20 Entropie, vzájemná entropie a podmíněná entropie diskrétních a spojitých rozdělení, základní vlastnosti a význam. Kódování zpráv, Kraftova-MacMillanova nerovnost. Souvislost entropie a střední délky kódového slova. Kódy s optimální střední délkou. Informační kanál a jeho kapacita. Shannonova věta o kódování. (A0B01PSI)

20.1 Entropie

20.1.1 Entropie diskrétního rozdělení

Za předpokladu, že X je diskrétní náhodná veličina s pravděpodobnostní funkcí p_X, je entropie diskrétní náhodné veličiny X rovna:

$$H(X) = -\sum_{x \in X} p_x(x) \log_2 p_x(x). \qquad (1.1)$$

Funkce H se nazývá Shannonova entropie nebo pouze entropie.

Příspěvek výsledku s pravděpodobností p je dán funkcí $\iota(p) = -p \log_2 p$. Ta je nezáporná a má nulové limity v 0 a 1. Důsledkem je, že entropie *diskrétního* náhodného pokusu je vždy nezáporná.

20.1.2 Vzájemná entropie diskrétního rozdělení

Dána vzorcem

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p_{XY}(x,y) \log_2 p_{XY}(x,y).$$

Platí nerovnost

$$H(X,Y) \leq H(X) + H(Y).$$

20.1.3 Podmíněná entropie diskrétního rozdělení

Dána vzorcem

$$H(Y \mid X = x) = -\sum_{y \in Y} p_{Y|X}(y|x) \log_2 p_{Y|X}(y|x).$$

20.1.4 Entropie spojitého rozdělení

Entropie spojité náhodné veličiny X s hustotou pravděpodobnosti $f(x)$ je definována takto:

$$H(X) = \int_{-\infty}^{\infty} f(x) \log_2 f(x) dx.$$

Zatímco entropie diskrétního veličiny je **absolutní** mírou neurčitosti, ve spojité verzi je entropie **relativní** mírou neurčitosti vzhledem ke zvolenému sytému souřadnic.

20.1.5 Vzájemná entropie spojitého rozdělení

Dána vzorcem

$$H(X,Y) = -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \log_2 f(x,y) dx dy.$$

20.1.6 Podmíněná entropie spojitého rozdělení

Dány vzorci

$$H(Y \mid X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \log_2 \frac{f(x,y)}{g(x)} dx dy$$

a

$$H(X \mid Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \log_2 \frac{f(x,y)}{h(x)} dx dy,$$

kde $f(x,y)$ je sdružená hustota pravděpodobnosti veličin x a y a *g(x), h(x)* jsou marginální hustoty pravděpodobnosti veličin x a y, pro které platí:

$$g(x) = \int_{-\infty}^{\infty} f(x,y) dy$$

a

$$h(y) = \int_{-\infty}^{\infty} f(x,y)dx.$$

20.1.7 Význam a vlastnosti entropie

Entropie je střední hodnota informace, kterou dostaneme, pokud se dovíme, který z k disjunktnch jevů nastal, jestliže jejich pravděpodobnosti byly p_1,..., p_k.

Shannonova entropie H:$D_h \to \langle 0, \infty\rangle$daná vzorcem (1.1) je jediná reálná funkce na D_h následujících vlastností:

1. H nezávisí na permutaci argumentů,
2. funkce $p \mapsto H(p, 1-p)$ je spojitá,
3. $H(1/2, 1/2) = 1$,
4. $H(p_1, p_2, p_3, ..., p_n) = H(p_1 + p_2, p_3, ..., p_n) + (p_1 + p_2)H(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$.

Entropie $H(X) = 0$ tehdy a jen tehdy, jsou-li všechy pravděpodobnosti kromě jedné rovny nule a jedna pravděpodobnost rovna jedné. Entropie dosahuje svého maxima, jsou li všechny pravděpodobnostu stejné.

20.2 Kódování zpráv

Definice: Nechť X je náhodná veličina s výběrovým prostorem $\mathcal{X}$. Nechť $\mathcal{D}$ je konečná abeceda a $\mathcal{D}^\star = \bigcup_{n=1}^{\infty} \mathcal{D}^n$. **Kód** pro X je zobrazení $\mathcal{C} : \mathcal{X} \to \mathcal{D}^*$.

Definice: **Střední délka** kódu C je $L(C) = \sum_{x\in X} p_X(x)l(C(x))$, kde $l(C(x))$ značí délku řetězce $C(x)$.

Příklad 1: Nechť $p_X(i) = \frac{1}{3}$, $i = 1, 2, 3$. Mějme tento binární kód:
$C(1) = 0$, $C(2) = 10$, $C(3) = 11$.
Zřejmě $L(C) = \frac{5}{3} = 1.\bar{6}$, přičemž $H(X) = \log 3 \doteq 1.58$.

Příklad 2: Nechť $p_X(i) = 2^{-i}$, $i = 1, 2, 3$ a $p_X(4) = 2^{-3}$. Mějme tento binární kód:
$C(1) = 0$, $C(2) = 10$, $C(3) = 110$, $C(4) = 111$.
Zřejmě $L(C) = H(X) = 1.75$.

20.2.1 Třídy kódů

Definice: Kód C je

- **nesingulární**, pokud je $\mathcal{C} : \mathcal{X} \to \mathcal{D}^*$prosté zobrazení.
- **jednoznačně dekódovatelný**, pokud je jeho rozšíření $\mathcal{C}^*$nesingulární, kde $\mathcal{C}^* : \mathcal{X}^* \to \mathcal{D}^*$ je definováno pomocí $\mathcal{C}^*(x_1...x_n) = \mathcal{C}(x_1)...\mathcal{C}(x_n)$, $x_1...x_n \in \mathcal{X}^*$.

- **instantní**, pokud žádné kódové slovo $\mathcal{C}(x)$ není počátečním úsekem kódového slova $\mathcal{C}(x')$ pro $x, x' \in \mathcal{X},\ x \neq x'$.

Vztahy mezi kódy:

1. Každý **instantní kód** je **jednoznačně dekódovatelný**.
2. Každý **jednoznačně dekódovatelý kód** je **nesingulární**.

	x	*singulární*	*nesingulární*	*jednoznačně d.*	*instantní*
	1	0	0	01	0
Třídy kódů	2	0	1	00	10
	3	0	00	11	110
	4	0	01	110	111

- neinstantní jednoznačně dekódovatelný kód obecně umožňuje dekódovaní až po přečtení **celého** rozšířeného kódového slova $C(x_1...x_n)$
- instantní kód umožňuje dekódování **ihned** po obdržení kódového slova

20.3 Kraftova-MacMillanova nerovnost

Věta (Kraft, 1949) Délky slov $l_1, ..., l_m$ libovolného **instantního** d-znakového kódu splňují nerovnost

$$\sum_{i=1}^{m} d^{-l_i} \leq 1.$$

Obráceně, splňují-li $l_1, ..., l_m \in \mathbb{N}$ tuto nerovnost, potom existuje **instantní** d-znakový kód s délkami slov $l_1, ..., l_m$.

Věta (MacMillan, 1956) Délky slov $l_1, ..., l_m$ libovolného **jednoznačně dekódovatelného** d-znakového kódu splňují nerovnost

$$\sum_{i=1}^{m} d^{-l_i} \leq 1.$$

Obráceně, splňují-li $l_1, ..., l_m \in \mathbb{N}$ tuto nerovnost, potom existuje **jednoznačně dekódovatelného** d-znakový kód s délkami slov $l_1, ..., l_m$.

Kombinací těchto dvou nerovností dostaneme:

Věta (Kraft-MacMillanova) Jednoznačně dekódovatelné kódování s předepsanou délkou slov existuje právě tehdy, když existuje instantní kó se stejnou délkou.

20.4 Souvislost entropie a střední délky kódovaného slova

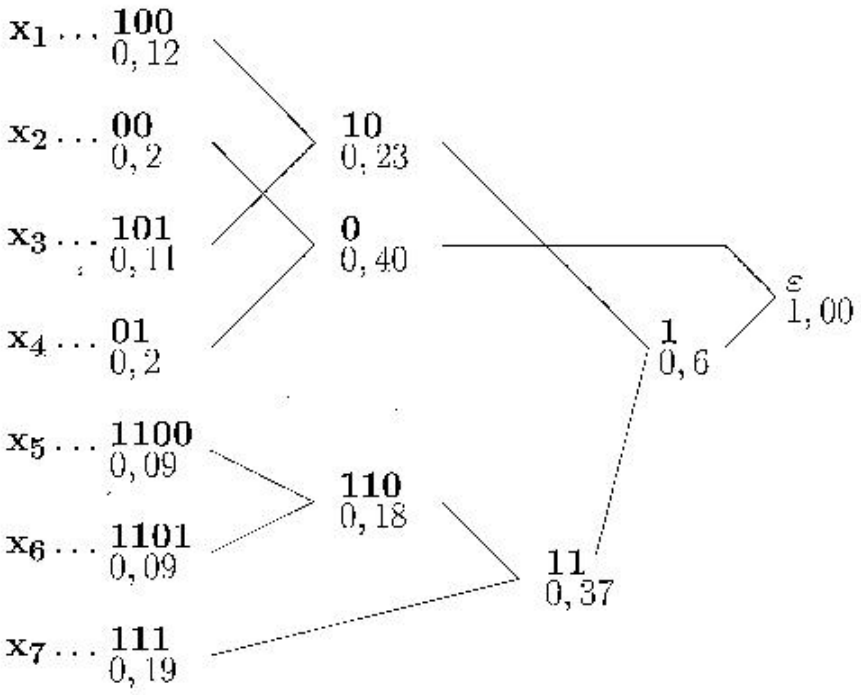
Jsou-li pravděpodobnosti výsledků známé a různé, můžeme to využít k úspornějšímu kódování. U úsporného kódu budou všechny použité znaky přibližně stejně pravděpodobné. Entropie určuje teoretickou dolní mez průměrné délky zprávy při nejúspornějším kódování. Tím nám ukazuje meze možné komprese dat. Častěji známe rozdělení jen přibližně, a proto nedosáhneme maximální účinnosti komprese.

20.5 Kódy s optimální střední délkou

20.5.1 Huffmanův kód

- Je to optimální kód, tedy instantní kód minimalizující střední délku na množině všech jednoznačně dekódovatelných kóků.
- Výsledný kód není určen jednoznačně (např. bitovou inverzí získáme jiný optimální kód).
- Jednoznačně není určena ani délka kódových slov.
- Aplikace : závěrečné zpracování formátů JPEG, MP3, DEFLATE, PKZIP

Příklad:



20.5.2 LZ algoritmy

- autoři Lempel a Ziv publikovali 2 základní varianty algoritmu, a to LZ77 a LZ78
- algoritmus má mnoho různých variací, jako je např. Lempel-Ziv-Welsch LZW (komprese v Unixu, ZIP, RAR, GIF, PDF,...)
- jde o třídu adaptivních kompresních algoritmů se slovníkem
- **optimalita**: LZ77 a LZ78 asymptoticky dosahují rychlosti entropie pro stacionární ergodické zdroje

LZ78 (Stromová verze LZ)

- řetězec $x_1...x_n \in X^n$ je sekvenčně testován na výskyt **nejkratších řetězců**, které se nevyskytly v předchozím kroku
- každý takový řetězec je označen a uložen do **slovníku**
- díky minimalitě ukládaného řetězce jsou jeho **prefixy** již ve slovníku: řetězec $xi_i...x_k$ byl uložen do slovníku před $x_i...x_kx_{k+1}$

Kód tvoří posloupnost dvojic (U_k, x_k), kde

- $x_k$je poslední znak řetězce $x_i...x_lx_k$,
- $U_k$je ukazatel na odpovídající prefix $x_i...x_l$.

Příklad

$X = \{A, B\}$, řetězec ABBABBABBBAABABAA

Dostaneme následující řetězce: A, B, BA, BB, AB, BBA, ABA, BAA

Výsledný kód: 0A 0B 2A 2B 1B 4A 5A 3A

20.6 Informační kanál a jeho kapacita

Definice

Informační kanál je trojice $K = \langle X, \mathbf{P}, Y \rangle$, kde

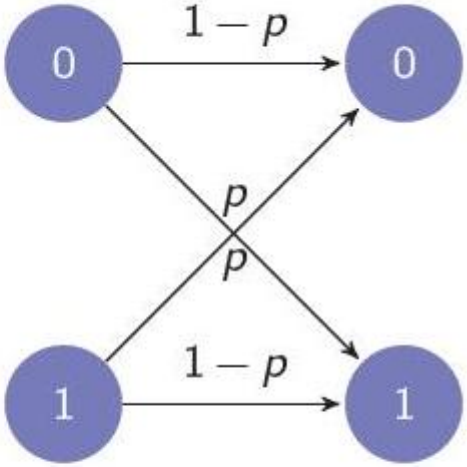
- X je m-prvková vstupní abeceda
- Y je n-prvkovávýstupní abeceda
- $\mathbf{P}$ je matice $m \times n$ podmíněných pravděpodobností:

$$\mathbf{P} = \begin{pmatrix} p_{Y|X}(y_1|x_1) & p_{Y|X}(y_2|x_1) & \dots & p_{Y|X}(y_n|x_1) \\ p_{Y|X}(y_1|x_2) & p_{Y|X}(y_2|x_2) & \dots & p_{Y|X}(y_n|x_2) \\ & \dots\dots\dots\dots\dots\dots & & \\ p_{Y|X}(y_1|x_m) & p_{Y|X}(y_2|x_m) & \dots & p_{Y|X}(y_n|x_m) \end{pmatrix}$$

Příklad

Binární symetrický kanál

$$\mathcal{X} = \mathcal{Y} = \{0, 1\}$$



$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

Vstup a výstup kanálu

Vstup: Informační zdroj X s pravděpodobnostmi

$$(p_X(x_1) \dots p_X(x_m))$$

Výstup: Informační zdroj Y s pravděpodobnostmi

$$(p_Y(y_1) \dots p_Y(y_n)) = (p_X(x_1) \dots p_X(x_m)).\mathbf{P}$$

Přenositelnost zdroje X daným kanálem popisuje **vzájemná informace**

$$I(X;Y) = H(X) - H(X|Y)$$

20.7 Shannonova věta o kódování

20.7.1 Shannonova věta o kódování za přítomnosti šumu

Tato věta říká, že existuje kódování opravující chyby, které dovoluje přenést informaci reálným kanálem (kanál s přítomností šumu) rychlostí libovolně blízkou kapacitě kanálu.

20.7.2 Shannonova věta o kódování bez přítomnosti šumu

Mějme $f : W \to A*$ kód se zdrojovými slovy $w_1, w_2, \ldots, w_m$, délek l_i, s pravděpodobností p_i vyslání slova w_i a kódovými slovy $a_i = f(w_i)$. Pak **průměrná délka kódového slova** je:

$$f = \sum_{i=1}^{m} p_i l_i.$$