

Samoopravné kódy

Alexandr Kazda

Univerzita Karlova

20. dubna 2020

Pár upozornění

- Budeme počítat s pravděpodobností
- Pravděpodobnost je těžká – pokud vám něco není jasné, ptejte se nebo hledejte
- Na chvíli opustíme blokové kódy
- Klasická teorie informace ignoruje výpočetní složitost
- Zdroje: Skripta prof. Drápala, kapitola „Teorie informace“
- Učební text Andrewa Kozlíka
<http://www.karlin.mff.cuni.cz/~kozlik/ks/shannon.pdf>

a_1	sedni	$1/4$
a_2	lehni	$1/4$
a_3	dej pokoj	$1/4$
a_4	do boudy	$1/8$
a_5	aport	$1/16$
a_6	hledej	$1/32$
a_7	pac	$1/64$
a_8	vem si ho	$1/64$

- Srovnejte s rovnoměrně rozdělenými příkazy s pstí $1/8$
- Lze tipovat, že zdrojové písmeno je nejspíš jedno z a_1, a_2, a_3
- Znak „vem si ho“ je překvapivější než „sedni“ – má větší informační obsah

- Cíl: Přiřadit náhodnému jevu („něco se stane“) číslo I
- I je nějaká funkce pravděpodobnosti jevu
- Chtěli bychom: Pokud A, B jsou dva nezávislé jevy, tak
$$I(A \cap B) = I(A = a) + I(B = b)$$
- $P[A \cap B] = P[A = a]P[B = b]$
- Kdo převádí součin na součet? Logaritmus!
- Volme $I(A) = -\log_2(P[A]) = \log_2(1/P[A])$, aby to nebylo záporné

Proč to dává smysl – příklady

- Necht' nám z krabičky padají nuly a jedničky nezávisle, každá s postí 50%
- Značme i -tý znak, co jsme dostali, jako z_i
- Pak $I(z_1 = 0) = \log_2(2) = 1$ (1 bit)
- $I(z_1 = 0, z_2 = 1, z_3 = 1) = 1 + 1 + 1$ (3 bity)
- Pokud by z krabičky padaly samé jedničky, tak
- $I(z_1 = 1) = \log_2(1) = 0$; $I(z_1 = 0) = \log_2(1/0) = \infty$
- Pro nás $0 \cdot \infty = 0$

Informační obsahy povelů

a_1	sedni	$1/4$	2
a_2	lehni	$1/4$	2
a_3	dej pokoj	$1/4$	2
a_4	do boudy	$1/8$	3
a_5	aport	$1/16$	4
a_6	hledej	$1/32$	5
a_7	pac	$1/64$	6
a_8	vem si ho	$1/64$	6

- “sedni” má třikrát méně informačního obsahu než “vem si ho”

- Σ buď abeceda
- **Informační zdroj** je pro nás „vysílač“, který vysílá nekonečná slova nad abecedou Σ náhodně s nějakým rozdělením
- Námitka: Když telefonuju, tak neříkám náhodná slova!
- Odpověď: Z hlediska výrobce telefonu to náhodná slova jsou.
- Otázka: Jak vypadá pravděpodobnost nekonečných slov?
- Odpověď: Pro nás bude stačit uvažovat konečná slova – předpony toho nekonečného slova
- Příklad 1: Zdroj, kde znaky na různých pozicích jsou náhodné nezávislé, pst písmena 1 i 0 je $1/2$
- Příklad 2: Zdroj, kde znaky na různých pozicích jsou náhodné nezávislé, pst písmena 1 je 1

- Bud' \mathcal{A} informační zdroj nad abecedou $\Sigma = \{1, 2, \dots, k\}$
- Znaky na různých pozicích nezávislé stejně rozdělené s rozdělením (p_1, p_2, \dots, p_k)
- Kolik je průměrný informační obsah zdroje na jeden znak?
- Tolik:

$$H(\mathcal{A}) = \sum_{i=1}^k p_i I(A_1 = i) = \sum_{i=1}^k p_i \log_2(1/p_i)$$

- Toto je entropie (fyzici by tam dali \ln)

- Zdroj, kde znaky jsou náhodné, na různých pozicích nezávislé a pst písmena 1 i 0 je $1/2$ má entropii $1/2 \cdot 1 + 1/2 \cdot 1 = 1$
- Zdroj, kde znaky jsou samé 1, má entropii $1 \cdot 0 = 0$
- Zdroj příkazů Hafíkovi má entropii

$$1/4 \cdot 2 + 1/4 \cdot 2 + 1/4 \cdot 2 + 1/8 \cdot 3 + 1/16 \cdot 4 + 1/32 \cdot 5 + \\ + 1/64 \cdot 6 + 1/64 \cdot 6 = 79/32 \doteq 2,47$$

- Zdroj rovnoměrných příkazů Hafíkovi s pstí $1/8$ má entropii

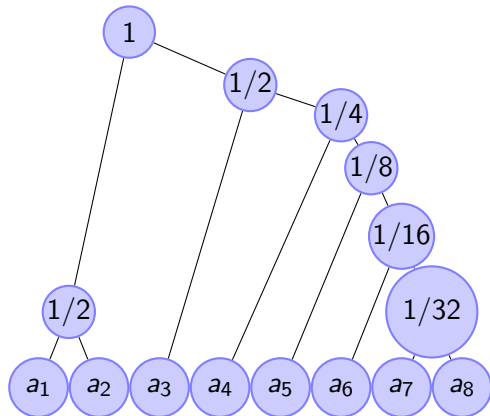
$$8 \cdot (1/8 \cdot 3) = 3$$

- n znaků z daného zdroje má informační obsah n -krát informační hodnota jednoho znaku
- Co když znaky nejsou nezávislé? Brali bychom entropii dlouhých slov
- Informační hodnota textu v přirozeném jazyce: Anglický text o délce n má informační obsah zhruba někde mezi $2n$ a $3n$
- Zdroj vysílající jeden z 27 rovnoměrně rozdělených (nezávislých) má entropii $\log_2(27)n \simeq 4,75n$

- Pokud různé zprávy mají různou pravděpodobnost, dává smysl časté zprávy kódovat kratším kódovým slovem
- Přirozený jazyk: „jíst“ versus „interferovat“
- Opouštíme blokové kódy
- Nový problém: Jak zjistit, kde kódové slovo končí?
- Máme zdrojové slovo nad abecedou Σ , kódovat budeme binárně
- Informační zdroj vysílá znaky nezávisle se známými pravděpodobnostmi možných písmen

Huffmanovo kódování – příklad

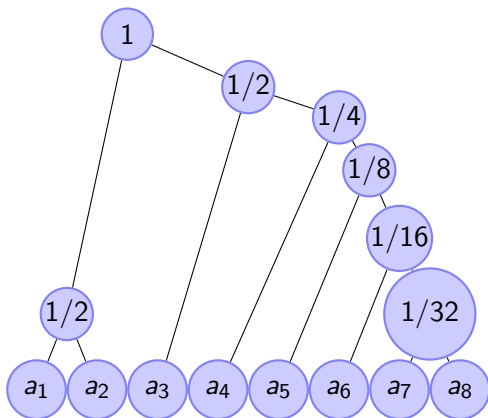
a_1	sedni	$1/4$
a_2	lehni	$1/4$
a_3	dej pokoj	$1/4$
a_4	do boudy	$1/8$
a_5	aport	$1/16$
a_6	hledej	$1/32$
a_7	pac	$1/64$
a_8	vem si ho	$1/64$



- Stavíme si strom, vždy spojíme kořeny s nejmenšími pravděpodobnostmi
- Strom nám dá kód

Huffmanovo kódování – příklad

a_1	sedni	$1/4$	11
a_2	lehni	$1/4$	10
a_3	dej pokoj	$1/4$	01
a_4	do boudy	$1/8$	001
a_5	aport	$1/16$	0001
a_6	hledej	$1/32$	00001
a_7	pac	$1/64$	000001
a_8	vem si ho	$1/64$	000000



- Strom nám dá kód (cestujeme od kořene k listům, vlevo je 1, vpravo je 0)
- Průměrná délka kódu je

$$\begin{aligned} & 1/4 \cdot 2 + 1/4 \cdot 2 + 1/4 \cdot 2 + 1/8 \cdot 3 + 1/16 \cdot 4 + 1/32 \cdot 5 + 1/32 \cdot 5 + \\ & + 1/64 \cdot 6 + 1/64 \cdot 6 = H(\mathcal{A}) \end{aligned}$$