



Zdroje informácie

Stanislav Palúch

Fakulta riadenia a informatiky, Žilinská univerzita

22. apríla 2021



Zdroj informácie – Objekt (osoba, zariadenie, prístroj), ktorý je schopný na svojom výstupe produkovať nejaký signál.

Príklady:

- človek signalizujúci baterkou znaky Morseovej abecedy
- klávesnica počítača, vysielajúca 8-bitové slová
- telefónny prístroj produkujúci analógový signál od 300 do 3400 Hz
- signál z prehrávača kompaktných diskov produkujúci 44000 16-bitových vzoriek za sekundu
- televízny obrazový signál obsahujúci 25 obrázkov za sekundu
- Geiger-Millerova trubica produkujúca jednotkové impulzy

Zdroje informácie môžu produkovať spojité alebo diskretný signál. Každý spojité signál však možno v diskretných časových okamihoch odmerať – diskkrétne vzorkovať.

Veta (Nyquist-Shannon sampling theorem)

Pokiaľ je vzorkovacia frekvencia aspoň dvojnásobná ako maximálna frekvencia signálu, stačia tieto diskkrétne vzorky na plnohodnotnú rekonštrukciu pôvodného signálu.

Môžeme teda predpokladať, že zdroj produkuje v časových okamihoch $t = t_1, t_2, t_3, \dots$ signály $X_{t_1}, X_{t_2}, X_{t_3}, \dots$, ktoré môžeme považovať za diskkrétne náhodné veličiny – nadobúdajú len konečne veľa rôznych hodnôt.

Konečnú množinu rôznych diskretných signálov produkovaných zdrojom nazveme **abecedou zdroja**, jednotlivé prvky abecedy zdroja nazveme **znakmi**.

Časové okamihy $t = t_1, t_2, t_3, \dots$ môžu, ale nemusia byť pravidelné.

Je však výhodné považovať časový interval medzi dvoma za sebou nasledujúcimi časovými okamihmi za jednotkový – potom pracujeme s náhodnými veličinami X_1, X_2, X_3, \dots .

Definícia

Diskrétny náhodný proces je postupnosť náhodných veličín

$$\mathcal{X} = X_1, X_2, X_3, \dots$$

Ak X_i nadobudne hodnotu a_i pre $i = 1, 2, \dots$, postupnosť

$$a_1, a_2, \dots$$

nazveme **realizáciou náhodného procesu** \mathcal{X} .

Každú konkrétnu postupnosť znakov zo zdroja môžeme považovať za realizáciu niektorého náhodného procesu \mathcal{X} .



Zdroje sa líšia

- frekvenciou vysielania znakov
- mohutnosťou abecedy zdroja
- pravdepodobnostným rozdelením náhodného procesu \mathcal{X}

Treba zistiť informačnú výdatnosť zdroja, t.j.

- koľko informácie produkuje zdroj za jednotku času
- koľko informácie produkuje zdroj na jeden výstupný znak

Závislosť množstva informácie za jednotku času na frekvencii je lineárna, a preto budeme skúmať množstvo informácie pripadajúce na jeden znak.

Toto závisí nielen na mohutnosti abecedy zdroja, ale aj od rozdelenia pravdepodobnosti náhodných veličín X_j .

Definícia

Nech X je konečná množina, nech X^* je množina všetkých konečných postupností prvkov z X vrátane prázdnej postupnosti, ktorú budeme značiť symbolom ϵ .

Množinu X nazveme **abecedou**, jej prvky **znakmi abecedy** X , prvky množiny X^* nazveme **slovami**, ϵ **prázdny slovom**.

Označme X^n množinu všetkých n -prvkových postupností znakov z X , jej prvky nazveme **slovami dĺžky n** .

Definícia

Nech $P : X^* \rightarrow \mathbb{R}$ je reálna nezáporná funkcia definovaná na X^* s nasledujúcimi vlastnosťami:

$$1. \quad P(e) = 1 \quad (1)$$

$$2. \quad \sum_{(x_1, \dots, x_n) \in X^n} P(x_1, \dots, x_n) = 1 \quad (2)$$

$$3. \quad \sum_{(y_{n+1}, \dots, y_{n+m}) \in X^m} P(x_1, \dots, x_n, y_{n+1}, \dots, y_{n+m}) = P(x_1, \dots, x_n) \quad (3)$$

Potom usporiadanú dvojicu $\mathcal{Z} = (X^*, P)$ nazveme **zdrojom informácie** alebo krátko **zdrojom**.

Číslo $P(x_1, x_2, \dots, x_n)$ nazveme **pravdepodobnosťou slova** x_1, \dots, x_n .

Zaujíma nás pravdepodobnosť $P_n(y_1, y_2, \dots, y_m)$, s akou zdroj vyšle slovo y_1, y_2, \dots, y_m v čase n , presnejšie v časových okamihoch

$$n, n + 1, \dots, n + m - 1.$$

Túto pravdepodobnosť vypočítame nasledovne:

$$P_n(y_1, y_2, \dots, y_m) = \sum_{(x_1, \dots, x_{n-1}) \in X^{n-1}} P(x_1, x_2, \dots, x_{n-1}, y_1, y_2, \dots, y_m) \cdot \quad (4)$$

Definícia

Hovoríme, že zdroj $\mathcal{Z} = (X^*, P)$ je **stacionárny**, ak pravdepodobnosti $P_i(x_1, x_2, \dots, x_n)$ nezávisia od i ,
t. j. ak pre každé i a každé $x_1, x_2, \dots, x_n \in X^n$

$$P_i(x_1, x_2, \dots, x_n) = P(x_1, x_2, \dots, x_n) \cdot$$

Označme X_i diskretnú náhodnú premennú, ktorá bude popisovať vyslanie jedného znaku zo zdroja $\mathcal{Z} = (X^*, P)$ v čase i .

Potom jav „v čase i zdroj vyslal znak x “ je vlastne javom $[X_i = x]$ a teda

$$P([X_i = x]) = P_i(x).$$

Vyslanie slova x_1, x_2, \dots, x_n v čase i možno pomocou náhodných veličín X_i modelovať ako jav

$$[X_i = x_1] \cap [X_{i+1} = x_2] \cap \dots \cap [X_{i+n-1} = x_n],$$

čo skrátene zapíšeme

$$[X_i = x_1, X_{i+1} = x_2, \dots, X_{i+n-1} = x_n],$$

z čoho máme

$$P([X_i = x_1, X_{i+1} = x_2, \dots, X_{i+n-1} = x_n]) = P_i(x_1, x_2, \dots, x_n).$$

Definícia

Hovoríme, že zdroj $\mathcal{Z} = (X^*, P)$ je **nezávislý**, ak pre ľubovoľné i, j, n, m také, že $i + n \leq j$ platí

$$\begin{aligned} &P\left([X_i = x_1, X_{i+1} = x_2, \dots, X_{i+n-1} = x_n] \cap \right. \\ &\quad \left. \cap [X_j = y_1, X_{j+1} = y_2, \dots, X_{j+m-1} = y_m]\right) = \\ &= P([X_i = x_1, X_{i+1} = x_2, \dots, X_{i+n-1} = x_n]) \cdot \\ &\quad \cdot P([X_j = y_1, X_{j+1} = y_2, \dots, X_{j+m-1} = y_m]) = \\ &= P_i(x_1, x_2, \dots, x_n) \cdot P_j(y_1, y_2, \dots, y_m). \end{aligned}$$

Zdroj je nezávislý, ak vyslanie ľubovoľného slova v ľubovoľnom čase j nezávisí od toho, čo zdroj vyslal do času j .

Niekedy sa takýmto zdrojom hovorí aj bezpamätové.

Majme stacionárny zdroj $\mathcal{Z} = (Z^*, P)$ s abecedou $Z = \{a_1, a_2, \dots, a_m\}$.

Vyslanie znaku v ľubovoľnom čase možno pri stacionárnom zdroji považovať za vykonanie pokusu

$$\mathbf{B} = \{\{a_1\}, \{a_2\}, \dots, \{a_m\}\}$$

s pravdepodobnosťami $p_1 = P(a_1)$, $p_2 = P(a_2)$, \dots , $p_m = P(a_m)$.

Entropia tohoto pokusu je $H(\mathbf{B}) = H(p_1, p_2, \dots, p_m)$, čo je stredná hodnota informácie získanej týmto pokusom.

Skúmajme teraz informáciu, ktorú dostaneme v dvoch po sebe idúcich znakoch vyslaných zo stacionárneho zdroja $\mathcal{Z} = (Z^*, P)$.

Príslušný pokus bude teraz

$$\mathbf{C}_2 = \{(a_{i_1}, a_{i_2}) \mid a_{i_1} \in Z, a_{i_2} \in Z\} .$$

Pokus \mathbf{B} môžeme prezentovať aj ako

$$\mathbf{B} = \{\{a_1\} \times Z, \{a_2\} \times Z, \dots, \{a_m\} \times Z\} .$$

Ak definujeme $\mathbf{D} = \{Z \times \{a_1\}, Z \times \{a_2\}, \dots, Z \times \{a_m\}\}$, potom

$$\mathbf{C}_2 = \mathbf{B} \wedge \mathbf{D}$$

a

$$H(\mathbf{D}) = H(\mathbf{B}) = H(p_1, p_2, \dots, p_m) .$$

Podľa vety platí

$$H(\mathbf{C}_2) = H(\mathbf{B} \wedge \mathbf{D}) \leq H(\mathbf{B}) + H(\mathbf{D}) = 2.H(\mathbf{B}) .$$

Teraz túto vlastnosť rozšírime na n -znakové slová. Budeme postupovať matematickou indukciou.

Predpokladajme, že pre pokus

$$\mathbf{C}_n = \{(a_{i_1}, a_{i_2}, \dots, a_{i_n}) \mid a_{i_k} \in Z, \text{ pre } k = 1, 2, \dots, n\}$$

už platí $H(\mathbf{C}_n) \leq n.H(\mathbf{B})$.

Pokus \mathbf{C}_n má rovnakú entropiu ako pokus

$$\mathbf{C}'_n = \{(a_{i_1}, a_{i_2}, \dots, a_{i_n}) \times Z \mid a_{i_k} \in Z, \text{ pre } k = 1, 2, \dots, n\}.$$

Označme

$$\mathbf{C}_{n+1} = \{(a_{i_1}, a_{i_2}, \dots, a_{i_{n+1}}) \mid a_{i_k} \in Z, \text{ pre } k = 1, 2, \dots, n+1\},$$

$$\mathbf{D} = \{Z^n \times \{a_1\}, Z^n \times \{a_2\}, \dots, Z^n \times \{a_m\}\},$$

potom $H(\mathbf{D}) = H(\mathbf{B})$ a

$$H(\mathbf{C}_{n+1}) = H(\mathbf{C}'_n \wedge \mathbf{D}) \leq H(\mathbf{C}'_n) + H(\mathbf{D}) \leq n.H(\mathbf{B}) + H(\mathbf{B}) = (n+1).H(\mathbf{B}).$$

Tým sme dokázali, že pre všetky prirodzené n platí

$$H(\mathbf{C}_n) \leq n.H(\mathbf{B})$$

Vidíme, že v prípade stacionárneho zdroja, ktorý nie je nezávislý, je priemerná entropia na jedno písmeno $\frac{1}{n} H(\mathbf{C}_n)$ vždy menšia ako entropia prvého písmena.

To nás vedie k myšlienke, definovať entropiu zdroja ako priemernú entropiu na jedno písmeno pre veľmi dlhé slová.

Definícia

Nech $\mathcal{Z} = (Z^*, P)$ je zdroj informácie. Nech existuje limita

$$H(\mathcal{Z}) = - \lim_{n \rightarrow \infty} \frac{1}{n} \cdot \sum_{(x_1, \dots, x_n) \in Z} P(x_1, x_2, \dots, x_n) \cdot \log_2(P(x_1, x_2, \dots, x_n)).$$

Potom číslo $H(\mathcal{Z})$ nazveme **entropiou zdroja** \mathcal{Z} .



Veta

Nech (Z^*, P) je stacionárny nezávislý zdroj. Potom

$$H(Z) = - \sum_{x \in Z} P(x) \cdot \log_2 P(x).$$

■ Dôkaz. Platí:

$$\begin{aligned}
 & \sum_{(x_1, \dots, x_n) \in Z} P(x_1, x_2, \dots, x_n) \cdot \log_2(P(x_1, x_2, \dots, x_n)) = \\
 & = \sum_{(x_1, \dots, x_n) \in Z} P(x_1) \cdot P(x_2), \dots, P(x_n) \cdot [\log_2 P(x_1) + \log_2 P(x_2) + \dots + \log_2 P(x_n)] = \\
 & = \sum_{(x_1, \dots, x_n) \in Z} P(x_1) \cdot P(x_2), \dots, P(x_n) \cdot \log_2 P(x_1) + \\
 & + \sum_{(x_1, \dots, x_n) \in Z} P(x_1) \cdot P(x_2), \dots, P(x_n) \cdot \log_2 P(x_2) + \\
 & + \dots + \\
 & + \sum_{(x_1, \dots, x_n) \in Z} P(x_1) \cdot P(x_2), \dots, P(x_n) \cdot \log_2 P(x_n) = \\
 & = \sum_{x_1 \in Z} P(x_1) \cdot \log_2 P(x_1) \cdot \underbrace{\sum_{(x_2, \dots, x_n) \in Z} P(x_2) \cdot P(x_3), \dots, P(x_n)}_{=1} + \dots = \\
 & = \sum_{x_1 \in Z} P(x_1) \cdot \log_2 P(x_1) + \sum_{x_2 \in Z} P(x_2) \cdot \log_2 P(x_2) + \dots + \sum_{x_3 \in Z} P(x_3) \cdot \log_2 P(x_3) = \\
 & = n \cdot \sum P(x) \cdot \log_2 P(x).
 \end{aligned}$$

Veta

Shannon – Mac Millan. Nech $\mathcal{Z} = (Z^*, P)$ je stacionárny nezávislý zdroj. Potom k ľubovoľnému $\varepsilon > 0$ existuje prirodzené číslo $n(\varepsilon)$ také, že pre všetky $n \geq n(\varepsilon)$ je

$$P \left\{ x_1, \dots, x_n \in Z^n \mid \left| \frac{1}{n} \cdot \log_2 P(x_1, \dots, x_n) + H(\mathcal{Z}) \right| \geq \varepsilon \right\} < \varepsilon. \quad (5)$$

Označme

$$E(n, \varepsilon) = \left\{ x_1, \dots, x_n \in Z^n \mid \underbrace{\left| \frac{1}{n} \cdot \log_2 P(x_1, \dots, x_n) + H(\mathcal{Z}) \right|}_{=H(\mathcal{Z}) - \frac{1}{n} I(x_1, \dots, x_n)} < \varepsilon \right\}$$

$-\log_2 P(x_1, \dots, x_n) = I(x_1, \dots, x_n)$ – informácia slova (x_1, \dots, x_n)

$-\frac{1}{n} \log_2 P(x_1, \dots, x_n) = \frac{1}{n} I(x_1, \dots, x_n)$ – priemerná informácia na jeden znak slova (x_1, \dots, x_n)

$E(n, \varepsilon)$ je množina n -znakových slov, ktorých informácia na jeden znak sa líši od $H(\mathcal{Z})$ menej než ε .

$\forall \varepsilon \quad \exists n(\varepsilon)$ také, že $\forall n > n(\varepsilon) \quad P(E(n, \varepsilon)) \geq 1 - \varepsilon$

Platí:

$$\begin{aligned} (x_1, \dots, x_n) \in E(n, \varepsilon) &\iff \\ -\varepsilon < \frac{1}{n} \log_2 P(x_1, \dots, x_n) + H(\mathcal{Z}) < \varepsilon &\iff \\ \iff -n(H(\mathcal{Z}) + \varepsilon) < \log_2 P(x_1, \dots, x_n) < -n(H(\mathcal{Z}) - \varepsilon) &\iff \\ \iff 2^{-n(H(\mathcal{Z})+\varepsilon)} < P(x_1, \dots, x_n) < 2^{-n(H(\mathcal{Z})-\varepsilon)} & \end{aligned}$$

Nech $|E(n, \varepsilon)|$ je počet prvkov v množine $E(n, \varepsilon)$.

Pretože pravdepodobnosť každého prvku množiny $E(n, \varepsilon)$ je väčšia než $2^{-n(H(\mathcal{Z})+\varepsilon)}$, je

$$1 \geq P(E(n, \varepsilon)) > |E(n, \varepsilon)| \cdot 2^{-n(H(\mathcal{Z})+\varepsilon)},$$

$$\boxed{|E(n, \varepsilon)| < 2^{n(H(\mathcal{Z})+\varepsilon)}}.$$

Na druhej strane je pravdepodobnosť každého prvku množiny $E(n, \varepsilon)$ menšia než $2^{-n(H(\mathcal{Z})-\varepsilon)}$, z čoho

$$1 - \varepsilon < P(E(n, \varepsilon)) < |E(n, \varepsilon)| \cdot 2^{-n(H(\mathcal{Z})-\varepsilon)}.$$

$$1 - \varepsilon < |E(n, \varepsilon)| \cdot 2^{-n(H(\mathcal{Z})-\varepsilon)}.$$

$$(1 - \varepsilon) \cdot 2^{n(H(\mathcal{Z})-\varepsilon)} < |E(n, \varepsilon)|$$

Je teda

$$(1 - \varepsilon) \cdot 2^{n(H(\mathcal{Z})-\varepsilon)} < |E(n, \varepsilon)| < 2^{n(H(\mathcal{Z})+\varepsilon)}$$

Ku každému ε existuje $n(\varepsilon)$ také, že pre všetky $n \geq n(\varepsilon)$ je počet n -znakových slov nesúcich na jeden znak informáciu blízku $H(\mathcal{Z})$ aproximovateľný hodnotou $2^{nH(\mathcal{Z})}$.

$$|E(n, \varepsilon)| \approx 2^{nH(\mathcal{Z})} \tag{6}$$

Slovenčina používa 26 písmen abecedy bez diakritiky a 15 písmen s diakritikou á, č, ď, é, í, ľ, í, ň, ó, ô, š, ť, ú, ý, ž.

Navyše sa používajú aj interpunkčné znamienka (čiarka, dvojbodka, bodkočiarka, pomlčka, úvodzovky, bodka, výkričník, otáznik a medzera). Aj keď prijmem námiatku, že slovenčina by vystačila bez písmen q, w, x, potrebuje jej abeceda Z minimálne 40 znakov. (A to ešte nepoužívame veľké písmená.)

Entropia slovenčiny určite neprevýši číslo 2.

Počet všetkých 8-znakových slov abecedy Z je teda 40^8 , $|E(8, \varepsilon)|$ odhadneme na $2^{8 \cdot 2} = 2^{16}$.

$$\text{Je } \frac{2^{16}}{40^8} = 6 \cdot 10^{-8}$$

Množina $E(8, \varepsilon)$ významných 8-znakových slov obsahuje približne 6 milióntin percenta počtu všetkých 8-znakových slov.

Definícia

Majme dva informačné zdroje $\mathcal{Z}_1 = (A^*, P_1)$, $\mathcal{Z}_2 = (B^*, P_2)$. **Produktom zdrojov** \mathcal{Z}_1 , \mathcal{Z}_2 nazveme zdroj $\mathcal{Z}_1 \times \mathcal{Z}_2 = ((A \times B)^*, P)$, kde $(A \times B)$ je karteziánskym súčinom množín A a B a kde

$$P(e) = 1$$

$$P((a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)) = P(a_1, a_2, \dots, a_n) \cdot P(b_1, b_2, \dots, b_n)$$

pre ľubovoľné $a_i \in A$, $b_j \in B$, $i, j \in \{1, 2, \dots, n\}$.

Pre pravdepodobnosti zdroja musí platiť

$$1. \quad P(e) = 1 \quad (7)$$

$$2. \quad \sum_{(x_1, \dots, x_n) \in X^n} P(x_1, \dots, x_n) = 1 \quad (8)$$

$$3. \quad \sum_{(y_{n+1}, \dots, y_{n+m}) \in X^m} P(x_1, \dots, x_n, y_{n+1}, \dots, y_{n+m}) = P(x_1, \dots, x_n) \quad (9)$$

Veta

Produkt zdrojov $\mathcal{Z}_1, \mathcal{Z}_2$ je korektne definovaný, t. j. pre pravdepodobnosť P platí (1), (2), (3) z definície zdroja.

Dôkaz.

Vzťahy (1), (2), (3) z definície zdroja prepíšeme nasledovne:

$$1. \quad P(e) = 1 \quad (10)$$

$$2. \quad \sum_{(a_1, b_1), \dots, (a_n, b_n) \in (A \times B)^n} P((a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)) = 1 \quad (11)$$

$$3. \quad \sum_{(p_1, q_1), \dots, (p_m, q_m) \in (A \times B)^m} P((a_1, b_1), (a_2, b_2), \dots, (a_n, b_n), (p_1, q_1), \dots, (p_m, q_m)) = \\ = P((a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)) \quad (12)$$

Prvý vzťah vyplýva z definície zdroja $\mathcal{Z}_1 \times \mathcal{Z}_2$.

Dokážeme platnosť druhého vzťahu.

$$\begin{aligned} & \sum_{(a_1, b_1), \dots, (a_n, b_n) \in (A \times B)^n} P((a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)) = \\ & = \sum_{(a_1, b_1), \dots, (a_n, b_n) \in (A \times B)^n} P(a_1, a_2, \dots, a_n) \cdot P(b_1, b_2, \dots, b_n) \\ & = \sum_{(a_1, a_2, \dots, a_n) \in A^n} P(a_1, a_2, \dots, a_n) \cdot \sum_{(b_1, b_2, \dots, b_n) \in B^n} P(b_1, b_2, \dots, b_n) = 1 \cdot 1 = 1 \end{aligned}$$

Dokážeme platnosť tretieho vzťahu.

$$\begin{aligned} & \sum_{(p_1, q_1), \dots, (p_m, q_m) \in (A \times B)^m} P((a_1, b_1), (a_2, b_2), \dots, (a_n, b_n), (p_1, q_1), \dots, (p_m, q_m)) = \\ & = \sum_{p_1 p_2 \dots p_m \in A^m} \sum_{q_1 q_2 \dots q_m \in B^m} P(a_1, \dots, a_n, p_1, \dots, p_m) \cdot P(b_1, \dots, b_n, q_1, \dots, q_m) = \\ & = \sum_{p_1 p_2 \dots p_m \in A^m} P(a_1, \dots, a_n, p_1, \dots, p_m) \cdot \sum_{q_1 q_2 \dots q_m \in B^m} P(b_1, \dots, b_n, q_1, \dots, q_m) = \\ & = P(a_1, a_2, \dots, a_n) \cdot P(b_1, b_2, \dots, b_n) = P((a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)) . \end{aligned}$$

Veta

Majme dva informačné zdroje \mathcal{Z}_1 , \mathcal{Z}_2 s entropiami $H(\mathcal{Z}_1)$, $H(\mathcal{Z}_2)$. Potom pre entropiu zdroja $\mathcal{Z}_1 \times \mathcal{Z}_2$ platí

$$H(\mathcal{Z}_1 \times \mathcal{Z}_2) = H(\mathcal{Z}_1) + H(\mathcal{Z}_2). \quad (13)$$

Dôkaz.

$$\begin{aligned} H(\mathcal{Z}_1 \times \mathcal{Z}_2) &= \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{(a_1, b_1), \dots, (a_n, b_n) \in (A \times B)^n} P((a_1, b_1), \dots, (a_n, b_n)) \cdot \log_2 P((a_1, b_1), \dots, (a_n, b_n)) = \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{(a_1, b_1), \dots, (a_n, b_n) \in (A \times B)^n} \left\{ P(a_1, \dots, a_n) \cdot P(b_1, \dots, b_n) \cdot \right. \\ &\quad \left. \cdot [\log_2 P(a_1, \dots, a_n) + \log_2 P(b_1, \dots, b_n)] \right\} = \end{aligned}$$

Produkt informačných zdrojov

$$\begin{aligned} &= \lim_{n \rightarrow \infty} \frac{1}{n} \left[\sum_{(a_1, b_1), \dots, (a_n, b_n) \in (A \times B)^n} P(a_1, \dots, a_n) \cdot P(b_1, \dots, b_n) \cdot \log_2 P(a_1, \dots, a_n) + \right. \\ &\quad \left. + \sum_{(a_1, b_1), \dots, (a_n, b_n) \in (A \times B)^n} P(a_1, \dots, a_n) \cdot P(b_1, \dots, b_n) \cdot \log_2 P(b_1, \dots, b_n) \right] = \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \left[\sum_{a_1, \dots, a_n \in A^n} P(a_1, \dots, a_n) \cdot \log_2 P(a_1, \dots, a_n) \cdot \underbrace{\sum_{b_1, \dots, b_n \in B^n} P(b_1, \dots, b_n)}_{=1} + \right. \\ &\quad \left. + \sum_{b_1, \dots, b_n \in B^n} P(b_1, \dots, b_n) \log_2 P(b_1, \dots, b_n) \cdot \underbrace{\sum_{a_1, \dots, a_n \in A^n} P(a_1, \dots, a_n)}_{=1} \right] = \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{a_1, \dots, a_n \in A^n} P(a_1, \dots, a_n) \cdot \log_2 P(a_1, \dots, a_n) + \\ &\quad + \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{b_1, \dots, b_n \in B^n} P(b_1, \dots, b_n) \log_2 P(b_1, \dots, b_n) = H(Z_1) + H(Z_2). \end{aligned}$$

Nech $\mathcal{Z} = (A^*, P)$ je informačný zdroj. Definujme

$$\mathcal{Z}^2 = \mathcal{Z} \times \mathcal{Z}$$

a ďalej indukciou

$$\mathcal{Z}^n = \mathcal{Z}^{n-1} \times \mathcal{Z}.$$

Zdroj $\mathcal{Z}^n = \underbrace{\mathcal{Z} \times \mathcal{Z} \times \cdots \times \mathcal{Z}}_{n\text{-krát}}$ je zdroj s abecedou A^n .

Použitím predchádzajúcej vety a matematickej indukcie dostaneme:

Veta

Nech \mathcal{Z} je informačný zdroj s entropiou $H(\mathcal{Z})$. Potom pre entropiu $H(\mathcal{Z}^n)$ zdroja \mathcal{Z}^n platí

$$H(\mathcal{Z}^n) = n.H(\mathcal{Z})$$

Definícia

Nech $\mathcal{Z} = (A^*, P)$. Označme $\mathcal{Z}_{(k)} = ((A^k)^*, P_{(k)})$ zdroj s abecedou A^k , kde $P_{(k)}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ pre $\mathbf{a}_i \in A^k$, $\mathbf{a}_i = a_{i1}a_{i2} \dots a_{ik}$ je definované ako

$$P_{(k)}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = P(\underbrace{a_{11}, a_{12}, \dots, a_{1k}}_{\mathbf{a}_1}, \underbrace{a_{21}, a_{22}, \dots, a_{2k}}_{\mathbf{a}_2}, \dots, \underbrace{a_{n1}, a_{n2}, \dots, a_{nk}}_{\mathbf{a}_n})$$

Informačný zdroj $\mathcal{Z}_{(k)}$ vznikne z informačného zdroja \mathcal{Z} tak, že zo zdroja \mathcal{Z} budeme odoberať každý k -ty okamih celé výstupné slovo dĺžky k v pôvodnej abecede, pričom budeme výstupné k -znakové slová brať ako znaky novej abecedy.

Združenie k znakov zdroja do k-znakových slov

Veta

Nech \mathcal{Z} je informačný zdroj s entropiou $H(\mathcal{Z})$. Potom pre entropiu $H(\mathcal{Z}_{(k)})$ zdroja $\mathcal{Z}_{(k)}$ platí

$$H(\mathcal{Z}_{(k)}) = k.H(\mathcal{Z})$$

Dôkaz.

Platí

$$\begin{aligned} H(\mathcal{Z}_{(k)}) &= - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\mathbf{a}_1, \dots, \mathbf{a}_n \in A^n} P_{(k)}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) \cdot \log_2(P_{(k)}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)) = \\ &= - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{a_{ij} \in A \text{ pre } 1 \leq i \leq n, 1 \leq j \leq k} P(\mathbf{a}_{11}, \dots, \mathbf{a}_{1k}, \mathbf{a}_{21}, \dots, \mathbf{a}_{2k}, \dots, \mathbf{a}_{n1}, \dots, \mathbf{a}_{nk}) \cdot \\ &\quad \cdot \log_2(P(\mathbf{a}_{11}, \dots, \mathbf{a}_{1k}, \mathbf{a}_{21}, \dots, \mathbf{a}_{2k}, \dots, \mathbf{a}_{n1}, \dots, \mathbf{a}_{nk})) = \\ &= - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{x_1, x_2, \dots, x_{n.k} \in A} P(x_1, x_2, \dots, x_{n.k}) \cdot \log_2(P(x_1, x_2, \dots, x_{n.k})) = \\ &= -k \cdot \left[\lim_{n \rightarrow \infty} \frac{1}{k.n} \sum_{x_1, x_2, \dots, x_{n.k} \in A} P(x_1, x_2, \dots, x_{n.k}) \cdot \log_2(P(x_1, x_2, \dots, x_{n.k})) \right] = k.H(\mathcal{Z}) \end{aligned}$$



Združenie k znakov zdroja do k-znakových slov



Združenie k znakov zdroja do k-znakových slov
