Kolmogorovská zložitosť

5.12.2013

$age(x) = \min_p\{2^{l(p)}t : U(p) = x \text{ v priebehu t krokov}\}$

Def. (Kt zložitosť)

UTS monotonne skenuje začiatok p kým vypíše x, $t(p, x)$je počet krokov, kým vypísal x

$$Kt_U(x) = \min_p\{l(p) + \log t(p, x)\}$$

$T_1, T_2, \ldots$ enumerácia prefixných TS, $\Phi_1, \Phi_2, \ldots$ príslušné čiastočne rekurzívne funkcie.

Φ čiastočne rekurzívna a $\Phi(y) = x$ potom y je svedok pre x

// $h(\Phi)$ - obor hodnôt Φ

Def.

Algoritmus A invertuje problém Φ ak pre dané $x \in h(\Phi)$ vypočíta Φ svedka y pre x a overí, že $\Phi(y) = x$. Pre $x \notin h(\Phi)$ diverguje.

Example

- splniteľná formula – chceme spĺňajúce priradenie
- k-zafarbiteľný graf – chceme to zafarbenie
- ...

Lemma

Ak existuje algoritmus A, ktorý invertuje Φ v čase $t(n)$, potom existuje algoritmus, ktorý invertuje Φ v čase $c_A t(n)$.

SIMPLE simuluje

- T_1 v každom druhom
- T_2 v každom druhom z toho, čo ostalo
- T_3 v každom druhom z toho, čo ostalo
- ...

1	⊔	1	⊔	1	⊔	1	⊔	1	⊔	1	⊔	1	⊔	1	⊔	1
1	2	1	⊔	1	2	1	⊔	1	2	1	⊔	1	2	1	⊔	1
1	2	1	3	1	2	1	⊔	1	2	1	3	1	2	1	⊔	1
1	2	1	3	1	2	1	4	1	2	1	3	1	2	1	⊔	1

T_k invertuje Φ v čase t, potom SIMPLE to spraví v

$$2^k t + 2^{k-1}$$
$$c_k = 2^{k+1}$$

Theorem

Ak existuje algoritmus A, ktorý invertuje Φ v čase $t(n)$, potom metóda SEARCH invertuje Φ v čase $c_A t(n)$.

$Kt'(w|x,\Phi) = \min\{l(p) + \log t(p,x)\}$ – pre dané x program p vypíše w a otestuje, či $\Phi(w) = x$ v čase $t(p,x)$

SEARCH — univerzálny prefixový U beží postupne na programoch p dĺžky $< i$ počas $2^i 2^{-l(p)}$ krokov a hľadá $\Phi(w) = x$

Fakt

$\forall w : Kt'(w|x,\Phi) \leq k$ *sú testované v čase* 2^{k+1} $\hookrightarrow$

$\hookrightarrow$ $m = \min\{Kt'(w|x,\Phi) : w$ je Φ-svedok pre $x,\ l(x) = n\}$.

$\hookrightarrow$ prefixný T invertuje Φ v čase $t(n)$ $\quad m \leq Kt'(T|x,\Phi)$

$\hookrightarrow$ SEARCH invertuje počas 2^{m+1} krokov, $Kt'(T|x,\Phi) \leq K(T) + \log t(n)$

$\hookrightarrow 2^{K(T)+1} t(n)$ krokov

$$c = 2^{K(T)+1}$$

$\forall w : Kt'(w) \leq k$ sú testované v čase 2^{k+1}

- $Kt'(w|x,p) \leq i \hookrightarrow l(p) + \log t(p,x) \leq i \hookrightarrow t(p,x) \leq 2^{i-l(p)}$
- Kraft $\sum 2^{-l(p)} \leq 1$

$$\hookrightarrow \sum_{1 \leq i \leq k} \sum_{0 < i-l(p)} 2^{i-l(p)} \leq \sum_{U(p)<\infty} 2^{-l(p)} \sum_{1 \leq i \leq k} 2^i \leq 2^{k+1}$$

□

Example

T_k invertuje v čase $t(n)$

SIMPLE $2^{k+1}t(n)$

SEARCH $2^{K(T)+O(1)}t(n) = O(k(\log k)^2)t(n)$

ak $K(T_k) = \log\log k$, čas pre SEARCH je len $O((\log k)t(n))$

SEARCH ? SIMPLE

$t(n)$ časovo konštruovateľná v $O(t(n))$

- $P^*(x)$= súčet pravdepodobností prvkov $\leq x$; počitateľná v $t(n)$ ak $\exists T : \forall x, k$ vypočíta y v $t(l(x)+k)$: $|P^*(x) - y| \leq 1/2^k$
- $P(x) = P^*(x) - P^*(x-1)$
- $K^t(x) = \min\{l(p) : \ U(p) = x$ v t(n) krokoch $\}$
- $m^t(x) = 2^{-K^t(x)}$, $m^{*t}(y) = \sum_{y \leq x} m^t(x)$

Theorem

Rozdelenie $m^{t'}$, kde $t'(n) = O(nt(n)\log(nt(n)))$, je univerzálne pre pravdepodobnostné mass funkcie P v čase t počítateľného rozdelenia P^*; $m^{t'}$ multiplikatívne dominuje P v nasledujúcom zmysle:
$\exists c_p : c_p m^{t'}(x) \geq P(x)$, pričom $\log c_p = K^{t'}(P^*) + O(1)$ závisí od P^*,ale nie t, x.

Tvrdenie Ak P^* je počitateľné v $t(n)$, tak existuje c_P

$$K^{t'}(x) \leq \log \frac{1}{P(x)} + \log c_P$$

- $[0,)$ rozdelíme na podintervaly tak, že kód $c(x)$ zaberá $I_x = [P^*(x-1), P^*(x))$
- binárny interval určený reťazcom r je $\Gamma_r = 0.r, 0.r + 2^{-l(r)}$
- ak Γ_r je najväčší binárny interval obsiahnutý v I_x, tak $c(x) \leftarrow r$
- $|\Gamma_r| \geq \frac{|I_x|}{4} \Longrightarrow l(c(x)) \leq \log 1/P(x) + 2$

dekódovanie

$k \leftarrow 1$

doubling

repeat $k \leftarrow 2k$ until $\Gamma_{c(x)} \subseteq [P^*(k-1), P^*(k))$

$l \leftarrow k/2;\ u \leftarrow k$

binarySearch

$m \leftarrow (u+l)/2$

if $\Gamma_{c(x)} \subseteq [P^*(m-1), P^*(m))$

then $x \leftarrow m$

else $\begin{cases} u \leftarrow m, & \Gamma_{c(x)} \text{ naľavo od } P^*(x-1); \\ l \leftarrow m, & \Gamma_{c(x)} \text{ napravo od } P^*(x) \end{cases}$

$\sum_{i=0}^{n} O(t(i)) = O(nt(n))$

rekonštrukcia x

- diskusia $O(1)$
- program q na výpočet $P^*(x)$ v $t(n)$
- $c(x)$

čas

$O(nt(n))$ //univerzálny TS v $t'(n) = O(nt(n)\log(nt(n)))$

KZ

- $K^{t'}(x) \leq l(c(x)) + \underbrace{l(q) + O(1)}_{\log c_P}$
- $K^{t'}(x) \leq \log 1/P(x) + \log c_P$
- q v $t(n)$, q' v $t'(n)$ $\quad l(q') \leq l(q)$, $\mathbf{l(q') = K^{t'}(P^*)}$
- $\log c_P = l(q') + O(1) = K^{t'}(P^*) + O(1)$ □

Logická hĺbka – počet krokov na ceste od pôvodu k objektu $\leftrightsquigarrow$ čas algoritmu na generovanie objektu z kratšieho popisu

x^* najkratší samoodeľujúci popis

1. počet krokov na výpočet x z x^* nie je stabilné

   pár bitov navyše, podstatné zníženie času

2. relaxujeme na minimum $\rightsquigarrow$ skoro minimálny program
   x má hĺbku d s toleranciou 2^{-b} ak x vypočítame v d krokoch z p, $|p| \leq |x^*| + b$

   $\hookrightarrow \quad \frac{2^{-l(p)}}{2^{-K(x)}} \geq 2^{-b}$

   stabilné, nevyhovujúce

$Q_U(x) = \sum_{U(p)=x} 2^{-l(p)}$

$$\log \frac{1}{Q_U(x)} = \log \frac{1}{m(x)} = K(x)$$

hĺbka reťazca x na hladine významnosti $\varepsilon = 2^{-b}$ je

$$depth_\varepsilon(x) = \min \left\{ d : \frac{Q_U^d(x)}{Q_U(x)} \geq \varepsilon \right\}$$

$Q_U^d(x) = \sum_{U^d(p)=x} 2^{-l(p)}$ $\quad$ $U_d(p) = x$ U po nanajvýš d krokoch zastane.

x je (d,b)-deep, ak $d = depth_\varepsilon(x)$, $\varepsilon = 2^{-b}$

b-nestlačiteľný reťazec: $K(x) > l(x) - b$

Theorem

Reťazec x je (d,b)-deep(b s presnosťou $K(d) + O(1)$) $\iff$ d je minimálny počet krokov potrebný na generovanie x b-nestlačiteľným programom

↪

chceme $\frac{1}{2^{b+K(d)+O(1)}} \underbrace{\leq}_{\Longleftarrow} \frac{Q_U^d(x)}{Q_U(x)} \underbrace{\leq}_{\Longrightarrow} \frac{1}{2^{b-O(1)}}$

d - čas potrebný na generovanie x b-nestlačiteľným programom

//menej ako d krokov, b-stlačiteľný program p

- $\forall p\ \exists p' : U(p') = p,\ l(p') \leq l(p) - b$
- q: $U(p') \rightsquigarrow p, U(p) \rightsquigarrow x$
- $l(q) \approx l(p) - b + O(1)$
- $\alpha = Q_U(x) - \sum_{U(q)=x} 2^{-l(q)} \geq 0$

$$\Longrightarrow \quad \frac{Q_U^d(x)}{Q_U(x)} = \frac{\sum_{U^d(p)2^{-l(p)}}}{\alpha + \sum_{U(q)=x} 2^{-l(q)}}$$

$$\leq \frac{\sum_{U^d(p)2^{-l(p)}}}{\sum_{U(q)=x} 2^{-l(q)}}$$

$$\leq \frac{\sum_{U^d(p)2^{-l(p)}}}{\sum_{U(q)=x} 2^{-(l(p)-b+O(1))}} \leq \frac{1}{2^{b-O(1)}}$$

$\Longleftarrow$ SPORom

$// \frac{1}{2^{b+K(d)+O(1)}} > \frac{Q_U^d(x)}{Q_U(x)}$

- x^*, d^*
  $\rightsquigarrow$ vymenovávame A = samoodd. programy generujúce x v čase max d
  $l(q) = K(x) + K(d) + O(1)$

- $\sum_{p \in A} 2^{-l(p)} = Q_U^d(x)$

- platí $Q_U(x) = 2^{-K(x)+O(1)}$

(*) $\sum_{p \in A} 2^{-l(p)} = Q_U^d(x) < \frac{Q_U(x)}{2^{b+K(d)+O(1)}} = \frac{2^{-K(x)+O(1)}}{2^{b+K(d)+O(1)}} = 2^{-K(x)-b-K(d)-O(1)}$

(**) B prefix-free s $\sum_{x \in B} 2^{-l(x)} < 2^{-m}$, enumerovateľná programom s. Potom B komprimovaná o $m - l(s) - O(1)$ bitov $\hookrightarrow$

- $B = A, s = q, m = K(x) + b + K(d) + O(1)$, (*) , (**)
  $\forall p \in A$ môže byť komprimovaný o $K(x) + b + K(d) + O(1) - l(q) - O(1) > b$ bitov
  SPOR s x je (d, b)-deep

B prefix-free s $\sum_{x \in B} 2^{-l(x)<2^{-m}}$, enumerovateľná programom s. B komprimovaná o $m - l(s) - O(1)$ bitov

$\sum_{x \in B} 2^{-l(x)} 2^m < 1 \quad \rightsquigarrow$ Shannon-Fano: $|c(x)| \leq l(x) - m + 2 + \quad l(s) + O(1)$

◇

reťazec x je d-shallow ak nie je $(d+1, b)$-deep. $n + O(1)$ shallow je shallow.

- náhodný reťazec je shallow
- 1^n je shallow

Theorem

Nech $R = \{x : C(x) \geq l(x) - \log^2 l(x)\}$. *Potom* $BPP^R = P^R$

konštrukcia x ∈ R, l(x)=m //induktívne

$x = \epsilon$

repeat $\forall y, l(y) = \log m$

if $xy \in R$ then $x \leftarrow xy$

until $l(x) \geq m$‘

y existuje

$x \in R, C(y|x) \geq l(y)$, potom symetria informácie:

$$\begin{aligned} C(xy) &\geq C(x) + C(y|x) - \text{clog}(C(xy)) \\ &\geq C(x) + C(y|x) - \text{clog } l(xy) \\ &\geq l(x) - \log^2 l(x) + l(y) - \text{clog } l(xy) \\ &\geq l(xy) - \log^2 l(xy) \end{aligned}$$

n^2 krát BPP výpočet dĺžky n + väčšina

$\rightsquigarrow$ pravdepodobnosť chyby $\leq 2e^{-O(n^2)}$

$\rightsquigarrow \leq 2^M 2^{-O(n^2)}$ reťazcov s chybou

$n^2 \leq M \leq n^3$ je počet "chybných" postupností

$\rightsquigarrow$ KZ chybných postupnost$\leq M$-$O(n^2)$

$C_{AB}(x \leftrightarrow y)$=$\min\{|p| : A(p, y) = x, B(p, x) = y\}$ informačná vzdialenosť

- $C_{AB}(x \leftrightarrow y) \geq \max\{C(x|y), C(y|x)\} + O(\log \max\{C(x|y), C(y|x)\})$

Dôkaz pomocou online farbenia grafov (susediace hrany majú rôznu farbu)

hra A a B

A generuje hrany pri zachovaní $d(G) \leq d$

B určuje farby
greedy - najmenšia možná farba $\rightsquigarrow 2d - 1$ farieb stačí

konštrukcia grafu G_k pre $k = \max\{C(x|y), C(y|x)\}$

vrcholy binárne reťazce dĺžky max k

hrany $(x, y) \in E \Leftrightarrow (C(x|y) \leq k \wedge C(y|x) \leq k)$

počet hrán, $d(G_k) \leq 2^{k+1} - 1$, počet farieb $\leq 2^{k+2} - 3$

algoritmus T: vstup k, výstup: postupnosť hrán grafu G_k a ich farieb

algoritmus A

- pozná y, vstup $\bar{k}i$
- simulácia $T(k)$, výstup $x :< (x, y), i >$ vo výstupe T

$$\begin{aligned}(x \leftrightarrow y) \leq C_{AA}(x \leftrightarrow y) &\leq |\bar{k}i| = 2\log k + 1 + k \\ &= O(\log \max\{C(x|y), C(y|x)\} + \max\{C(x|y), C(y|x)\})\end{aligned}$$

□