

Kolmogorovská zložitosť

24.10.2013

Bloky v reťazcoch

//opakovanie

Theorem \exists konštanta c taká, že \forall deficiency funkciu δ , $\forall n$ a $x \in \{0, 1\}^n$

$$\text{ak } \mathbf{C(x)} \geq \mathbf{n - \delta(n)} \quad \text{tak} \quad \left| \#_1(\mathbf{x}) - \frac{\mathbf{n}}{2} \right| \leq \sqrt{\frac{3}{2} \frac{\delta(\mathbf{n}) + \mathbf{c} \cdot \mathbf{n}}{\log e}}$$

Theorem \exists konštanta c taká, že $\forall n$ a $x \in \{0, 1\}^n$

$$\text{ak } \left| \#_1(\mathbf{x}) - \frac{\mathbf{n}}{2} \right| \leq 2^{-\delta(\mathbf{n}) - \mathbf{c}} \sqrt{\mathbf{n}} \quad \text{tak} \quad \mathbf{C(x)} \leq \mathbf{n - \delta(n)}$$

Theorem \exists konštanta c taká, že $\forall n$ a $x \in \{0, 1\}^n$

$$\text{ak } \left| \#_1(\mathbf{x}) - \frac{\mathbf{n}}{2} \right| = \mathbf{j} \quad \text{tak} \quad \mathbf{C(x|n)} \leq \mathbf{n - 1/2 \log n + K(j|n) + c}$$

Theorem Nech $l(y) = \ell$, $\ell \leq \log n$. \exists konštanta c tak, že $\forall n \forall x \in \{0, 1\}^n$

$$\text{ak } \mathbf{C(x)} \geq \mathbf{n - \delta(n)} \quad \text{potom} \quad \left| \#_y(\mathbf{x}) - \mathbf{pn} \right| \leq \sqrt{\alpha \mathbf{pn}}$$

$$\text{kde } \alpha = \lceil \mathbf{K(y|n) + \log \ell + \delta(n) + c} \rceil 3\ell / \log e$$

Theorem (Blok núl v náhodných reťazcoch)

Nech x dĺžky n spĺňa $C(x) \geq n - \delta(n)$. Potom pre dostatočne veľké n platí, že v x sa vyskytuje každý blok veľkosti

$$\ell = \log n - \log \log n - \log(\delta(n) + \log(n)) - O(1)$$

- Ak $\delta(n) = O(\log n)$, tak sa každý blok dĺžky $\log n - 2 \log \log n - O(1)$ v x vyskytuje aspoň raz
- Ak $\delta(n) = O(\log \log n)$, tak $\forall \epsilon > 0$ a dost' veľké n každý reťazec x dĺžky n obsahuje $y = 0^\ell$, pričom

$$K(y|n) = O(\log n), \ell = \log n - (1 + \epsilon) \log \log n + O(1)$$

//o kvalite odhadu bloku núl...

 $x = uvw$ také, že $C(x) \geq n - \delta(n)$, popíšeme

- v ($K(v)$ bitov), uw ($n - l(v)$ bitov)
- popis $l(u)$ $\log n + \log \log n + 2 \log \log \log n + O(1)$

$$C(x) \leq \underbrace{n - l(v)}_{uw} + K(v) + \log n + \underbrace{\log \log n + 2 \log \log \log n + O(1)}_{(1+o(1)) \log \log n}$$

pre $K(v) = o(\log \log n)$

$$n - \delta(n) \leq C(x) \leq n - l(v) + o(\log \log n) + \log n + (1 + o(1)) \log \log n + O(1)$$

$$l(v) \leq \delta(n) + \log n + (1 + o(1)) \log \log n$$

Zložitý reťazec $C(x) = n + O(1)$

- **neobsahuje** podreťazec núl dĺžky $\log n + (1 + \epsilon) \log \log n$ pre dostatočne dlhé a pravidelné n .
- **Musí obsahovať** reťazec núl dĺžky $\log n - (1 + \epsilon) \log \log n + O(1)$

Algoritmické vlastnosti C

Vieme, že C

- neohraničene rastie
- rastie pomalšie ako každá monotónne rastúca neohraničená RE
- nie je rekurzívna, ale dá sa aproximovať

Theorem

O Kolmogorovskej zložitosti C platí

- 1** $A = \{(x, a) : C(x) \leq a\}$ je RE ale nie R
- 2** \forall čiastočne rekurzívna $\Phi(x)$, kt. je dolným odhadom $C(x)$, je ohraničená
- 3** Nech $f(x)$ je totálne rekurzívna, $g(x) \leq f(x) \leq l(x) \forall x$ a nejakú neohraničenú monotónnu funkciu $g(x)$. Potom

$$B = \{x : C(x) \leq f(x)\}$$

je jednoduchá/simple (RE, B^C nekonečná, kt. neobsahuje žiadnu nekonečnú RE podmnožinu)

Každá PRF $\Phi(x)$, kt. je dolným odhadom $C(x)$, je ohraničená

Nech $D = \{x : \Phi(x) \leq C(x)\}$

- D konečná
- D nekonečná, $\Phi = \Phi_k$, $\Phi(x)$ **neohraničená**

T **enumeruje** jej definičný obor bez opakovania

$g(n) = \min\{x : \Phi(x) \geq n\}$ //totálna rekurzívna

$$n \leq \Phi_k(x) \leq C(x) \leq l(n) + l(\bar{k}) + O(1)$$

$$n \leq l(n) + l(\bar{k}) + O(1) \quad \text{spor}$$

Algoritmické vlastnosti C

Nech $f(x)$ je totálne rekurzívna, $g(x) \leq f(x) \leq l(x) \forall x$ a nejakú neohraničenú monotónnu funkciu $g(x)$. Potom

$$B = \{x : C(x) \leq f(x)\}$$

je jednoduchá/simple (RE, B^C nekonečná, kt. neobsahuje žiadnu nekonečnú RE podmnožinu)

- B je RE:

$f(x)$ rekurzívna: testom programov dĺžky $\leq f(x)$ vymenováваме $x \in B$

- B^C je nekonečná (nestlačiteľné reťazce)
- B^C neobsahuje nekonečnú RE podmnožinu

NECH D je nekonečná RE, $D \subseteq B^C$

- zúženie $f_D(x)$ je PRF
- pre $x \in D \subseteq B^C$ $f_D(x) = f(x) < C(x)$, preto je $f_D(x)$ ohraničená
- pritom $f(x)$ je neohraničená \rightsquigarrow D je konečná

Example (Nerozhodnuteľnosť nestlačiteľnosťou)

Nech T je formálny systém. Potom existuje konštanta c_T tak, že tvrdenie typu $C(x) \geq c_T$ je nedokázateľné

- T axiomatizovateľná \rightsquigarrow k_T bitov na popis $C(T) \leq k_T$
- T bezsporná
- $S_c(x) \sim x$ je lexikograficky najmenšie x dĺžky c pre ktoré $C(x) \geq c$

$$C(S_c(x)) \leq \log c + O(1) \quad O(1) \sim s$$

- $\forall c \exists! x : S_c(x) = \text{true}$

popis x \bar{T} popis S_c

$$C(x) \leq 2k_T + 1 + \log c + s + O(1)$$

spor s tým, že $C(x) > c \forall c > c_T$

Nerozhodnuteľnosť nestlačiteľnosťou

Dôsl.

Nech B je simple (RE, B^C nekonečná bez nekonečnej RE podmnožiny). Potom množina $D = \{n \mid \text{dá sa dokázať, že } n \in B^C\}$ je konečná.

- D je enumerovateľná
- žiadna nekonečná podmnožina B^C – ani D – nie je RE

Dôsl.

- nekonečne veľa nedokázateľných formúl " $C(x) \geq c_T$ "
- nemáme efektívnu procedúru na hľadanie c_T

Hoci vieme, že nerozhodnuteľných príkladov je veľa, nevieme ich hľadať

enumerovateľné vs. rekurzívne množiny

charakteristická postupnosť množiny A , $\chi = \chi_1\chi_2\dots$, $\chi_i = \begin{cases} 1, & i \in A \\ 0, & i \notin A \end{cases}$

A aj A^C RE, potom $f(i) = \chi_i$ rekurzívna a $C(\chi_{1\dots n}|n) \leq c_A$

Lemma (Barzdinova lema)

- 1 ak A je RE, χ jej charakteristická postupnosť, tak $C(\chi_{1\dots n}|n) \leq \log n + c_A$
- 2 existuje RE množina, pre ktorú $C(\chi_{1\dots n}) \geq \log n$

1 ak A je RE, χ jej charakteristická postupnosť, tak $C(\chi_{1\dots n}|n) \leq \log n + c_A$

enumerácia A pomocou PR funkcie Φ

$$A = \{x \mid \Phi(x) < \infty\}$$

ukončenie pomocou počtu jednotiek $m \leq n$

2 existuje RE množina B , pre ktorú $C(\chi_{1\dots n}) \geq \log n$

existencia B diagonalizáciou (pomocou aditívne optimálnej Φ_0)

$$\chi_i = \begin{cases} 1, & \Phi_0(i, i) = 0 \\ 0, & \Phi_0(i, i) \neq 0 \text{ alebo } \Phi_0(i, i) = \infty \end{cases}$$

Ak $C(\chi_{1\dots n}) < \log n$ potom \exists program $p : l(p) < \log n$

$$\text{problém } \Phi_0(p, p) = \chi_p$$

Example (Diofantické rovnice)

$\Delta = \Delta_1 \Delta_2 \dots$, $\Delta_i = 1$ ak i -ta rovnica má riešenie, inak 0

$$C(\Delta_{1\dots n}|n) \leq \log n + O(1)$$

Example (nestlačiteľné reťazce cez zastavenie)

$$K_0 = \{\langle x, y \rangle : \Phi_x(y) < \infty\}$$

d – počet prvkov $\in K_0$, ktoré sú menšie ako 2^n

AK poznáme d , vieme nájsť \forall výpočty $\Phi_x(y)$, $|\langle x, y \rangle| < 2^n$, ktoré zastanú

AK χ je charakteristická pre K_0 , tak $\chi_{1\dots 2^n}$ vypočítame z d : $l(p) \leq l(d) + O(1) \leq n + c$

ALE \exists enumerovateľná množina s χ , $C(\chi_{1\dots m}) \geq \log m \forall m$

\rightsquigarrow existuje konštanta c' , že program p je c' -nestlačiteľný

Algoritmická informácia o y obsiahnutá v x

$$I_c(x : y) = C(y) - C(y|x)$$

- $C(x|x) = 0 \quad I_c(x : x) = C(x)$
- \exists reťazce $C(x|n) > n$, je veľa $C(n) \geq I(n)$
- Ak $x, I(x) = n$ a n je náhodné číslo, tak

$$I_c(x : n) = C(n) - C(n|x) \geq I(n)$$

$$I_c(n : x) = C(x) - C(x|n) \leq n - n = 0$$

Theorem

$$\forall x, y \in \mathbb{N} \quad C(x, y) = C(x) + C(y|x) + O(\log C(x, y))$$

$$\leq \log C(x, y) \geq \max\{\log C(x), \log C(y|x)\} \quad \checkmark$$

pre $c \geq 0$, $C(x, y) \geq C(x) + C(y|x) - c \log C(x, y)$

kvôli sporu: $\forall c \exists x, y : C(y|x) > C(x, y) - C(x) + c \log C(x, y)$

- $A = \{ \langle u, z \rangle : C(u, z) \leq C(x, y) \}$

ak vieme $C(x, y)$, tak A je enumerovateľná

- $A_x = \{ z : C(x, z) \leq C(x, y) \}$

ak vieme $C(x, y)$, tak A_x je enumerovateľná

- pre **popis** y stačí x a poradové číslo y v A_x , $C(x, y)$

$$C(y|x) \leq l(d(A_x)) + 2l(C(x, y)) + O(1)$$

$$\rightsquigarrow d(A_x) > 2^\ell, \ell = C(x, y) - C(x) + (c - 2)l(C(x, y))$$

- **krátky popis** x

$$\left. \begin{array}{l} C(x, y) \\ \ell \end{array} \right\} u \text{ je kandidát na } x \text{ ak } A_u = \{ z : C(u, z) \leq C(x, y) \} \text{ a } 2^\ell < d(A_u)$$

U množina kandidátov, $x \in U$

$$\{\langle u, z \rangle : u \in U, z \in A_u\} \subseteq A$$

$$d(A) \leq 2^{C(x,y)+O(1)} \rightsquigarrow d(U) < \frac{d(A)}{2^\ell} \leq \frac{2^{C(x,y)+O(1)}}{2^\ell}$$

x vieme zrekonštruovať z $C(x, y), \ell$, poradia x v U

$$C(x) < 2I(C(x, y)) + 2I(\ell) + C(x, y) - \ell + O(1)$$

$$C(x) < 2I(C(x, y)) + 2I(\ell) + C(x, y) - \mathbf{C(x, y)} + \mathbf{C(x)} - \mathbf{(c + 2)I(C(x, y))} + O(1)$$

Pre veľké c dostaneme spor $C(x) < C(x)$

□

Dôsledok

Až na aditívnych $O(\log C(x, y))$ platí $C(x) - C(x|y) = C(y) - C(y|x)$
 Preto $|I_c(x : y) - I_c(y : x)| = O(\log C(x, y))$

Algoritmická prefixová zložitosť

čiasťočne rekurzívna prefixová funkcia $\Phi : \{0, 1\}^* \rightarrow \mathbb{N}$

AK $\Phi(p) < \infty$ & $\Phi(q) < \infty$ TAK p nie je vlastný prefix q

Enumerácia PR prefixových fcíí

- T_1, T_2, \dots enumerácia PR Φ_1, Φ_2, \dots konečný vstup
- T'_1, T'_2, \dots enumerácia PRP Ψ_1, Ψ_2, \dots nekonečný vstup $b_1 b_2 \dots$
- **halting input** T' zastane po dočítaní $b_1 \dots b_m$, ale pred čítaním b_{m+1}

modifikácia T na T' , vstup $b_1 b_2 \dots$

- 1 $p \leftarrow \epsilon$
- 2 dovetail $T(pq) \forall q \in \{0, 1\}^*$; nech $\Phi(q)$ zastane prvé
- 3 ak $q = \epsilon$, na výstup $\Phi(p)$ a stop
ak $q \neq \epsilon$, načítaj ďalšie b zo vstupu; $p \leftarrow pb$; choď na 2

prefixová zložitost'– vlastnosti/príklady

- existuje univerzálny $T'_0 \Psi_0$

$$C_{\psi_0}(x|y) \leq C_{\psi}(x|y) + c_{\psi}$$

Univerzálny prefix stroj $U(\langle y, \langle n, p \rangle \rangle) = T'_n(y, p)$

- $K(x, y) := K(\langle x, y \rangle)$

$$K(x, y) \leq K(x) + K(y) + O(1)$$

- $C(x|y) \leq K(x|y) \leq C(x|y) + 2 \log C(x|y) + O(1)$

$$\begin{aligned} K(x|y) &\leq C(x|y) + C(C(x|y)) + \dots + O(1) + r \\ &\leq C(x|y) + C(C(x|y)) + O(\log C(C(x|y))) + O(1) \\ &\leq C(x|y) + I^*(C(x|y)) + O(1) \end{aligned}$$

- $K(x) \leq \log^* n + n + l(n) + l(l(n)) + \dots + O(1)$

$$\begin{aligned} K(x) &\leq K(x|n) + K(n) + O(1) \\ &\leq K(x|n) + \log^* n + l(n) + l(l(n)) + \dots + O(1) \end{aligned}$$

Theorem

Platí

- $\forall n \max\{K(x) : l(x) = n\} = n + K(n) + O(1)$
- $\forall r$ je počet slov x dĺžky n , pre ktoré $K(x) \leq n + K(n) - r$, najvyššie $2^{n-r+O(1)}$

Dôkaz

- $\leq K(n)x$
 $\geq 2^n$ reťazcov dĺžky n , podľa 2. $n + K(n) - r$ nestačí
- Nech reťazec x spĺňa $K(x) \leq n + K(n) - r$
využijeme niečo, čo ešte nevieme ...

$$K(x) + K(n|x, K(x)) = K(n) + K(x|n, K(n)) + O(1)$$

Vezmeme $K(n|x, K(x)) = O(1)$ a pre $n=l(x)$ potom

$$K(x|n, K(n)) \leq n - r + O(1)$$

Len $2^{n-r+O(1)}$ to môže spĺňať.

$$C(x) = \min\{i : K(x|i) \leq i\} + O(1) = K(x|C(x)) + O(1)$$



- x^* najkratší program pre x má dĺžku $C(x)$
- $K(x|C(x)) \leq C(x) + O(1)$
- $C(x) \geq \min\{i : K(x|i) \leq i\} + O(1)$



$C(x) \leq \min\{i : K(x|i) \leq i\} + O(1)$, resp. $K(x|i) \leq i$ potom $C(x) \leq i + O(1)$

$$\begin{cases} i - 1 > l(p), & \underbrace{0\dots 01p}_i \\ i - 1 \leq l(p), & 1p \end{cases}$$

$$C(x) \leq C_T(x) + c_T \leq i + O(1)$$

Algoritmické vlastnosti C

K, C sú nerekurzívne

ak f je rekurzívna, tak $K(f(x)|x) = O(1)$

ak $f(x) \in \{0, 1\}$, tak $K(f(x)|x) = O(1)$

} $K(f(x)|x)$ hovorí niečo o f

Def.

Zložitosť $K(f)$ funkcie f je $K(f(x)|x)$

$K(K(x)|x)$

//zložitosť zložitosti

1 ak $l(x) = n$ tak $K(x) \leq n + 2 \log \log n + O(1)$

$1^{l(n)} n x$

2 $K(K(x)|x) \leq K(K(x)|n) + O(1) \leq \log n + O(1)$

ak viem n , $|n - K(x)| < n$, preto stačí povedať $d = n - |n - K(x)|$

$$\log d + 2 \log \log d + O(1) \leq \log n + O(1)$$

3 $\forall n \exists x$ dĺžky n také, že $K(K(x)|x) \geq \log n - \log \log n + O(1)$

Theorem

$\forall n \exists x$ *dĺžky n také, že* $K(K(x)|x) \geq \log n - \log \log n + O(1)$

//analogicky pre $C(C(x)|x)$

Dôkaz

- U je referenčný stroj $U(\langle y, \langle n, p \rangle \rangle) = T'_n(y, p)$
- fixneme dostatočne veľké n , všetky reťazce sú dĺžky n
- $s = \max_{I(x)=n} \min\{I(p) \mid U(p, x) = K(x)\}$

// maximálna hodnota $K(K(x)|x)$

Chceme $s \geq \log n - \log \log n + O(1)$

// $K(K(x)|x) \leq s \leq \log n + O(1)$

vhodný program p pre x - ak pre nejaké q

- $l(p) \leq s$ určite existuje
- U počíta $l(q)$ z p , ak pozná x $l(p) = K(K(x)|x) \leq s$
- U počíta x z q $l(q) = K(x)$

M_i je množina tých x , že \exists aspoň i vhodných programov pre x

$$\emptyset = M_{j+1} \subseteq M_j \subseteq \dots \subseteq M_0 = \{0, 1\}^n, \quad M_j \neq \emptyset$$

Ukážeme $l(d(M_i)) \leq l(d(M_{i+1})) + 5 \log n$, čo spolu s $j \leq 2^{s+1}$ dáva

$$s \geq \log n - \log \log n + O(1)$$

Chceme $I(d(M_i)) \leq I(d(M_{i+1})) + 5 \log n$,

Ak vieme $i, s, n, d(M_{i+1}), I(d(M_i))$

- 1 enumerujeme prvky z M_{i+1}
- 2 vygenerujeme dostatočne veľa prvkov z $M_i - M_{i+1}$
 $\forall z \in M_i - M_{i+1}$ nájdeme *všetkých* i vhodných programov
 vhodný program p , pre kt. $U(p, z)$ je minimálne, spĺňa $K(z) = U(p, z)$
 $\log d(M_i - M_{i+1}) \geq \log d(M_i) - 1$ inak platí, čo chceme, triviálne
 zoberieme $z_{\max} : K(z_{\max}) \geq I(d(M_i)) - 1$

- 3 nech $x = z_{\max}$. Potom pre popis x stačí
 diskusia $\rightsquigarrow O(1)$
 popis $d(M_{i+1}) \rightsquigarrow I^*(n) + I(d(M_{i+1}))$
 popis $I(d(M_i)) \rightsquigarrow I^*(n)$
 popis $i, n, s \rightsquigarrow I^*(n), I^*(n), I^*(\log n + 2 \log \log n)$

$$I(d(M_i)) - 1 \leq K(x) \leq 4I^*(n) + O(I^*(\log n)) + I(d(M_{i+1})) + O(1)$$

$$I(d(M_i)) \leq I(d(M_{i+1})) + 5 \log n + O(1)$$

Algoritmické vlastnosti C

Vieme,

- $C(C(x)) \leq \log n + O(1) \forall x$
- $C(C(x)|x) \geq \log n - \log \log n + O(1)$ pre niektoré x – pre aké?

$x, C(x) \geq n - k$

- $C(C(x)|x) \leq C(k) + O(1), C(k) \leq \log k + O(1)$
- ak platí veta 9, tak

$$C(k) \geq C(C(x)|x) \geq \log n - \log \log n + O(1)$$

$$\Downarrow$$

$$k = \Omega\left(\frac{n}{\log n}\right) \rightsquigarrow \boxed{C(x) \leq n - \Omega\left(\frac{n}{\log n}\right)} \quad x \text{ nie je náhodné}$$

Nech c' je konštanta.

$$C(C(x)|x) \geq \log n - \log \log n + O(1)$$

$$\Downarrow$$

$$\forall k \leq n : C(k|x) \leq c', C(x) \notin [k - \delta, k + \delta], \delta = O(n/\log n)$$

Lemma

Na rozpoznanie xx^R treba rádovo n^2 krokov

- x dĺžky n tak, že $C(x|T, n) \geq n$
- $l(T)$, $l(pp)$ dĺžka popisu stroja a prechodovej postupnosti
- výpočet na $x0^{2n}x^R$
 ak $\forall l(pp(n+1)), \dots, l(pp(2n)) \geq \frac{n}{2l(T)}$ ✓
 ak $\exists n < i_0 \leq 2n : l(pp(i_0)) < \frac{n}{2l(T)}$, spravíme **kratší popis x** :
 $\forall y \in \{0, 1\}^n$ skúšame $y0^{2n}$
 $C(x|T, n) \leq l(pp(i_0)) + O(1) \leq n/2 + O(1)$? pozícia ?