MI-SPOL-17

Základy teorie informace a kódování, entropie.

Entropie diskrétní náhodné veličiny

Míra neuspořádanosti

Entropie diskrétní náhodné veličiny X:

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

Logaritmus má bázi 2. Při obecné bázi b je to $H_b(X)$

Jednotka při bázi 2: **bit**

Entropii lze chápat jako **střední hodnotu míry neurčitosti:** $H(X) = -E \log p(X) = EI(X)$, kde $I(X) = -\log p(x)$ pro každé $x \in \mathcal{X}$ je **míra neurčitosti**

Platí $H(X) \geq 0$

Sdružená entropie

Sdružená entropie $H(X,Y)$ diskrétních **náhodných veličin** X, Y se sdruženým rozdělením $p(x,y)$:

$$H(X,Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(x,y)$$

Sdruřená entropie diskrétního **náhodného vektoru** X se sdruženým rozdělením $p(x)$:

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

Podmíněná entropie

Podmíněná entropie $H(Y|X)$ diskrétních náhodných veličin X, Y se sdruženým rozdělením:

$$H(Y|X) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x)$$

, kde $p(y|x) = \frac{p(x,y)}{p(x)}$

Řetězové pravidlo: $H(X, Y) = H(X) + H(Y|X)$
$H(Y|X)$ určuje, co je v Y navíc oproti X

Relativní entropie (Kullback-Leiblerova vzdálenost)

Relativní entropie $D(p||q)$ mezi diskrétním rozdělením p a diskrétním rozdělením q na množině $\mathcal{X}$:

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$$

Vzájemná informace

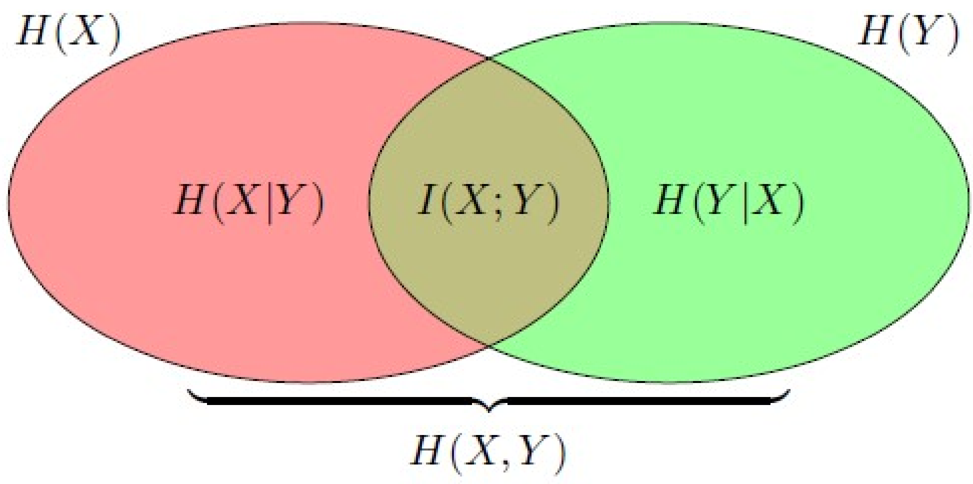
Vzájemná informace $I(X; Y)$ diskrétním náhodných veličin X, Y:

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

Platí, že $I(X; Y) = D(p(x, y)||p(x)p(y))$

Vztahy vzájemné informace a entropie

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



Jensenova neorvnost: f konvexní funkce, X náhodná veličina, potom: $Ef(x) \geq f(EX)$

Informační nerovnost: $p(x), q(x)$ pro $x \in \mathcal{X}$ dvě možná rozdělení diskrétní náhodné veličiny X. Potom: $D(p||q) \geq 0$. ROvnost nastává pouze pokud $p(x) = q(x)$ pro každé $x \in \mathcal{X}$
Důsledek: Pro dvojici diskrétních náhodných veličin X, Y platí $I(X;Y) \geq 0$. Rovnost nastává, právě když jsou X a Y nezávislé ($p(x,y) = p(x)p(y)$)

Maximalizace entropie: Pro diskrétní náhodnou veličinu X s hodnotami z konečné množiny $\mathcal{X}$ platí: $H(X) \leq \log|\mathcal{X}|$. Rovnost nastává, právě když X má rovnoměrné rozdělení na $\mathcal{X}$
$\Rightarrow$ Entropie **maximalizována rovnoměrným rozdělením**

Podmiňování redukuje entropii: Pro diskrétní náhodné veličiny X a Y platí $H(X|Y) \leq H(X)$. Rovnost nastává, pouze pokud X, Y nezávislé
(znalost další veličiny Y může v průměru pouze zredukovat neurčitost X)

Teorie kódování

Zabývá se problémem, jak zapsat zdrojovou zprávu do posloupnosti symbolů, které jsme schopni přenášet, tak, aby byl následný přenos co nejefektivnější (největší komprese, nejmenší náchylnost k chybám)

$\mathcal{D}$**-ární abeceda:** abeceda $\mathcal{D}$ obsahující D symbolů

Zpráva: $x_1x_2...x_n$ složená z konečné posloupnosti znaků z nějaké množiny $\mathcal{X}$

Kódování: Přiřazení kódového slova (konečné posloupnosti znaků z $\mathcal{D}$) každému symbolu z $\mathcal{X}$

(D-ární) Kód náhodné veličiny X: Zobrazení $C : \mathcal{X} \to \mathcal{D}*$ množiny $\mathcal{X}$ do množiny $\mathcal{D}*$ konečných řetězců symbolů D-ární abecedy $\mathcal{D}$

Kódové slovo přislušející prvku x: Obraz $C(x)$ prvku $x \in \mathcal{X}$
Délka: $l(x)$

Střední délka $L(C)$ kódu C: X náhodná veličina s rozdělením $p(x)$, $l(x)$ délka kódového slova příslušjícího k x. Střední délka je:

$$L(C) = \sum_{x \in \mathcal{X}} l(x)p(x)$$

Optimální kód: kód s nejmenší střední délkou

Vlastnosti kódu

Kód C diskrétní náhodné veličiny X je **nesingulární**, pokud C je prosté zobrazení $x \neq x' \Rightarrow C(x) \neq C(x')$

Rozšíření C^* kódu C: zobrazení množiny $\mathcal{X}^*$ do množiny $\mathcal{D}^*$ definované jako

$$C^*(x_1 x_2 ... x_n) = C(x_1)C(x_2)...C(x_n)$$

Kód je **jednoznačně dekódovatelný**, pokud je C^* nesingulární
(možnost jednoznačně dekódovat libovolnou zprávu, ale musí být přijatá celá -- nelze dekódovat postupně přijaté znaky)

Kód je **instantní (prefixový)**, pokud žádné kódové slovo není prefixem jiného slova

Kraftova nerovnost: Pro libovolný instantní kód nad D-ární abecedou musí délky kódových slov $l_1, l_2, ..., l_n$ splnit nerovnost

$$\sum_i D^{-l_i} \leq 1$$

Ke každé n-ici délek, které tuto nerovnost splňují, existuje instnatní kód s kódovými slovy těchto délek

McMillanova věta: Pro libovolný *jednoznačně dekódovatelný* kód nad D-ární abecedou musí délky kódových slov $l_1, l_2, ..., l_n$ splnit nerovnost

$$\sum_i D^{-l_i} \leq 1$$

Ke každé n-ici délek, které tuto nerovnost splňují, existuje *jednoznačně dekódovatelný* kód s kódovými slovy těchto délek
$\Rightarrow$ ke každému jednoznačně dekódovatelnému kódu lze sestrojit instantní kód, který má **stejně dlouhá slova**

Střední délka $L(C)$ instantního D-árního kódu C diskrétní náhodné veličiny X je

$$L(C) \geq H_D(X)$$

Rovnost nastává, právě když $D^{-l_i} = p_i$ pro všechna $i = 1, ..., |\mathcal{X}|$

Střední délka optimálního kódu: Optimální instantní D-ární kód C^* diskrétní náhodné veličiny X. Platí:

$$H_D(X) \leq L(C^*) < H_D(X) + 1$$

Optimálním kódem se tedy lze od dolní meze dané entropií vzdálit maximálně o 1

Huffmanovo kódování

Sestrojení binárního Huffmanova kódu:

- Spojit dvě nejméně pravděpodobné hodnoty do jedné -- vzniká náhodná veličina s o 1 menším očtem hodnot a novým rozdělením
- Opakovat, dokud nezůstane jediná hodnota -- přiřadit jí prázdný řetězec jako kódové slovo
- Zpětným chodem konstruovat kódová slova všech původních hodnot:
  - Hodnota x vzniklá spojením u a v
  - Pro méně pravděpodobnou z hodnot u a v se vyvoří kódové slovo připojením symbolu 1 za $C(x)$
  - Pro více pravděpodobnou připojením symoblu 0 za $C(x)$

Sestrojení D-árního Huffmanova kódu:

- analogicky

- v dopředném chodu agregace D hodnot místo dvou
- ve zpětném chodu připojovány symboly $0, ..., D-1$

Pokud počet hodnot v $\mathcal{X}$ není roven $D+k(D-1)$, musí se počet doplnit pomocnými hodnotami s pravděpodobností 0

Huffmanův kód je **optimální**
Pokud C^* Huffmanův kód a C' libovolný unikátně dekódovatelný kód, $L(C^*) \leq L(C')$