Kolmogorovská zložitosť

21.11.2013

kompresia jazykov

- $\Sigma = \{0,1\}$. $\mathbf{f} : \Sigma^* \to \Sigma^*$ je kompresia jazyka $L \subseteq \Sigma^*$ ak je one-to-one na L a až na konečne veľa $x \in L$ platí $l(f(x)) < l(x)$
- Funkcia f, $f^{-1} : f(L) \to L$ pričom $\forall x \in L\ f^{-1}(f(x)) = x$. Jazyk je komprimovateľný v čase $T(n)$, ak existuje kompresia f pre L počitateľná v čase $T(n)$ ktorej inverzná f^{-1} je tiež počitateľná v $T(n)$
- Kompresia f optimálne komprimuje L ak $\forall x \in L$ dĺžky n,

$$l(f(x)) \leq \lceil \log \sum_{i=0}^{n} d(L^{=i}) \rceil$$

- prirodzená kompresia — ranking

$$r_L : L \to N \qquad \forall x \in L : r_L(x) = \text{ index } x \text{ v } L$$

Theorem

Nech A množina. Potom existuje konštanta c a polynóm p tak, že

$$\forall x \in A^{=n} \quad CD^p(x|A^{=n}) \leq 2\log d(A^{=n}) + 2\log n + c$$

$A, d = d(A^{=n})$

Lemma

Nech $S = \{x_1, \ldots, x_d\} \subseteq \{0, \ldots, 2^n - 1\}$. Potom $\forall x_i \in S$ existuje prvočíslo $p_i \leq 2dn$ tak, že $\forall j \neq i \; x_i \neq x_j \mod p_i$. ↪

- $A = S$, $\forall x \in A$ máme p_x
- CD program p pre x

  vstup y

  if $y \notin A^{=n}$ reject y else if $y = x \mod p_x$ then accept y else reject y

  $l(p) = l(p_x) + l(x \mod p_x) + O(1)$

  $p \leq 2\log d(A^{=n}) + 2\log n + O(1)$ rovnako dlhé $p_x, x \mod p_x$

  akceptuje **len x** Thm.

Dôsledok

$A \in P$. Potom existuje polynóm p $\forall x \in A^{=n}$

$$CD^p(x|n) \leq 2\log d(A^{=n}) + 2\log n + O(1)$$

$S = \{x_1, \ldots, x_d\} \subseteq \{0, \ldots, 2^n\text{-}1\}$. $\forall x_i \in S$ existuje prvočíslo $p_i \leq 2dn$ tak, že $\forall j \neq i$ $x_i \neq x_j \mod p_i$.

- $N = 2^n$
- $\forall x_i \neq x_j \in S$ existuje nanajvýš $\log_c N = \log N / \log c$ rôznych prvočísel p takých, že $c \leq p \leq 2c$ a $x_i = x_j \mod p$
- len $d-1$ dvojíc obsahuje x_i, preto medzi $\frac{(d-1)\log N}{\log c} + 1$ prvočíslami $\exists p_i$, že $\forall j,\ i \neq j,\ x_i \neq x_j \mod p_i$
- PrimesNumberThm: $\pi(n) \backsim \frac{n}{\ln n}$,
  resp. $\pi(n) = \frac{n}{\ln n} + \frac{n}{(\ln n)^2} + \frac{2!n}{(\ln n)^3} + \frac{3!n}{(\ln n)^4} + O(\frac{n}{(\ln n)^5})$
  - $\hookrightarrow$ medzi $c, 2c$ je aspoň $\frac{c}{\log c}$ prvočísel
  - $\hookrightarrow$ ak $c > (d-1)\log N$ tak $p_i \leq 2d \log N = 2dn$ $\square$

Theorem

$\exists$ polynóm $p(n)$ taký, že $\forall A$ a dostatočne veľké $n \in \mathbb{N}$, ak $x \in A^{=n}$, tak

$$CD^p(x|A^{=n}, s) \leq \log d(A^{=n}) + \log\log d(A^{=n}) + O(1)$$

pričom $A^{=n}$ je dané ako orákulum, dĺžka $s \sim n \log d(A^{=n})$

↪

$h : \Sigma^n \to \Sigma^m$ lineárna transformácia daná náhodnou binárnou maticou $R = \{r_{i,j}\}$
$\forall x \in \Sigma^n \; Rx = y \in \Sigma^m : \quad y_i \equiv (\sum_j r_{i,j} x_j) \mod 2$

H..množina kódovacích funkcií.

- Nech $B, C \subseteq \Sigma^n, x \in \Sigma^n$. h separuje x v B ak $\forall y \in B, y \neq x, h(y) \neq h(x)$
- h separuje C v B ak $\forall y \in C$ h separuje y v B
- H separuje C v B ak $\forall x \in C \; \exists h \in H$ tak, že h separuje x v B.

Lemma (o kódovaní)

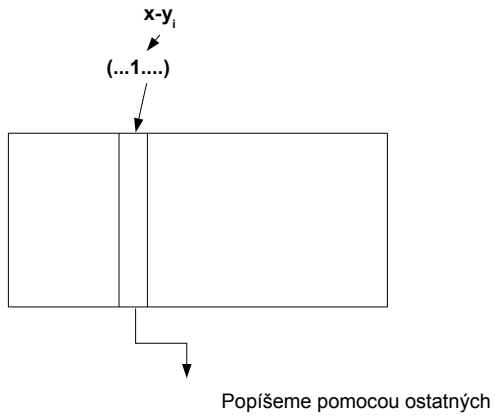
$B \subseteq \Sigma^n, d(B) = k, m = 1 + \lceil \log k \rceil$. Existuje množina H m náhodných lineárnych transformácií $\Sigma^n \to \Sigma^m$ takých, že H separuje B v B.

$B \subseteq \Sigma^n, d(B) = k, m = 1 + \lceil \log k \rceil$. Existuje množina H m náhodných lineárnych transformácií $\Sigma^n \to \Sigma^m$ takých, že H separuje B v B

fixneme náhodný reťazec s dĺžky nm^2 : $C(s|B, P, m, n) \geq l(s)$,
P program na popis s; //s reprezentuje H

H separuje B v B SPOROM:
$\exists x \in B$ že žiadne $h \in H$ neseparuje x v B:

$\exists y_1, \ldots, y_m \in B \; h_i(x) = h_i(y_i) \;\Rightarrow\; h_i(x - y_i) = 0$



krátky popis s

- index x v B $\log k$
- indexy $y_1, \ldots, y_m$ $m\lceil \log k \rceil$
- $h_1, \ldots, h_m$ bez "redundantných" stĺpcov $nm^2 - m^2$

$$C(s|B, P, m, n) \leq m^2 n - m^2 + m\lceil \log k \rceil + \lceil \log k \rceil \leq m^2 n - 1$$

◇

$\exists$ polynóm $p(n)$ taký, že $\forall A$ a dostatočne veľké $n \in \mathbb{N}$, ak $x \in A^{=n}$, tak

$$CD^p(x|A^{=n}, s) \leq \log d(A^{=n}) + \log\log d(A^{=n}) + O(1)$$

pričom $A^{=n}$ je dané ako orákulum, dĺžka $s \sim n \log d(A^{=n})$

— H z kódovacej lemy, s popisuje H; $m = 1 + \lceil \log d(A^{=n}) \rceil$

— $\forall z \in A^{=n} \exists h_i \in H$ separujúce z v $A^{=n}$

— program, kt. z $ih_i(z)$ akceptuje z $\qquad$ $A^{=n}, s$ poznáme

- over, či $x \in A^{=n}$
- ak $x \in A^{=n}$, $ih_i(z) \rightsquigarrow i, h_i(z)$, nájde h_i
- vypočíta $h_i(x)$
- akceptuje $z \quad \Leftrightarrow \quad h_i(x) = h_i(z)$

$$CD^p(z|A^{=n}, s) \leq m + \log m + O(1) = \log d(A^{=n}) + \log\log d(A^{=n}) + O(1)$$

□

Theorem

Existuje polynóm p, že $\forall A$ a dostatočne veľké n ak $x \in A^{=n}$, tak

1. $C^p(x|A^{=n}, s, NP^A) \leq \log d(A^{=n}) + \log\log d(A^{=n}) + O(1)$
2. $C^p(x|A^{=n}, \Sigma_2^{P,A}) \leq \log d(A^{=n}) + \log\log d(A^{=n}) + O(1)$

keď $A^{=n}$, NP^A-úplná mn., $\Sigma_2^{P,A}$-úplná mn. sú orákulá, s je reťazec dĺžky $\sim n \log d(A^{=n})$

$$\Sigma_2^p = NP^{NP} \begin{cases} \Sigma_1^P = & NP \\ \Sigma_{i+1}^P = & NP^{\Sigma_i^P} \end{cases}$$

(1.) $C^p(x|A^{=n}, s, NP^A) \leq$ $CD^q(x) + O(1)$
$CD^p(x|A^{=n}, s) \leq \log d(A^{=n}) + \log\log d(A^{=n}) + O(1)$

(2.) pomocou orákula **generujeme s**

je $s_1 \ldots s_i 1$ prefix H kt. separuje $A^{=n}$ v $A^{=n}$?

□

riedka množina A — $d(A^{=n}) \leq n^c + c$

Theorem

Existuje riedky jazyk L taký, že ak L skomprimujeme pomocou pravdepodobnostného polynomiálneho stroja s orákulom pre L, tak kompresná funkcia mapuje reťazce dĺžky n na reťazce dĺžky $n - O(\log n)$.

- L obsahuje jediné nestlačiteľné slovo dĺžky $2^{2^{\cdots^{\mathbf{2}}}}$, $\mathbf{I(x) \leq C(x)}$
- T pravdepodobnostný orákulovský stroj, kt. rozpoznáva L v čase n^k
- kvôli SPORu: existuje kompresia f : $I(f(x)) \leq n - c_1 - (k + c_1)\log n$. Ukážeme, že T nemôže zrekonštruovať x z $f(x)$ s pravdep. aspoň $1/2$

  Ak áno:

  – R je množina rozhodnutí korektnej odpovede — $1/2$ z 2^{n^k}

  – $r \in R$ s $\mathbf{C(r|x) \geq I(r) - 1}$ — existuje!

  – $\mathbf{C(x) \geq I(x)}$, $C(r) \leq I(r) + O(1)$, symetria KZ — $\mathbf{C(x|r) \geq n - c_2 \log n}$

krátky popis $x|r$

informácia potrebná pre odpovede orákula keď T počíta pri vstupe $f(x)$

$$(k+2)\log n + c_1$$

simulácia T so vstupom $f(x)$ pri znalosti výsledkov hádzania r

- diskusia
- $2\log n$ bitov pre slová dĺžky $< n$
- T sa pýta $?x \in L?$ - $k\log n$ bitov na určenie času, keď sa to pýta

$$C(x|r) \leq c_1 + 2\log n + k\log n + l(f(x))$$

pre $c_1 > c_2 + 2$ dostaneme spor $//l(f(x)) \leq n - c_1 - (k+c_1)\log n$

□

Theorem

Ak $L \in 1DLOG$, potom r_L môžme vypočítať v P.

- T rozpoznáva L
- $L_x = \{y \in L : y \leq x\}$; zrejme $r_L(x) = d(L_x)$
- x do vnútornej pamäte a rozpoznávame L_x jednosmerne v DLOG s polynomiálnym počtom stavov — $\mathbf{T_x}$
- $G_x = (V_x, E_x)$ — graf konfigurácií pre T_x;
  $V_x = \{[hlava, stav, paska]\}$
  E_x— zachytáva možné prechody

$\hookrightarrow$ $r_L(x)$ je počet akceptujúcich ciest v G_x

Nech $L \subseteq \Sigma^*$ je regulárny, $L_x = \{y \mid xy \in L\}$. Potom existuje konštanta c, že $\forall x$: ak y je n-té slovo v L_x, tak $C(y) \leq C(n) + c$

- stav DKA po dočítaní x
- n
- diskusia ◇

$\Sigma^* = \{y_1, y_2, \ldots\}$;
pre $L \subseteq \Sigma^*, x \in \Sigma^*$ je $\mathbb{X} = X_1 X_2 \ldots$ charakteristická postup. L_x: $X_i = 1 \Leftrightarrow xy_i \in L$.

Theorem

Nech $L \subseteq \Sigma^*$. Existuje konštanta c_L, že nasledujúce podmienky sú ekvivalentné

1. *L je regulárny*
2. $\forall x \in \Sigma^* \forall n \in \mathbb{N}: \; C(\mathbb{X}_{1..n}|n) \leq c_L$
3. $\forall x \in \Sigma^* \forall n \in \mathbb{N}: \; C(\mathbb{X}_{1..n}) \leq c_L + C(n)$
4. $\forall x \in \Sigma^* \forall n \in \mathbb{N}: \; C(\mathbb{X}_{1..n}) \leq \log n + c_L$

Ak $\forall x \in \Sigma^* \forall n \in \mathbb{N}: \; C(\mathbb{X}_{1..n}) \leq \log n + c_L$ tak L je regulárny

Lemma

$\forall c$ *existuje konečne veľa postupností* $w \in \{0,1\}^*$ *takých, že*
$\forall n \; C(w_{1..n}) \leq \log n + c$

$\hookrightarrow$ pravá kongruencia konečného indexu $\sim$ $\quad x \sim x' \Leftrightarrow \mathbb{X} = \mathbb{X}'$
$\hookrightarrow$ konečný automat □

k dôkazu lemy:

- $A_n = \{x \in \{0,1\}^n : C(x) \leq \log n + c\}$;
  $A = \{w \in \{0,1\}^\infty : \forall n \; C(w_{1..n}) \leq \log n + c\}$
- ak $d(A_n) \leq c'$ pre nekonečne veľa n, tak $d(A) \leq c'$
- fixnime $\ell \in \mathbb{N}$: y dĺžky $2\ell + c + 1$, $C(y) \geq 2\ell + c + 1$
- vezmime i maximálne také, že $y = mn$, $l(m) = i$, $m \leq d(A_n)$
  nech $y = sr$, $l(s) = i + 1$
- pomocou programov $l(p) \leq \log n$ vymenovávanie A_n; y_0- m-tý v A_n;
  rekonštrukcia $y = mn$ z $y_0 \hookrightarrow n = l(y_m), m : y_0 = y_m$

$$l(n) + l(m) = 2\ell + c + 1 \leq C(y) \leq C(y_0) + O(1) \leq \log n + c + O(1)$$

on-line TS - pred pohybom hlavy na vstupe akcept/reject na príslušný prefix

$L = \{y\#x_1 * x_2 * \ldots x_k : \ \exists i \ y = ux_i^R v\}$

Theorem

Viacpáskový TS akceptuje L online v čase $\Omega(n^2/\log n)$

Lemma

Nech $n = l(x)$, p program. Ak $C(x|n,p) \geq n$, potom žiaden podreťazec dĺžky $> 2\log n$ sa v x nevyskytuje viac ako raz.

Lemma

Ak sa v reťazci neopakuje reťazec dĺžky m, tak je jednoznačne určený množinou svojich podreťazcov dĺžky $m+1$.

- $\hookrightarrow$ $m = 3\log n \quad l(x_l) = m$
- $\hookrightarrow$ pre y, $l(y) = n$, $C(y) \geq n$ vyrobíme vstup taký, že žiadne x_l nie je reverzom podreťazca v y ale jeho spracovanie vyžaduje $n\epsilon$ krokov;
  pre $k = \Omega(n/\log n)$ vynútime $\Omega(n^2/\log n)$ krokov

Nech $n = l(x)$, p program. Ak $C(x|n,p) \geq n$, potom žiaden podreťazec dĺžky $> 2\log n$ sa v x nevyskytuje viac ako raz.

- $x = uvw$, $l(v) > \log n$, v sa v uv vyskytuje presne dvakrát
- povieme i, j– indexy začiatku v v x $2\log n$
- povieme uw $//l(v) = n - l(uw)//$ $l(uw)$

$$C(x|p,n) \leq n - 2\log n + \log n(n-1) < n$$

Ak sa v reťazci neopakuje reťazec dĺžky m, tak je jednoznačne určený množinou $S_{x,m+1}$ svojich podreťazcov dĺžky $m+1$.
$a, b \in \{0,1\}, u, v, w \in \{0,1\}^*$

- prefix ua je jednoznačne určený podmienkou $\forall b$, $bu \notin S_{x,m+1}$
- $\forall$ prefix vw, $l(w) = m$ existuje **práve jedno** $b \in \{0,1\}$: $wb \in S_{x,m+1}$ a vwb je prefix x

$\hookrightarrow$ rekonštrukcia x zo znalosti $S_{x,m+1}$

$m = 3 \log n \quad l(x_l) = m$; pre y, $l(y) = n$, $C(y) \geq n$ vyrobíme vstup taký, že žiadne x_l nie je reverzom podreťazca v y ale jeho spracovanie vyžaduje $n\epsilon$ krokov

majme to po $y\#x_1 * \ldots * x_{i-1}*$

1. ak $i - 1 = k$ ✓
2. ak také x_i neexistuje, vyrobíme krátky popis na generovanie $S_{y,m+1}$ //a teda y
   - diskusia $O(1)$
   - obsah pracovných pások do vzdialenosti $n\epsilon$ od hláv v čase t_0, keď T ukončil spracovanie prefixu $y\#x_1 * \ldots x_{i-1}*$ $O(n\epsilon)$
   - polohy hláv v čase t_0 $O(\log n\epsilon)$
   - popis $T, n, stav$ v čase t_0 $O(1) + \log n$

$$C(y) \leq O(1) + O(n\epsilon) + O(\log n\epsilon) + logn < n$$

□

Pre FA je $(k+1)$ hláv viac ako k

$$L_b = \{w_1\# \cdots \#w_b\#\#w_b\# \cdots \#w_1 \mid w_i \in \{0,1\}^*\}, b = \binom{k}{2} + 1$$

1. $L_b \in (k+1)DFA$
2. $L_n \notin (k)NFA$ SPOROM:
   - vezmeme nestlačiteľné w, $C(w) \geq l(w), l(w) \ggg \log l(w)$
   - $w = w_1 \dots w_b \quad \rightsquigarrow \quad$ korektný vstup I
   - NKA kontroluje w_i ak súčasne jedna hlava na ľavej a jedna na pravej kópii w_i

Lemma

M musí skontrolovať každé w_i

- $[q, p(h_1), \dots, p(h_k)]$ — postupnosť hlavových konfigurácií v okamihoch, keď jedna z hláv prvýkrát prichádza na p $\qquad O(k \log(l(I)) + l(M)) = O(\log l(w))$
- r_l, r_r-L a R pozícia pravej kópie w_i, $p(r_l), p(r_r)$ príslušné prechodové postupnosti

$\hookrightarrow$ rekonštrukcia w_i z

- popis $w - w_i$
- popis $p(r_l), p(r_r)$

$$C(w_i|w - w_i) \leq O(\log l(w))$$

$$C(w) \leq l(w) - l(w_i) + O(\log l(w)) < l(w)$$