Obsah 1. kapitoly (pojmy, které by měly být zvládnuty)

Obsah – princip použití samoopr. kódů – schema

- redundance (př. opakovací kód)
- abeceda kódu F_q, příklad q-ární kód (anglická slova)
- blokový kód
- $(F_q)^n$ – množina všech vektorů (slov) délky n nad abecedou F_q, z q symbolů.
- příklady kódů – 10-ciferné telefonní čísla v UK
  - $C_1 - (2,4,1)$, $C_2 - (3,4,2)$ ← detekuje 1 chybu
  $C_3 - (5,4,3)$ ← opraví 1 chybu
- Hammingova vzdálenost
  $d(\bar{x},\bar{y})$ = počet míst na kterých se $\bar{x}$ liší od $\bar{y}$

← dekódování na nejbližšího souseda – maximalizuje pravděpodobnost opravy chyb za předpokladu symetrického q-árního komunikačního kanálu.

p – pravděpodobnost chyby (pravděp, že dojde k chybě v symbolu) 1-p – pravděp. že symbol bude doručen bez chyby

$P_{err}(C)$ – word error probability = pravděpod. chyby kódu
– (Pravděp, že přijaté slovo bude chybně dekódováno, tj. v odeslaném kódovém slově dojde k alespoň $\left\lfloor \frac{d-1}{2} \right\rfloor + 1$ chybám)

- minimální vzdálenost kódu $d(C)$

$$d(C) = \min \{ dist(\bar{x},\bar{y}) \mid \bar{x},\bar{y} \in C, \bar{x} \neq \bar{y} \}$$

Věta 1.2.

(i) C detekuje až t chyb jestliže $d(C) \geq t+1$

(ii) C opraví až t chyb jestliže $d(C) \geq 2t+1$

Důkaz: (i) přímo, Předpokládejme $d(C) = t+1$ a že $\bar{x}$ je odeslané slovo. Dojde-li při přenosu k t nebo méně chybám, nemůže přijaté slovo $\bar{x}$ být jiným kódovým slovem.

$dist(\bar{u},\bar{x}) \leq t < t+1 \leq dis(\bar{x},\bar{x}') \quad \forall \bar{x}' \in C, \bar{x}' \neq \bar{x}$.

To znamená, že lze detekovat až t chyb.

(ii) přímo. Předpokládejme, že $d(C) = 2t+1$ a $\bar{x}$ je vysílané slovo, ve kterém došlo k nejvýše t chybám. Přijato bylo slovo $\bar{u}$ takové, že $dist(\bar{u}, \bar{x}) \leq t$. Pak musí být $dist(\bar{u}, \bar{x}') \geq t+1$ pro každé kódové slovo $\bar{x}' \in C$, $\bar{x}' \neq \bar{x}$. Kdyby platilo, že $dist(\bar{u}, \bar{x}') \leq t$, tak z trojúhelníkové nerovnosti dostaneme $dist(\bar{x}, \bar{x}') \leq dist(\bar{x}, \bar{u}) + dist(\bar{u}, \bar{x}') \leq t + t = 2t < 2t+1$ a to je spor s min vzdál. kódu $D(C) = 2t+1$. Proto vysílané slovo $\bar{x}$ je nejblíže přijatému slovu $\bar{u}$ a kód C opraví až t chyb.

Důsledek

Pro kód C takový, že $d(C) = d$ platí

(i) C detekuje až $d-1$ chyb

(ii) C opraví až $\left\lfloor \frac{d-1}{2} \right\rfloor$ chyb.

Důkaz: $d \geq t+1 \iff t \leq d-1$ t - počet detek. chyb

$d \geq 2t+1 \iff t \leq \left\lfloor \frac{d-1}{2} \right\rfloor$ t - počet opravených chyb

Příklady řešené ve cvičení

(1) Předpokládejme, že binární opakovací kód délky 5

(i) je použit pro komunikaci skrze binární symetrický kanál s pravděpodobností chyby v symbolu p.

Ukažte, že "Word error probability" (pravděpodobnost, že kódové slovo bude špatně dekódováno) je:

$$P_{err}(C) = 10p^3 - 15p^4 + 6p^5$$

$(F_q)^n \sim (F_2)^5$ — množina všech binárních vektorů délky 5

$|(F_2)^5| = 2^5 = 32$ $\quad C = \{ 00000, 11111 \}$ $\quad (5,2,5)$-kód

předpokládáme binární symetrický kanál $\quad p$ – symbol error probability (pravděp., že symbol bude doručen s chybou)

nearest neighbour decoding ~ maximum likelihood decoding

dekódování na nejbližšího souseda odpovídá dekód. na maximálně pravděpodobné slovo

Word error probability ~ $P_{err}(C)$ - Pravděp., že přijaté slovo bude špatně dekódováno.

Odeslána mohou být slova $\overset{s_1}{00000}$ nebo $\overset{s_2}{11111}$

Předpoklad odeslání s_1

- pravděp, že bude přijato slovo, které bude dekódováno na s_1 je

$$(1-p)^5 + 5(1-p)^4 \cdot p + \binom{5}{2}(1-p)^3 \cdot p^2 =$$

$$= 1 - 5p + 10p^2 - 10p^3 + 5p^4 - p^5 + 5p - 20p^2 + 30p^3 - 20p^4 + 5p^5$$
$$+ 10p^2 - 30p^3 + 30p^4 - 10p^5 = 1 - 6p^5 + 15p^4 - 10p^3$$

- pravděp, že bude přijato slovo, které bude dekódováno na jiné než bylo odesláno.

$$P_{err}(s_1) = 1 - (1 - 6p^5 + 15p^4 - 10p^3) = \underline{6p^5 - 15p^4 + 10p^3}$$

Předpokl. odeslání $s_2 \Rightarrow P_{err}(s_2) = P_{err}(s_1) = P_{err}(C)$

(ii) předpokládejme, že $p = 0{,}01$ určete $P_{err}(C)$

$$P_{err}(C) = 6 \cdot 0{,}01^5 - 15 \cdot 0{,}01^4 + 10 \cdot 0{,}01^3 = 0{,}0000985 \approx 10^{-5}$$

Tj. asi jedno ze 100 000 přijatých slov bude dekódováno špatně (zbylo ze 100 000 kód. slov nebude přijato správně.)

(2)

(i) Ukažte, že ternární $(3, M, 2)$-kód musí mít $M \leq 9$

(ii) Ukažte, že ternární $(3, 9, 2)$-kód existuje.

(iii) Zobecněte výsledek (i) a (ii) na lib. q-ární $(3, M, 2)$-kód, pro $q \geq 2$, $q \in \mathbb{Z}$.

ternární kód (tři symboly)

$(3, M, 2)$-kód délka $n = 3$ $\quad |(F_3)^3| = 3^3 = 27$

$M = ?$

$d(C) = 2$

(i) Jestliže vypustíme ze všech slov kódu poslední (nebo v pořadí pevný kterýkoliv) symbol, zůstane M uspořádaných párů, které nutně musí být navzájem různé. Kdyby některé dva páry byly stejné, znamenalo by to, že se původní slova lišila jen na posledním místě, tj. platilo by, že $d(C) \leq 1$.
Počet různých párů, které lze utvořit ze 3 symbolů je 3^2.
Proto $\underline{M \leq 3^2 = 9}$

(ii) Sestrojením kódu - přidáme čtvrtý symbol který je kontrolní součet (mod 3))

000 101 202
011 112 210 dostaneme $(3, 9, 2)$-kód
022 120 221

(iii) Lze ukázat, že $\{(a, b, a+b) \mid a, b \in F_q^2\}$, kde $F_q = \{0, 1, \dots, q-1\}$ a $a + b = a +_{\text{mod } q} b$ (počítá se modulo q) je q-ární $(3, q^2, 2)$-kód.

Zřejmě je $n = 3$, $M = q^2$. Je třeba ukázat, že $d(C) > 1$.

Sporem: Předpokl, že $d(x, x') = 1$, kde

$x = (a, b, a+b)$ a $x' = (a', b', a'+b')$

musí být ① $a = a' \wedge b \neq b'$ nebo ② $a \neq a' \wedge b = b'$ a $a + b \equiv a' + b' \pmod q$

aby $d(x, x') = 1$ případ ①

pokud $a = a' \wedge b \neq b' \wedge a + b \equiv a' + b' \pmod q$

pak $b \equiv b' \pmod q$

a pro $0 \leq b, b' < q \Rightarrow b = b'$

případ ② - analogický argument což je spor s $d(x, x') = 1$

Kód ekvivalentní s C_3 z přednášky

Dva různé důkazy, že pro $(5,M,3)$-kód je $M \leq 4$. (Druhý viz další strana)

(3) Ukažte, že pro $(5,M,3)$-kód platí, že $M \leq 4$.

Předpokládejme, že každý kód je "ekvivalentní" kódu, který obsahuje nulové slovo $\bar{x} = (0,0,0,0,0) = (00000) = 00000$

jiné možné značení

Dále víme, že existuje $(5,4,3)$-kód. } $\Rightarrow M \geq 4$

Např. kód robota $C_R = \begin{Bmatrix} 00000 \\ 01101 \\ 10110 \\ 11011 \end{Bmatrix}$

Je třeba ukázat, že $M < 5$.

- Jestliže $\bar{x}_1 = (0,0,0,0) \in C$ pak C nemůže obsahovat slova s jednou nebo dvěma jedničkama. ($d(\bar{x},\bar{y})$ musí být ≥ 3)
- C nemůže obsahovat $x = (1,1,1,1,1)$ – nešlo by přidat žádné další slovo
- C může obsahovat slova s 3-ma nebo 4-ma jedničkama. (váhy 3 a 4)
- Předpokl. že $\bar{x}_2 = (1,1,1,1,0) \in C \Rightarrow$ žádné další slovo $\bar{y}$ váhy 4 nemůže patřit do C, nastalo by $d(\bar{x}_2,\bar{y}) \leq 2$
- Lze přidat slova váhy 3 – např. $11110 = \bar{x}_2$, $11001 = \bar{x}_3$, $00111 = \bar{x}_4$
- C nemůže obsahovat třetí slovo váhy 3 ← Proč?

Pro 3 slova váhy 3 nelze zařídit $d(C) \geq 3$.

11100
10011
? ← nelze přidat tři 1 a dvě 0
00111 aby $d(C) = 3$

(4) Ukažte, že neexistuje binární kód délky $n = 4$ s minimální vzdáleností $d = 3$ o velikosti $M = 4$. (Neexistuje $(4, 4, 3)$-kód)

1. Můžeme začít slovem 0000. Permutací symbolů na dané pozici kódu dostaneme vždy ekvivalentní kód. Tj. každý kód je ekvivalentní s kódem obsahujícím nulové slovo.
2. Každé další slovo musí obsahovat alespoň 3 jedničky, aby min. vzdálenost od 0000 byla alespoň 3.
   Můžeme vzít slovo 1110. (Permutací pozic kódu dostaneme ekviv. kód)
   Další slovo s 3-mi jedničkami přidat nelze. Lišilo by se na nejvýše 2 pozicích od 1110.
3. Z možných 16 binárních sekvencí délky 4 zbývá jen jedno 1111 a to také přidat nelze, jinak $d \neq 3$.

(3) - znovu trochu jinak
Ukažte, že pro $(5, M, 3)$-kód platí, že $M \leq 4$.

Označme C $(5, M, 3)$-kód

1.) Předpokládejme, že $\underline{00000} \in C$

2.) Další slova musí obs. alespoň 3x jedničku.

Začneme slovem $\underline{11100}$.

Pak lze přidat jedno ze slov $\longrightarrow$ 00111, 01011, 10011 například $\underline{00111}$

3.) Slov se 4-ma jedničkama je právě 5 a navzájem se liší na 2 pozicích. Tj. pro kód C lze využít jen jedno z nich a musí to být právě slovo $\underline{11011}$, aby se lišilo od vybraných slov na 3 pozicích

4.) Slovo 11111 přidat nelze.

Odtud $M(C) \leq 4$.