓
(011)(110)(000)
↓
(011)(100)(000)

ок ↓ (01) не ок :(↓ ок (00)

000 110 101 если 1 ош.

отбросим, т.к. наим. вероятн. → 011

Код только обнаруж. ошибку, но не исправляет

Ex $k=2$, $n=5$

$x_1 = u_1$
$x_2 = u_2$
$x_3 = u_1$
$x_4 = u_1 \oplus u_2$
$x_5 = u_2$

00	000
01	011
10	110
11	101

$u \to x \xrightarrow{\text{к.св.}} x' = 01111$
↓
01011
↓
01

ищем наиб. вер. сит.

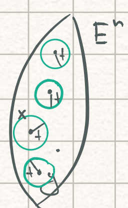
$u \longrightarrow x \xrightarrow{\text{к.св.}} x' = 11110$
↓
10110
↓
10

В этом коде всегда ошибку исправим

Параметры лин. кода: $[n, k, d]$

- n - длина кода (разм-ть пр-ва, в ктр рассм. код)
- k - размерн-ть кода
- d - min расст-е Хэмминга м/у код. словами

Для Ex1: $[3, 2, 2]$ Для Ex2: $[5, 2, 3]$



$x \in C$

$t = \lfloor \frac{d-1}{2} \rfloor$ — тогда ошибки испр-ся

Опр Порождающая м-ца лин. кода G $k \times n$, её строки — базис лин. кода

$\boxed{u \cdot G = x}$ $\forall u \in E^k$ — генерация кода

Опр Проверочная м-ца кода H $(n-k)\times n$, её строки — базис ортог. кода.

$$\boxed{Hx^T = 0 \iff x \in L}$$