

Samoopravné kódy

Alexandr Kazda

Univerzita Karlova

6. května 2020

Problem

Odhadněte pomocí Singletonova a pomocí Hammingova odhadu (viz začátek semestru), jak velké musí být n , aby existoval binární

- a) $(n, 100, 3)$ -kód,
- b) $(n, 100, 5)$ -kód.

Co z toho plyne pro hustotu takových kódů?

- Singleton: $100 \leq n - 3 + 1$ resp. $100 \leq n - 5 + 1$
- Tedy $n \geq 102$ resp. $n \geq 104$

Hammingův odhad pro $k = 100$

- Hamming pro $d = 3$: $V_{1,n} = 1 + n$
- $100 + \log_2(1 + n) \leq n$
- $\log_2(1 + n) \in (6, 7]$
- Nejmenší celočíselné řešení $n = 107$

Hammingův odhad pro $k = 100$

- Hamming pro $d = 5$: $V_{2,n} = 1 + n + \binom{n}{2}$
- $100 + \log_2(1 + (n+1)n/2) \leq n$
- $\log_2(1 + (n+1)n/2) \geq \log_2(n^2/2) = 2\log_2(n) - 1$
- $99 + 2\log_2(n) \leq n$
- Nejmenší n je zhruba 113
- Numerický výpočet pro přesné hodnoty to potvrdí ($n = 112$ nejde, $n = 113$ jde)

Co to znamená pro hustotu?

- Chybovost $p = 0,01$, $n \geq 100$, tedy aspoň jedna chyba nastává často
- Pokud chceme kódovat s jistotou opravy 1 chyby, potřebujeme hustotu kódu nejvýš $100/107 \doteq 0,93$
- Pokud chceme kódovat s jistotou opravy 2 chyb, potřebujeme hustotu kódu nejvýš $100/113 \doteq 0,88$
- Lze těchto hodnot dosáhnout?

Zkrácení kódu bez zmenšení d

- Kód H_7 je $[127, 120, 3]$ -kód, ale je moc dlouhý...
- Měli jsme propíchnutí kódu, typicky $(n, k, d) \rightarrow (n - 1, k, d - 1)$
- Ale d je „drahé“, co kdybychom raději zmenšili k ?
- Buď C lineární systematický $[n, k, d]$ -kód
- Kódujeme $0, w_1, \dots, w_{k-1}$ na kódové slovo $0c_2c_3\dots$
- První informační symbol je vždy 0 – zahodíme ho
- Dostaneme $[n - 1, k - 1, d']$ kód a $d' \geq d$

- Vezmeme H_7 , kódujme systematicky
- Opakovaně zmenšujeme n, k :

$$[127, 120, 3] \rightarrow [126, 119, 3] \rightarrow [125, 118, 3] \rightarrow \cdots \rightarrow [107, 100, 3]$$

- Dosáhli jsme meze z Hammingova odhadu!

Použití pro $d = 5$

- Dosáhneme $n = 114$ (o jedna horší než Hammingův odhad)
- Potřebujeme $[127, 113, 5]$ -kód
- Uvažme binární kód $\text{BCH}_{2,7,5}$
- Tvrdíme, že je to $[127, 113, 5]$ -kód
- Kódová slova jsou polynomy p stupně ≤ 126 z $\mathbb{F}_2[x]$, které mají kořeny $\alpha, \alpha^2, \alpha^3, \alpha^4$ pro α generátor \mathbb{F}_{128}^*
- $n = 127$ jasné, $d \geq 5$ bylo na přednášce, $k = ?$

- Připomínám, že dimenze \mathbb{F}_{128} nad \mathbb{F}_2 je 7
- Na přednášce jsme měli $k \geq 127 - 7 \cdot 4 = 99$; to nestačí
- $x \mapsto x^2$ je automorfismus \mathbb{F}_{128} , který fixuje \mathbb{F}_2 , tedy $p(\alpha) = 0 \Rightarrow p(\alpha^2), p(\alpha^4) = 0$
- Tedy ekvivalentní podmínky na kódový polynom p : $\deg p \leq 126$, $p(\alpha) = 0, p(\alpha^3) = 0$
- To je v \mathbb{Z}_2 soustava 7 + 7 rovnic pro 127 neznámých
- Proto $k \geq 127 - 14 = 113$ (počítání generujícího polynomu ukáže $k = 113$)

Dokončení konstrukce $[114, 100, 5]$ -kódu

- Máme $[127, 113, 5]$ -kód
- Dosazováním a zapomínáním nul dostaneme

$$[127, 113, 5] \rightarrow [126, 112, 5] \rightarrow \cdots \rightarrow [114, 100, 5]$$

- Spolehlivost aspoň 0,89, hustota $100/114 \doteq 0,877$

Věta

Nechť $p \in (0, 1/2)$. Mějme kanál s chybovostí p a kapacitou $C(p) = 1 - H(p)$. Nechť $\kappa < C(p)$. Potom pro každé $\varepsilon > 0$ existuje binární kód s hustotou k/n aspoň κ a spolehlivostí $> 1 - \varepsilon$.

- Asymptotický, nekonstruktivní výsledek
- Šlo by to obrátit?
- Chceme něco jako: „Pokud C je dlouhý a spolehlivý pro kanál kapacity R , tak C musí mít hustotu nejvýš R “.

Věta

Bud'te $\varepsilon, p > 0$ a $R > 1 - H(p)$. Pak existuje N takové, že každý kód hustoty aspoň R a délky aspoň N má spolehlivost nejvýš ε při chybovosti kanálu p .

- Věta nic neříká o krátkých kódech
- Ale krátký kód opakovaný hodněkrát je dlouhý kód. . .
- Srovnejte s naší „větou“: Kód spolehlivosti 1 musí mít hustotu nejvýše $1 - H(p)$
- Dohromady se Shannonem to říká, že kapacita kanálu určuje (pro dlouhé zprávy) hustotu (rychlost) posílání zpráv skrz kanál

Idea důkazu (nepřesná)

Věta

Bud'te $\varepsilon, p > 0$ a $R > 1 - H(p)$. Pak existuje N takové, že každý kód hustoty aspoň R a délky aspoň N má spolehlivost nejvýš ε při chybovosti kanálu p .

- Typicky udělá kanál při přenosu asi pn chyb
- Pro fixní odeslané slovo je pravděpodobnost každého přijatého slova s pn chybami asi $p^{pn}(1-p)^{n-pn} = 2^{n(p \log(p) + (1-p) \log(1-p))} = 2^{-nH(p)}$
- Bud' $D^{-1}(c)$ množina slov \tilde{c} , co se dekódují na c
- Pak spolehlivost kódu je asi

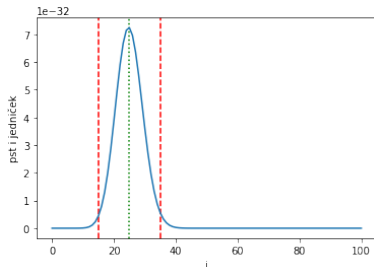
$$\frac{1}{|C|} \sum_{c \in C} |D^{-1}(c)| 2^{-nH(p)} = 2^{-Rn - nH(p)} \sum_{c \in C} |D^{-1}c| \leq 2^{n(-R - H(p) + 1)}$$

- Pro $R > 1 - H(p)$ jde pravá strana k nule

S kolika chybami počítat II

Věta

Bud' $p \in [0, 1], n \in \mathbb{N}, \alpha > 0$. Necht' $z_1, \dots, z_n \in \{0, 1\}$ jsou volené nezávisle tak, že $\text{pst } z_i = 1$ je p . Pak $\sum_{i=1}^n z_i$ leží v intervalu $[n(p - \alpha), n(p + \alpha)]$ s pravděpodobností aspoň $1 - 2e^{-n\alpha^2/2}$.



- Obrázek pro $n = 100, p = 1/4, \alpha = 1/10$
- Pro velká n to plyne ze zákona velkých čísel
- I toto je důsledek Černovovy nerovnosti (Kaiser Věta 11.2.1)

Věta

Bud'te $\varepsilon, p > 0$ a $R > 1 - H(p)$. Pak existuje N takové, že každý kód hustoty aspoň R a délky aspoň N má spolehlivost nejvýš ε při chybovosti kanálu p .

- Důkaz je podrobnější verze důkazu pro $p = 1/2$, co jsme měli
- Sporem. Zvolme ε . Búno $p \in (0, 1/2]$
- Nechť C je kód, D jeho dekódovací funkce
- Ať C má velkou délku n , hustotu $R > 1 - H(p)$, spolehlivost aspoň $\beta > \varepsilon$
- S pravděpodobností aspoň $1 - 2e^{-n\alpha^2/2} > 1 - \varepsilon/2$ leží počet chyb v $[n(p - \alpha), n(p + \alpha)]$

Slova s $n(p - \alpha)$ až $n(p + \alpha)$ chybami

- Fixujme $c \in C$ a značme $S(c) \subset \{0, 1\}^n$ všechny body vzdálené od c aspoň $n(p - \alpha)$ a nejvýš $n(p + \alpha)$
- Jaká je pravděpodobnost, že $c + e = \tilde{c}$ pro $\tilde{c} \in S(c)$?
- Je to $p^{d(c, \tilde{c})}(1 - p)^{n - d(c, \tilde{c})}$
- $p \leq 1/2$, tedy je to nerostoucí funkce $d(c, \tilde{c})$
- Pro $\tilde{c} \in S(c)$ máme $d(c, \tilde{c}) \geq n(p - \alpha)$, tedy

$$\begin{aligned} P[\tilde{c} = c + e | \text{odesláno } c] &\leq p^{n(p - \alpha)}(1 - p)^{n - n(p - \alpha)} \\ &\leq p^{np}(1 - p)^{n(1 - p)}(1 - p)^{n\alpha} p^{-n\alpha} \\ &\leq 2^{np \log p + n(p - 1) \log(1 - p)} \left(\frac{1 - p}{p} \right)^{n\alpha} \\ &\leq 2^{-nH(p) + n\alpha(\log(1 - p) - \log(p))} \end{aligned}$$

Přechod k případu X

- Značme X jev „počet chyb je v $[n(p - \alpha), n(p + \alpha)]$ “
- Chceme uvažovat jenom případy, kdy nastane X

$$P[D(\tilde{c}) = c] = P[D(\tilde{c}) = c, X] + P[D(\tilde{c}) = c | X^c]P[X^c]$$

- Přitom $P[X^c] < \varepsilon/2$, $P[D(\tilde{c}) = c | X^c] \leq 1$
- Tedy odhadujeme

$$P[D(\tilde{c}) = c] < P[D(\tilde{c}) = c, X] + \varepsilon/2$$

- Konečně $P[D(\tilde{c}) = c] \geq \varepsilon$, takže

$$\varepsilon/2 \leq P[D(\tilde{c}) = c] - \varepsilon/2 < P[D(\tilde{c}) = c, X]$$

$P[D(\tilde{c}) = c, X] \geq \varepsilon/2$ vede ke sporu

- Značme Y_c jev „odeslané slovo je c “, $Z_{\tilde{c}}$ jev „přijaté slovo je \tilde{c} “
- Vyšlo nám

$$\varepsilon/2 \leq P[D(\tilde{c}) = c, X] = \frac{1}{|C|} \sum_{c \in C} P[D(\tilde{c}) = c, X | Y_c]$$

- Přitom X je sjednocení jevů $Z_{\tilde{c}}$ pro $\tilde{c} \in S(c)$, takže

$$\begin{aligned} P[D(\tilde{c}) = c, X | Y_c] &= \sum_{\tilde{c} \in S(c)} P[D(\tilde{c}) = c, Z_{\tilde{c}} | Y_c] = \sum_{\substack{\tilde{c} \in S(c) \\ D(\tilde{c}) = c}} P[Z_{\tilde{c}} | Y_c] \\ &\leq |S(c) \cap D^{-1}(c)| 2^{-nH(p) + n\alpha(\log(1-p) - \log(p))} \end{aligned}$$

$P[D(\tilde{c}) = c|X]P[X] \geq \varepsilon/2$ vede ke sporu II

- Dosadíme odhad pro $P[D(\tilde{c}) = c, X] \geq \varepsilon/2$:

$$P[D(\tilde{c}) = c, X] \leq \frac{1}{|C|} \sum_{c \in C} |S(c) \cap D^{-1}(c)| 2^{-nH(p) + n\alpha(\log(1-p) - \log(p))}$$

- Přitom $\sum_{c \in C} |S(c) \cap D^{-1}(c)| \leq \sum_{c \in C} |D^{-1}(c)| \leq 2^n$
- Tedy

$$\varepsilon/2 \leq \frac{1}{|C|} 2^n 2^{-nH(p) + n\alpha(\log(1-p) - \log(p))}$$

- Vzpomeňme si, že $|C| \geq 2^{Rn}$ pro $R > 1 - H(p)$
- Tedy

$$\varepsilon/2 \leq 2^{-Rn + n - nH(p) + n\alpha(\log(1-p) - \log(p))}$$

- Máme

$$\varepsilon/2 \leq 2^{-Rn+n-nH(p)+n\alpha(\log(1-p)-\log(p))}$$

- Vezměme dvojkový logaritmus obou stran:

$$\log(\varepsilon/2) \leq n(-R + 1 - H(p) + \alpha(\log(1-p) - \log(p)))$$

- Víme $R > 1 - H(p)$
- Volme $\alpha > 0$ tak, aby $-R + 1 - H(p) + \alpha(\log(1-p) - \log(p)) < 0$
- Pak pro velké n jde pravá strana k $-\infty$, což je méně než libovolné číslo $\log(\varepsilon/2)$
- Máme spor. . .

- Výsledek je opět asymptotický
- Řekněme, že bychom pro kanál s chybovostí $p = 0,01$ chtěli kód s hustotou 0,99 a spolehlivostí $\varepsilon = 1/2$
- Volím α , aby

$$-0,99 + 0,92 + \alpha(\log(0,99) - \log(0,01)) < 0$$

- Řekněme $\alpha = 0,015$
- Volím n , aby $e^{-n\alpha^2}/2 < 1/4$ a $\log(1/4) > n(-R + 1 - H(p) + \alpha(\log(1 - p) - \log(p)))$
- Pro $n \geq 3081$ kód neexistuje, pro kratší kódy nevíme

- Shannon a inverzní Shannon nám pro $p = 0,01$ a n velké dávají hustotu spolehlivého kódu asi 0,92
- Pro malé n lze najít všelijaké „protipříklady“
- Hammingovy kódy umí opravit jednu chybu, což zhruba stačí
- Kód H_7 má pro $p = 0,01$ spolehlivost asi 0,64 a hustotu $120/127 \doteq 0,94$
- Reed-Solomonův $[31, 29, 3]_{32}$ -kód přeložený do $\{0, 1\}$ je binární $[155, 145, 3]$ -kód s hustotou 0,93 a spolehlivostí asi 0,57
- Kód H_8 má hustotu asi 0,97, ale spolehlivost už jen asi 0,28