Kolmogorovská zložitosť

10.10.2013

párovacia funkcia

$$\langle x, y\rangle = y + (x + y + 1)(x + y)/2$$
$$\langle x, y, z\rangle = \langle x, \langle y, z\rangle\rangle$$

$\mathcal{B}^*$ $\{\epsilon, 0, 1, 00, 01, 10, 11, \ldots\}$

reťazce–prirodzené čísla, $\mathcal{B}^* \leftrightarrow \mathcal{N}$

$$\{(\epsilon, 0), (0, 1,), (1, 2), (00, 3), (01, 4), \ldots\}$$

//nerozlišujeme medzi reťazcom 01 a číslom 4

//niekedy medzi reťazcom 100 a číslom 4

dĺžka reťazca w, kardinalita množiny A: $l(w), d(A)$

$D : \{0, 1\}^* \to \mathcal{N}$ $D(y) = x$, y je kód, x je vzor

prefixný kód $\mathbf{E_i(x)} = \begin{cases} 1^x0, & \text{i=0} \\ E_{i-1}(l(x))x, & \text{i>0} \end{cases}$

$\overline{x} = E_1(x) = E_0(l(x))x = 1^{l(x)}0\mathbf{x} \quad \rightsquigarrow \quad l(E_1(x)) = 2l(x)+1$

$E_1(100000001011) = 1^{12}0100000001011,\ l(x) = 25$

$E_1(x) = 11111000010 \quad \rightsquigarrow x = 00010$

$E_2(x) = \overline{l(x)}x = 1^{l(l(x))}0l(x)\mathbf{x} \quad \rightsquigarrow l(E_2(x)) = l(x) + 2l(l(x)) + 1$

$E_2(100000001011) = E_1(l(100000001011))100000001011 =$

$111101100100000001011,\ l(x) = 21$

$E_2(x) = 11101111000001 \quad \rightsquigarrow$ x=1000001

$\overline{x}y = 1110\mathbf{110}11 \quad \Rightarrow \quad x = 110,\ y = 11$

$\overline{x}\overline{y}z = 1110\mathbf{110}1110\mathbf{0}1\mathbf{0}11 \quad \Rightarrow \quad x = 110,\ y = 010,\ z = 11$

!!!viem, aké kódovanie používam!!!

popis objektu // má zmysel, ak z neho objekt vieme zrekonštruovať

- enumerácia objektu x z množiny S prirodzeným číslom $n(x)$
- $p \in \mathbb{N}$, potom $l(p)$ označuje dĺžku binárneho zápisu p

$$\mathbf{C_f(x) = \min_p\{l(p) \mid f(p) = n(x)\}}$$

zložitosť objektu x vzhľadom k metóde f

$\mathbf{p}$ $\sim$ program $\mathbf{f}$ $\sim$ počítač

- **f (aditívne) minorizuje g** ak $\exists c\ \forall x : \mathbf{C_f(x) \leq C_g(x) + c}$

  ak uvažujeme r metód $f_1, \ldots, f_r$, na identifikáciu metódy stačí povedať $\log r$ bitov

- metódy sú **ekvivalentné** ak sa navzájom aditívne minorizujú

- C je trieda čiastočných funkcií na Z^+. Funkcia **f je univerzálna (aditívne optimálna) pre C** ak

$$\mathbf{f \in C} \quad \& \quad \left[\forall \mathbf{g \in C}\ \exists \mathbf{c_{f,g}} :\ \mathbf{C_f(x) \leq C_g(x) + c_{f,g}}\right]$$

- univerzálne funkcie sú "ekvivalentné:"

$$|C_f(x) - C_g(x)| \leq \max\{c_{f,g}\ c_{g,f}\}$$

Example

Trieda $\mathcal{F}$ všetkých čiastočných funkcií/popisných metód nemá minimálny prvok

Sporom - nech f je univerzálna pre $\mathcal{F}$

- uvažujme postupnosť reťazcov $\mathcal{S} = x_1, x_2, \ldots$: $\mathbf{C_f(x_i) \geq i}$
- definujme funkciu $\mathbf{g}$: $\mathbf{g(i) = x_i}$
- Kolmogorovská zložitosť reťazcov z $\mathcal{S}$

$\mathbf{C_g(x_i)} \leq 2 \log i + O(1)$

$2 \log i + O(1) \lll \mathbf{i \leq C_f(x_i)}$

$\rightsquigarrow$ f nie je univerzálna $\diamond$

V triede čiastočne rekurzívnych/computable funkcií minimálny prvok existuje $\hookrightarrow$

metódy — $\{\Phi \mid \Phi$ je čiastočne rekurzívna funkcia$\}$

Theorem (invariance)

Existuje univerzálna čiastočne rekurzívna funkcia.

- usporiadanie TS $\rightsquigarrow$ usporiadanie PRF
- Φ_0, funkcia, kt. počíta UTS;
  vstup: $1^{l(n)}$ **0 n p** $1^{l(n)}$**0n** — kód TS T_n **p** — program
- $\Phi_0(\langle n, p\rangle) = T_n(p)$
  ak $UTS = T_0$, potom $U(0p) = U(p)$

$$C_{\Phi_0}(x) \leq C_{\Phi_n}(x) + c_{\Phi_n}, \quad c_{\Phi_n} = 2l(n) + 1$$

□

$x, y, p \in \mathbb{N}$: PRF Φ spolu s **p, y je popis x** $\Leftrightarrow$ **Φ<y,p>=x**

podmienená Kolmogorovská zložitosť

$$C_\Phi(x|y) = \begin{cases} \min\{l(p) : \ \Phi(\langle y, p\rangle) = x\} & \text{ak } p \text{ existuje} \\ \infty, & \text{ak } p \text{ neexistuje} \end{cases}$$

Theorem (conditional invariance)

*Existuje univerzálna PRF Φ_0 pre triedu PFR, ktoré počítajú **x** z **y**, pričom*

$$C_{\Phi_0}(x|y) \leq C_\Phi(x|y) + c_\Phi$$

- $\Phi_0(\langle y, \langle n, p\rangle\rangle) = \Phi_n(\langle y, p\rangle)$
- $C_{\Phi_0}(x|y) \leq C_{\Phi_n}(x|y) + c_{\Phi_n}, \quad c_{\Phi_n} = 2l(n) + 1$ □

na voľbe univerzálnej funkcie nezáleží

- $\forall$ aditívne optimálne funkcie Ψ, Ψ' $\quad \exists c_{\Psi,\Psi'}$

$$\forall x, y \quad |C_\Psi(x|y) - C_{\Psi'}(x|y)| \leq c_{\Psi,\Psi'}$$

$$\boxed{C(x|y) = C_{\Phi_0}(x|y)} \qquad \boxed{C(x) = C_{\Phi_0}(x|\epsilon)}$$

hľadanie správneho modelu

- $\mathbf{T}$ (samoodeľujúco) popisuje stroj; model, resp. regularita
  $\mathbf{p}$ popisuje program; neregularita

$$\boxed{\mathbf{C(x) = \min\{l(T) + l(p) : T(p) = x\} + O(1)}}$$

- $\min_{\mathbf{T}}\{\mathbf{l(T) + C(x|T)} \mid \mathbf{T} \in \{\mathbf{T_0}, \mathbf{T_1}, \ldots\}\}$

Theorem

$\exists c, \forall x, y \quad C(x) \leq l(x) + c, \qquad C(x|y) \leq C(x) + c$

- $C_T(x|y) = C(x)$ keď T so vstupom $\langle z, y \rangle$ vypočíta x práve vtedy, keď UTS vypočíta x pri vstupe $\langle z, \epsilon \rangle$
- $C(x|y) \leq C_T(x|y) + c = C(x) + c$ □

Example

- $C(xx) = C(x) + O(1)$
- ak Φ je totálna injektívna rekurzívna, potom $\exists c: \quad |C(\Phi(x)) - C(x)| \leq c$
- $C(x + C(x)) \leq ?$

Example

- $C(x, y) \leq C(x) + C(y) + 2\log(\min\{C(x), C(y)\}) + 1$
- $C(x, y|C(x)) \leq C(x) + C(y) + O(1)$

x^* je "prvý najkratší" program, kt.generuje x. Vieme vypočítať x^* , keď poznáme $x, C(x)$?

- ÁNO — systematicky simulujeme programy dĺžky $C(x)$; x^* je "prvý", kt. zastane s x
- výpočet $C(x)$ pri znalosti x sme zredukovali na výpočet x^* pri znalosti $x, C(x)$

$$C(x^*|x) = C(C(x)|x) \leq C(C(x)) \leq \log l(x) \qquad \boxed{\mathbf{C(x) = l(x^*)}}$$

Theorem

Nech $A \subseteq \mathbb{N} \times \mathbb{N}$ *je RE,* $y \in \mathbb{N}$. *Nech* $Y = \{x : \ (x, y) \in A\}$ *konečná. Potom* $\exists c_A \quad \forall x \in Y \ C(x|y) \leq l(d(Y)) + c_A$

- postupne vymenovávame $(x_1, y_1), \ldots$ a počítame tie, keď $y_j = y \quad (x_{i_1}, y), (x_{i_2}, y), \ldots$
- $+$ poradové číslo i_j stačí na určenie $x = x_{i_j}$ □

$\hookrightarrow$ **Ak** $A \subseteq \mathbb{N}$, $\mathbf{d(A^{\leq n}) \leq p(n)}$, je RE, potom

- $\forall x, l(x) \leq n \quad C(x|n) \leq l(p(n)) + O(1)$
- $\mathbf{C(x)} \leq C(x|n) + 2l(n) + O(1) \leq l(p(n)) + 2l(n) + O(1)$
- $\mathbf{C(x) = O(\log n)}$

Definition

Nech $\Phi_1, \Phi_2, \ldots$ a $\Psi_1, \Psi_2, \ldots$ sú enumerácie PRF, pričom $\Psi_i = \Phi_{f(i)}, \Phi_i = \Psi_{g(i)}$. Ak f, g sú čiastočne rekurzívne, potom Φ, Ψ sú **rekurzívne izomorfné**.

$\forall$ rekurzívne izomorfné $\Psi, \Phi \; \exists \; c_{\Phi,\Psi} : \forall x \; |C_\Phi(x) - C_\Psi(x)| < c_{\Phi,\Psi}$

ALE

Existujú enumerácie Ψ, Φ také, že rozdiel $|C_\Phi(x) - C_\Psi(x)|$ je neohraničený.

$\forall$ rekurzívne izomorfné Ψ, Φ $\exists$ $c_{\Phi,\Psi} : \forall x \; |C_\Phi(x) - C_\Psi(x)| < c_{\Phi,\Psi}$

- f, g čiastočne rekurzívne $\Rightarrow \exists n(f), m(g)$ tak, že $\Phi_{n(f)}, \Psi_{m(g)}$ počíta f, resp. g.
- $C_{\Phi_0}(x) \leq C_{\Phi(i)}(x) + c_i \leq C_{\Psi_{g(i)}}(x) + c_i + c_{m(g)}$
  $C_{\Psi_0}(x) \leq C_{\Psi(i)}(x) + c_i \leq C_{\Phi_{f(i)}}(x) + c_i + c_{n(f)}$

$$C_\Phi(x) = l(p), \Phi_0(p) = x \qquad c_f$$

$$C_\Psi(x) = l(q), \Psi_0(q) = x \qquad c_g$$

$$C_\Psi(x) \leq l(q) + 2l(c_g) + 1 \qquad C_\Phi(x) \leq l(p) + 2l(c_f) + 1$$

Pre $c = \max\{2l(c_g), 2l(c_f)\} + 1$ $\qquad |C_\Phi(x) - C_\Psi(x)| \leq c$

$\diamond$

Existujú enumerácie Ψ, Φ také, že rozdiel $|C_\Phi(x) - C_\Psi(x)|$ je neohraničený.

Definujeme Ψ k Φ:

$$\begin{array}{rcll} \Psi_{2i}(1) & = & y_i & C_\Phi(y_i) \geq i^2 \\ \Psi_{2i}(x) & = & \Phi_i(x) & \text{pre } x > 1 \\ \Psi_{2i+1}(x) & = & \Phi_i(x) & \end{array}$$

$$C_\Psi(y_i) \leq C_{\psi_{2i}}(y_i) + c_{\Psi_{2i}} = 1 + c_{\Psi_{2i}},\ c_{\Psi_{2i}} \leq 2\log 2i + O(1)$$

$$|C_\Phi(y_i) - C_\Psi(y_i)| \geq i^2 - c_{\Psi_{2i}} - O(1) \rightsquigarrow \infty$$

$\diamond$

Definition

Reťazec x je **c-nestlačiteľný** ak $\mathbf{C(x) \geq l(x) - c}$

- počet reťazcov dĺžky n je 2^n
- počet programov dĺžky menšej ako n je $\sum_{i=0}^{n-1} 2^i = 2^n - 1$

  $\hookrightarrow$ $\forall n \; \exists$ aspoň jeden nestlačiteľný reťazec
- $\#$ programov dĺžky menšej ako $n - c$ je $\sum_{i=0}^{n-c-1} 2^i = 2^{n-c} - 1$

  $\frac{2^n - (2^{n-c} - 1)}{2^n}$ aspoň (1-2^{-c})-tina reťazcov dĺžky n je c-nestlačiteľná

Theorem (incompressibility)

Nech $c \in \mathbb{N}$, y je fixované a A je konečná množina kardinality aspoň m. Potom aspoň $m(1 - 2^{-c}) + 1$ prvkov $x \in A$ má $C(x|y) \geq \log m - c$.

- programov dĺžky menšej ako $\log m - c$ je $\sum_{i=0}^{\log m - c - 1} 2^i = 2^{\log m - c} - 1 = \frac{m}{2^c} - 1$
- v A existuje aspoň $m - (\frac{m}{2^c} - 1) > m(1 - \frac{1}{2^c})$ prvkov s $C(x) \geq \log m - c$

□

Pritom $x \in A$ stačí identifikovať indexom a popisom A

$$C(x) \leq c_A + l(d(A)) \geq c_A + l(m)$$

Sú podreťazce nestlačiteľných reťazcov nestlačiteľné?

c-nestlačiteľné $x = uvw$, $|x| = n$, môžme popísať programom q

program p_v pre generovanie v
reťazec uw
separácia uw na u, w - nech $|u| \geq |w|$

$$\Big\} \; \mathbf{q = \overline{l(p_v)} p_v \overline{l(u)} uw}$$

$$l(q) = C(v) + 2l(C(v)) + 2l(u) + n - l(v) + 2$$

$$\mathbf{n - c \leq C(x) \leq l(q) + O(1)}$$

$$n - c \leq C(x) \leq C(v) + n - l(v) + 4 \log n + O(1)$$

$$\boxed{\mathbf{C(v) \geq l(v) - O(\log n)}}$$

Môžme predpokladať, že $C(v) \geq l(v) - O(1)$? **NIE**

Example

Nech $C(x) = l(p)$. Potom p je c-nestlačiteľné.

Sporom

- nech $\exists$ program $q : U(q) = p$ & $\mathbf{l(q) < l(p) - c}$.
- vezmime $V = T_i : V(q) = U(U(q))$
- potom $\mathbf{U(1^i 0 q) = x}$

$$\mathbf{C(x) \leq l(q) + i + 1 < l(p) - c + i + 1}$$

pre $c > i + 1$ SPOR ◇

Example

O vzťahu $C(x, y)$ k $C(x), C(y)$ — C nie je subaditívna.

horný odhad
$C(x, y) \leq C(x) + C(y) + 2\min\{\log(C(x)), \log(C(y))\} + 1$

dolný odhad: Uvažujme $A = \{(x, y) \mid |xy| = n\}$.

- $d(A) = (n + 1)2^n$
- existuje (x, y) : $C(x, y) > \log d(A) - c = n + \log n - c$
- pritom $C(x) + C(y) \leq l(x) + l(y) + d = n + d$

$C(x, y) > n + \log n - c \geq C(x) + C(y) - d + \log n - c = C(x) + C(y) + \log n - e$

Example

C nie je monotónna na prefixoch.

$\exists\ n > m\ x = 1^n,\ y = 1^m$, y vlastný prefix x ale $C(y) > C(x)$
$n = 2^k :\rightsquigarrow C(1^n) \leq \log\log n + O(1)$
m: $A = \{1, 11, \ldots, 1^n\}$

Incompressibility Thm: aspoň $n/2 + 1$ prvkov má $C(z) \geq \log n - 1$

Zvolíme m: $n/2 \leq m < n$

$C(1^m) \geq \log n - 1$ ◇

KZ pri znalosti dĺžky – C(x|l(x)) $\rightsquigarrow C(x|l(x)) \leq C(x) + c$

n-string je reťazec $n0^{n-l(n)}$, pričom $|n0^{n-l(n)}| = n$

- Ak x je n-string, tak $C(x|n) \leq c$
- **ani $C(x|l(x))$ nie je na prefixoch monotónna**
  n nestlačiteľné: $C(n) \geq l(n)$; $x = n0^{n-l(n)}$ $C(x|l(x)) \leq c$
  $C(n|l(n)) \geq C(n) - C(l(n)) \geq \log n - 2 \log \log n + O(1)$

Čo ak vieme, že $x \in A$? // $C(x|A) \leq l(d(A)) + c_A$

Definition

Randomness deficiency x vzhľadom k množine A je

$$\delta(x|A) = l(d(A)) - C(x|A)$$

- $\delta(x|A) \geq -c_A$
- Čo hovorí veľké $\delta(x|A)$?
- Koľko je prvkov s $\delta(x|A) \geq k$?

  $\delta(x|A) = l(d(A)) - C(x|A) \geq k \quad \rightsquigarrow \quad C(x|A) \leq l(d(A)) - k$

  $|x : C(x|A) \leq l(d(A)) - k| < 2^{l(d(A))-k+1} = \frac{2^{l(d(A))}}{2^{k-1}}$

  $d(\{x : \delta(x|A) \geq k\}) \leq d(A)/2^{k-1}$

randomness deficiency a náhodnosť

Nech: $\Phi = \Phi_r$, //Φ partial recursive, computable

$R = \{(x, y) : \Phi(i) = \langle x, y \rangle, i \geq 1\}$ //R recursively enumerable

Nech $\mathbf{A} = \{\mathbf{x} : (\mathbf{x}, \mathbf{y}) \in \mathbf{R}\}$ je konečná

$\hookrightarrow C(x|y) \leq \log d(A) + \log r + 2 \log \log r + O(1)$

$\hookrightarrow \delta(x|y) = \log d(A) - C(x|y)$

x je náhodný v A ak $\delta(x|y) = O(1)$

Nech $A = \{x : \ l(x) = n\}$, $R = \{(x, n) : \ l(x) = n\}$

$\hookrightarrow \delta(x|n) = \underbrace{n}_{\log d(A)} - C(x|n) + O(1)$

x je náhodný **iff** $\delta(x|n) = O(1)$

Theorem

Pre Kolmogorovskú zložitosť platí

1. C *je neohraničená*
2. $m(x) = \min\{C(y) : y \geq x\}$ *je neohraničená*
3. $\forall$ *PRF* Φ *monotónne rastúcu donekonečna (od nejakého* x_0*) platí* $m(x) < \Phi(x)$ *až na konečne veľa* x

!!! 2,3 pre $C(x|l(x))$ neplatia

// nekonečne veľa razy skočí na konštantu (n-stringy)

1. vyplýva z 2.
2. $\forall i \; \exists$ min. x_i také, že $\forall x > x_i : \;\; C(x) \geq i$
   $\rightsquigarrow \quad m(x) = i,\ x_i < x \leq x_{i+1}$

3. $\forall$ PRF Φ monotónne rastúcu donekonečna (od nejakého x_0) platí $m(x) < \Phi(x)$ až na konečne veľa x

Nech $\Phi{=}\Phi_r$ monotónne neklesajúca, $\Phi(x) \leq m(x)$ pre nekonečne veľa x;

- $D(\Phi) = A = \{x : \ \Phi(x) < \infty\}$ nekonečná, RE, **preto** $\exists$ nekonečná $B, B \subseteq A$, s rekurzívnou χ_B
- $$\Psi(x) = \begin{cases} \Phi(x) & x \in B \\ \Phi(y) & y = \max\{z : z \in B, z < x\} \text{ inak} \end{cases}$$
  $$\Psi(x) \leq \Phi(x) \leq m(x)$$
- $M(a) = \max\{x : C(x) \leq a\}$ $M(a)+1 = \min\{x : m(x) > a\}$
  $\max\{x : \Psi(x) \leq a+1\} \geq \min\{x : m(x) > a\} = M(a)+1 > M(a)$
- $F(a) = \max\{x : \Psi(x) \leq a+1\}$ je totálna rekurzívna
  $\hookrightarrow$ $F(a) > M(a)$ pre nekonečne veľa a $\hookrightarrow$ $C(F(a)) > a$

ALE $\mathbf{C(F(a))} \leq C_F(F(a)) + O(1) \leq l(a) + c_F + O(1) = \mathbf{l(a) + O(1)}$
$\mathbf{a < C(F(a)) < l(a) + c}$ □

intermezzo o rekurzívnych fciách $D(\Phi) = A = \{x : \Phi(x) < \infty\}$

nekonečná, RE, **preto** má nekonečnú podmnožiny $B, B \subseteq A$, s rekurzívnou χ_B:

- f RE s oborom hodnôt A, g rekurzívna:

  $g(0) = f(0)$

  $g(x+1) = f(y)$, y najmenšie také, že $f(y) > g(x)$

B je obor hodnôt g

A nekonečná, preto B nekonečná
g vymenováva B v rastúcom poradí
} g je rekurzívna

- nekonečná A je rekurzívna $\iff$ ak je enumerovateľná v rastúcom poradí

$\Rightarrow$ postupne počítame charakteristickú fciu

$\Leftarrow$ $y \overset{?}{\in} A$ postupne počítame, až kým $\chi(x) \in B$ pre $x \geq y$

A je RE A je obor hodnôt totálnej rekurzívnej fcie
A je R χ_A je rekurzívna ◇