

Kolmogorovská zložitost

7.11.2013

- $g(\langle\langle y, z \rangle, k \rangle) = \langle p/q \rangle$ píšeme ako $g(y/z, k) = p/q$
- (čiastočná) funkcia $f : \mathbb{Q} \rightarrow \mathbb{R}$ je **upper semicomputable/co-enumerable** ak existuje rekurzívna $g(x, k)$ nerastúca v k tak, že $f(x) = \lim_{k \rightarrow \infty} g(x, k)$

f je **enumerovateľná/lower semicomputable** ak $-f$ je upper semicomputable; semicomputable ak je lower alebo upper semicomputable

- funkcia $f : \mathbb{Q} \rightarrow \mathbb{R}$ je **rekurzívna** ak existuje rekurzívna $g(x, k)$ tak, že $|f(x) - g(x, k)| < 1/k$
- enumerovateľná f je **univerzálna**, ak existuje enumerácia f_1, f_2, \dots enumerovateľných tak, že

$$f(i, x) = f_i(x) \qquad // f(i, x) = f(\langle i, x \rangle)$$

Lemma

Existuje univerzálna enumerovateľná funkcia.

$$f(i, x) = f_i(x)$$

enumerácia Φ_1, Φ_2, \dots PRF $\Phi(\langle x, k \rangle) = \langle p, q \rangle \iff \Phi(x, k) = p/q$

transformácia enumerácie Φ na enumeráciu f :

$$f_i(x) = \begin{cases} \sup_{k \in \mathbb{N}} \{\Phi_i(x, k)\}, & \text{ak existuje} \\ 0, & \text{ak neexistuje} \end{cases}$$

univerzálna $\Phi_0(i, \langle x, k \rangle) = \Phi_i(x, k)$ pritom $\exists j \Phi_0(i, \langle x, k \rangle) = \Phi_j(\langle i, x \rangle, k)$

$$f_j(\langle i, x \rangle) = f_i(x)$$

Lemma

Nech $f(x, y)$ je co-enumerovateľná. Potom

$$\forall x, y \ C(x|y) \leq f(x, y) + O(1) \Leftrightarrow d(\{x : f(x, y) \leq m\}) = O(2^m) \forall y, m$$

$$\Rightarrow \text{AK } \forall c \exists y, m : d(\{x : f(x, y) \leq m\}) > c2^m$$

$$\text{TAK } \exists x : m \geq C(x, y) \geq m + \log c$$

$$\Leftarrow g \text{ aproximuje } f \text{ zhora; } m := f(x, y) \text{ a } A := \{x : f(x, y) \leq m\}$$

$$\text{popis } x \text{ - znalosť } g, y, m \text{ a indexu do } A \rightsquigarrow C(x|y, m) \leq m + O(1)$$

$$\text{pre } h : C(x|y, m) = m - h$$

$$C(x|y) \leq C(x|y, m) + 2 \log h + O(1) \leq f(x, y) - h + 2 \log h + O(1)$$

- **diskrétna semimiera** $P : N \rightarrow \mathbb{R}$ $\sum_{x \in N} P(x) \leq 1$

//Ak rovnosť, pravdepodobnosť

- \mathcal{M} -trieda diskretných semimier. P_0 je **univerzálna pre \mathcal{M}** ak

- $P_0 \in \mathcal{M}$

- $\forall P \in \mathcal{M} \exists c_P \forall x \in N \ c_P P_0(x) \geq P(x)$

P_0 multiplikatívne dominuje

Theorem

*Existuje univerzálna enumerovateľná diskretná semimiera **m***

- 1 enumerovateľné diskretné semimiery sa dajú vymenovávať P_1, P_2, \dots
- 2 $P_0(x) = \sum_{n \geq 1} \alpha(n) P_n(x)$, $\sum \alpha(n) \leq 1$, $\alpha(n) > 0$ je enumerovateľná

!!

Lemma

Trieda rekurzívnych semimier nemá univerzálny prvok.

■ x_i najmenšie také, že $P_0(x_i) < 2^{-i}/i$

■ $Q(x) = \begin{cases} 2^{-i}, & x = x_i \\ 0, & \text{inak} \end{cases}$

$$\sum_x Q(x) = \sum_i 2^{-i} = 1$$

Q je semimiera

$$\forall i \quad Q(x_i) = Q(x) = 2^{-i} > iP_0(x)$$

Theorem

m nie je rekurzívna a $\sum_x m(x) < 1$

Univerzálne pravdepodobnostné rozdelenie

- spočítateľný priestor S ; $S = N \cup \{u\}$
- $P : S \rightarrow \mathbb{R}$ také, že $\sum_{x \in S} P(x) \leq 1$ // zvyšok priradíme $u \notin N$
- P je **enumerovateľná**, ak $\{(x, y) \mid x \in \mathbb{N}, y \in \mathbb{Q}, y \leq P(x)\}$ je RE
 P je zdola aproximovateľná
- **Levin:**
 - enumerovateľné rozdelenia vieme vymenúvať: P_1, P_2, \dots
 - **existuje univerzálne rozdelenie m**

$$\forall k \in \mathbb{N} \exists c \in \mathbb{Q} \forall x \in \mathbb{N} [c \cdot m(x) \geq (P_k(x))]$$

$$m(x) = 2^{-K(x)}$$

$$c = 2^{K(P_k)+O(1)} = 2^{K(k)+O(1)} = O(k \log^2 k)$$

Priemerná zložitosť

$x \in N$, $l(x)$ – dĺžka binárneho zápisu x , $t(x)$ – čas algoritmu na vstupe x

$$T(n) = \max\{t(x) \mid l(x) = n\} \quad \text{najhorší prípad}$$

$$T_{av}^P(n|P) = \frac{\sum_{l(x)=n} P(x)t(x)}{\sum_{l(x)=n} P(x)} \quad \text{priemerný prípad}$$

- rovnomerné rozdelenie $L(x) = 2^{-2(l(x)) - 1}$, $L(x|l(x) = n) = 2^{-n}$
- univerzálne rozdelenie $m(x)$

Quicksort

- $T_{av}^L(n) = \Theta(n \log n)$
- $T_{av}^m(n) = \Omega(n^2)$

Theorem

Nech A je algoritmus, ktorý na každom vstupe zastane, vstupy podľa m . Potom priemerná zložitosť je rádovo rovná najhoršej.

Nech A je algoritmus, ktorý na každom vstupe zastane, vstupy podľa m . Potom priemerná zložitosť je rádovo rovná najhoršej.

- Definujme $P(x)$

$$\forall n = 1, 2, \dots \quad a_n = \sum_{l(x)=n} m(x)$$

ak $l(x) = n$ a x je lexicograficky najmenšie také, že $t(x) = T(n)$,
tak $P(x) = a_n$, inak $P(x) = 0$

- P je enumerovateľné
- dodefinujeme $P(u) = m(u)$
- keďže $\sum_{x \in S} P(x) = \sum_{x \in S} m(x)$, tak $c_P \cdot \mathbf{m}(x) \geq \mathbf{P}(x)$

$$\begin{aligned} T_{av}^m(n) &= \frac{\sum_{l(x)=n} m(x)t(x)}{\sum_{l(x)=n} m(x)} \geq \frac{1}{c_P} \frac{\sum_{l(x)=n} P(x)t(x)}{\sum_{l(x)=n} m(x)} \\ &= \frac{1}{c_P} \frac{a_n T(n)}{\sum_{l(x)=n} m(x)} = \frac{1}{c_P} T(n) \geq \frac{1}{c_P} T_{av}^m(n) \quad \diamond \end{aligned}$$

$$P \rightsquigarrow P_k, \text{ potom } c_P \leq k \log^2 k \quad T_{av}^m(n) \geq \frac{T(n)}{k \log^2 k}$$

Náhodný graf s vysokou pravdepodobnosťou nemá izolované vrcholy

štandardne

konkrétny vrchol je izolovaný s pravdepodobnosťou $\frac{1}{2^{n-1}}$
 nejaký vrchol je izolovaný s pravdep. nanajvýš $\frac{n}{2^{n-1}}$
 graf nemá izolovaný vrchol s pravdep. $\geq 1 - \frac{n}{2^{n-1}}$

nestlačiteľnosťou

$\chi_1 \dots \chi_e$, $e = \binom{n}{2}$ -charakteristická postupnosť množiny hrán \sim grafu
 graf G s $C(G|n) \geq \binom{n}{2} - \delta(n)$

graf s izolovaným vrcholom i - nie susedia i , áno identifikácia i

$$\binom{n}{2} - \delta(n) \leq C(G|n) \leq \binom{n}{2} - (n-1) + \log n \rightsquigarrow \boxed{\delta(n) \geq n - \log n - 1}$$

nanajvýš $2^{-(n-\log n-1)}$ -tina má izolovaný vrchol, preto s pravdep. aspoň $1 - \frac{n}{2^{n-1}}$
 náhodný graf nemá izolovaný vrchol

Example

$n, r, s \in N$, $2 \log n \leq r, s \leq n/4$, s párne; potom $\forall n$ existuje $n \times n$ matica nad $GF(2)$ taká, že \forall podmatica rozmeru s riadkov a $n - r$ stĺpcov má hodnotu $\geq s/2$.

náhodné x dĺžky n^2 tvorí štvorcovú maticu R ; $C(x) \geq n^2$
 ak existuje podmatica s iba $s/2 - 1$ lineárne nezávislými riadkami, tak **skrátene popisu** x

- prvky odpovedajúce podmatici
- + charakteristická postupnosť s riadkov vzhľadom k nezávislým
- + zoznam nezávislých riadkov
- + závislé určené koeficientami nezávislých
- + samoodel'ujúco n, r, s
- + identifikácia podmatice

ak existuje podmatica s iba $s/2 - 1$ lineárne nezávislými riadkami, tak **skrátene popis** x

- prvky odpovedajúce podmatici $(n - r)s$
- + charakteristická postupnosť s riadkov vzhľadom k nezávislým s
- + zoznam nezávislých riadkov $(s/2 - 1)(n - r)$
- + závislé určené koeficientami nezávislých $(s/2 - 1) \times (s/2 + 1)$
- + samoodelujúco n, r, s $3 \log n + 6 \log \log n + O(1) \leq 4 \log n$
- + identifikácia podmatice $2n$

$$n^2 - (n - r)s + s + (s/2 - 1)(n - r) + (s/2 - 1) \cdot (s/2 + 1) + 4 \log n + 2n$$

$2 \log n \leq s, r \leq n/4$ -pre veľké n to klesne pod n^2

vlastnosť s veľkou pravdepodobnosťou \Leftrightarrow vlastnosť objektov s malou deficienciou

- 1 všetky objekty zahŕňajú aj objekty s malou (Kolmogorovskou) zložitnosťou
- 2 objekt s malou deficienciou je prvkom množiny, na ktorej je založená platnosť vlastnosti s veľkou pravdepodobnosťou
- 3 Ak P, Q platia s pravdepodobnosťou aspoň $1 - \epsilon$, tak $P \wedge Q$ s pravdepodobnosťou aspoň $1 - 2\epsilon$



Ak P, Q platia na objektoch s $\delta(n)$, tak $P \wedge Q$ platia na objektoch s $\delta(n)$

\rightsquigarrow situácia s n rôznymi vlastnosťami

veľká pravdepodobnosť a nestlačiteľnosť nie sú vo všeobecnosti totožné, ale za nejakých predpokladov skoro áno

Lemma

Nech $S \subseteq \{0, 1\}^n$

- 1 ak P spĺňajú $\forall x \in S$ s $\delta(x|S) \leq \delta(n)$, potom P platí aspoň na $(1 - 1/2^{\delta(n)})$ -tine prvkov S
- 2 n, S fixované, P vlastnosť, kt. platí aspoň na $(1 - 1/2^{\delta(n)})$ -tine prvkov S . Potom existuje c , že každá taká vlastnosť P platí **súčasne** $\forall x \in S$ s $\delta(x|S) \leq \delta(n) - K(P|S) - c$

Dôsledok

Objekty s random deficiency $\leq \delta(n)$ majú všetky vlastnosti P , ktoré platia aspoň s pravdepodobnosťou $1 - 2^{-\delta(n) - K(P|n) - O(1)}$

Dôsledok

Všetky rekurzívne vlastnosti P s $K(P|n) = O(1)$, ktoré samostatne platia s pravdepodobnosťou $\rightarrow 1$ pre $n \rightarrow \infty$, platia súčasne s pravdepodobnosťou $\rightarrow 1$ pre $n \rightarrow \infty$.

Nech $S \subseteq \{0, 1\}^n$

1 ak P spĺňajú $\forall x \in S$ s $\delta(x|S) \leq \delta(n)$, potom P platí aspoň na $(1 - 1/2^{\delta(n)})$ -tine prvkov S
 # programov dĺžky $\leq \log d(S) - \delta(n) \leq \sum_{i=0}^{\log d(S) - \delta(n)} 2^i$

2 n, S fixované, P vlastnosť, kt. platí aspoň na $(1 - 1/2^{\delta(n)})$ -tine prvkov S . Potom existuje c , že každá taká vlastnosť P platí **súčasne** $\forall x \in S$ s $\delta(x|S) \leq \delta(n) - K(P|S) - c$
 Sporom: . . . nech existuje $x \in S$: $\delta(x|S) \leq \delta(n) - K(P|S) - c$ a P **neplatí**. Potom rekonštrukcia x :

– popis P

– index j v množine M objektov, pre kt. P neplatí $|M| \leq \frac{d(S)}{2^{\delta(n)}}$

$$\exists c_1, c_2 \quad K(x|S) \leq \log j + c_1 \leq \log d(S) - \delta(n) + c_2$$

$$\delta(n) - K(P|S) - c \geq \delta(x|S) = I(d(S)) - K(x|S)$$

$$c_2 - c \geq K(P|S)$$

Γ - množina turnajov na $\{1, 2, \dots, n\}$

zorientovaný úplný graf

- $E : T \rightarrow \{0, 1\}^{n(n-1)/2}$

- tranzitívny turnaj $T(i, j), T(j, k) \Rightarrow T(i, k)$

$T(i, j) \iff i < j$

$v(n)$ - max. číslo také, že $\forall T \in \Gamma : T$ obsahuje tranzitívny podturnaj o $v(n)$ vrcholoch

Theorem

$$v(n) \leq 1 + \lfloor 2 \log n \rfloor$$

Fixnime

- **T** $\in \Gamma : C(E(T)|n, p) \geq n(n-1)/2$

- **S** tranzitívny podturnaj na $v(n)$ vrcholoch

popis T

+ zoznam vrcholov S podľa dominance

- odstránime hrany patriace S z $E(T)$

$$n(n-1)/2 \leq C(E(T)|n, p) \leq n(n-1)/2 + v(n)\lfloor \log n \rfloor - v(n)(v(n)-1)/2$$

□

counting

tranzitívny podturnaj-pôvodne

- Γ' - turnaje, ktoré obsahujú tranzitívny podturnaj na $v = 2 + 2\lfloor \log n \rfloor$ vrcholoch
- $A \subseteq \{1, 2, \dots, n\}$ $d(A) = v$ σ je permutácia na A
- $d(\Gamma_{A,\sigma}) = 2^{\binom{n}{2} - \binom{v}{2}}$

$$d(\Gamma') \leq \sum_{A,\sigma} d(\Gamma_{A,\sigma}) = \binom{n}{v} v! 2^{\binom{n}{2} - \binom{v}{2}} < 2^{\binom{n}{2}} = d(\Gamma)$$

existuje $T \in \Gamma - \Gamma'$, ktorý nemá tranzitívny podturnaj na v vrcholoch

pravdepodobnostné metódy

\mathbb{T} náhodná premenná s hodnotami z Γ , $\Pr(\mathbb{T} = \mathbf{T}) = 2^{-\binom{v}{2}}$, $v = 2 + 2\lfloor \log n \rfloor$

$$\sum_A \sum_{\sigma} \Pr(\mathbf{T}|A \text{ generované podľa } \sigma) = \binom{n}{v} v! 2^{-\binom{v}{2}} < 1$$

existuje taká hodnota T , že T nemá tranzitívny podturnaj na v vrcholoch

turnaj s k -dominátormi

vlastnosť $S(k)$ - $\forall A \subseteq N, d(A) = k$, existuje v $N - A$ hráč x , ktorý dominuje všetkým v A

$s(k)$ - minimálny počet hráčov turnaja T s vlastnosťou $S(k)$

Theorem

$$s(k) \leq 2^k k^2 (\ln 2 + o(1))$$

Zoberme

$$n = 2^k k^2 (\ln 2 + o(1)), \text{ pričom } C(E(T)|n, k) \geq n(n-1)/2$$

turnaj T , v ktorom existuje $A, d(A) = k$, bez dominujúceho v $N - A$

kratší popis T

+ vrcholy A

$$d(A) \log n$$

- odstránime popis hrán $A \times (N - A)$

$$k(n - k)$$

+ popíšeme ich indexom j v množine možností

$$\log(2^k - 1)^{(n-k)}$$

$$n(n-1)/2 \leq C(E(T)|n, k, p) \leq n(n-1)/2 + \underbrace{k \log n - (n-k)k + \log(2^k - 1)^{(n-k)}}_{<0}$$

veľmi pravdepodobné vlastnosti

c konštanta, y fixné. Potom každá konečná množina A veľkosti m má aspoň $m(1 - 2^{-c}) + 1$ prvkov s $C(x|y) \geq \log m - c$

\Leftrightarrow aspoň $2^{n(n-1)/2}(1 - 1/n)$ turnajov s $C(E(T)|n, p) \geq n(n-1)/2 - \log n$

Dôsledok

Takmer všetky n -vrcholové turnaje (aspoň $(1 - 1/n)$ -tina) majú najväčší tranzitívny podturnaj veľkosti nanajvýš $1 + 2\lfloor 2 \log n \rfloor$

$n \rightarrow \infty$

$$C(E(T)|n, k, p) \geq n(n-1)/2 - \log n$$

Dôsledok

Pre dostatočne veľké k existuje $n, n \leq 2^k k^2 (\ln 2 + o(1))$ tak, že aspoň $(1 - 1/n)$ -tina turnajov má vlastnosť $S(k)$

$\mathcal{D} = \{D_1, \dots, D_j\}$, $D_i \subseteq N = \{1, \dots, n\}$ rozlišuje N , ak $\forall M, M' \subseteq N, M \neq M'$, existuje $i, 1 \leq i \leq j$ tak, že $d(D_i \cap M) \neq d(D_i \cap M')$

$f(n)$ - minimálne $d(\mathcal{D})$

(Vie sa) $f(n) = \frac{2n}{\log n} + O\left(\frac{n \log \log n}{\log^2 n}\right)$

Theorem

$$f(n) \geq (2n / \log n)[1 + O(\log \log n / \log n)]$$

$M, M' \subseteq N$ tak, že $C(E(M)|\mathcal{D}) \geq n$;

$$d_i = d(D_i),$$

$$m_i = d(D_i \cap M)$$

- s_i je podpostupnosť $E(M)$ odpovedajúca jednotkám v $E(D_i)$;
 $l(s_i) = d_i, \#_1(s_i) = m_i, C(s_i|\mathcal{D}) \geq d_i - O(\log i)$

- $d_i/2 - O(\sqrt{d_i \log i}) \leq m_i \leq d_i/2 + O(\sqrt{d_i \log i})$

$$C(x) \geq n - \delta(n), \text{ tak } \underbrace{|\#ones(x) - n/2|}_{A} \leq \sqrt{\frac{3}{2}(\delta(n) + c)n / \log e}$$

ak vieme d_i, m_i určíme poradím v $A \rightsquigarrow C(m_i|D_i) \leq \frac{1}{2} \log d_i + O(\log \log i)$

$$n \leq C(E(M)|\mathcal{D}) \leq C(m_1, \dots, m_j|\mathcal{D}) \leq \sum_{i=1}^j \left(\frac{1}{2} \log n + O(\log \log n) \right)$$

Kolmogorovsky náhodný graf - graf je $\delta(n)$ -náhodný ak má randomness deficiency $\leq \delta(n)$

$$C(E(G)|n, \delta) \geq n(n-1)/2 - \delta(n)$$

Lemma

1. Aspoň $(1 - 1/2^{\delta(n)})$ -tina grafov je $\delta(n)$ -náhodná.
2. Ak G je $\delta(n)$ -náhodný, d stupeň ľubovoľného vrchola, potom

$$|d - (n-1)/2| = O\left(\sqrt{(\delta(n) + \log n)n}\right)$$

1 počet programov

- 2
 - $C(E(G)|n, \delta) \geq n(n-1)/2 - \delta(n)$
 - i je vrchol s $|d - (n-1)/2| > k$
 - $V_i = \{j \in V - \{i\} \mid (i, j) \in E\}$

- + pôvodné kódovanie $E(G)$ $C(E(G)|n, \delta) \geq n(n-1)/2 - \delta(n)$
- + identifikácia i, k $2 \log n$
- odstránime pôvodný bitový vektor pre V_i $-n$
- + namiesto bitového vektora pre V_i index do množiny množín susedov mohutnosti m

$$\log \underbrace{\left[\sum_{|d-(n-1)/2| > k} \binom{n-1}{d} \right]}_m \leq \log \left[2 \cdot 2^{n-1} e^{-2k^2/3(n-1)} \right] = n - (2k^2/3(n-1)) \log e$$

$$\log m + 2 \log n + n(n-1)/2 - n + O(1) \geq n(n-1)/2 - \delta(n)$$

$$n - (2k^2/3(n-1)) \log e \geq \log m \geq n - 2 \log n - \delta(n) - O(1)$$

$$k \leq \sqrt{\frac{3}{2} (\delta(n) + 2 \log n + O(1)) (n-1) / \log e} = O\left(\sqrt{(\delta(n) + \log n)n}\right)$$

□

Lemma

Všetky $o(n)$ -náhodné očíslované grafy majú medzi ľubovoľnou dvojicou vrcholov $n/4 + o(n)$ rôznych ciest dĺžky 2.

1 priemer 1 majú len K_n

2 $i, j \in V$, spojené r cestami dĺžky 2; potom popis G

popis

$O(1)$

$i, j, i < j$

$2 \log n$

$E(G)$ bez bitov popisujúcich hrany $(i, k), (j, k)$

$\binom{n}{2} - 2(n-2)$

najkratší popis zoznamu $e_{i,j}$ odstránených hrán

$C(e_{i,j}|n)$

$$O(\log n) + n(n-1)/2 - 2(n-2) + C(e_{i,j}|n) \geq n(n-1)/2 - o(n)$$

$$C(e_{i,j}|n) \geq I(e_{i,j}) - o(n) = 2(n-2) - o(n)$$

3 frekvencia 11 v $e_{i,j}$ podľa $|\#y - \frac{n}{2^{l(y)}}| \leq \sqrt{\alpha pn}, \alpha = K(y|n) + \log I(y) + \delta(n)$

$$n/4 + o(n)$$

\rightsquigarrow $o(n)$ -náhodné grafy sú $n/4 + o(n)$ -súvislé.

Lemma

$G = (E, V)$ má random.def. $\delta(n) = O(\log n)$. Potom maximálna klika, ktorú G obsahuje, má veľkosť nanajvyš $\lfloor 2 \log n \rfloor + O(1)$.

$$-\frac{k(k-1)}{2} + k \log n + O(1)$$

Lemma

Nech G je aspoň $c \log n$ -náhodný očíslovaný, $i \in V$. $\forall i \neq j$ alebo $(i, j) \in E$ alebo existuje k spomedzi $(c + 3) \log n$ najmenších $(i, k) \in E$, pričom $(k, j) \in E$

A je množina $(c + 3) \log n$ najmenších spomedzi susedov i ;

SPOROM: Nech k nie je spojený so žiadnym z $A \cup \{i\}$; potom krátky popis G :

- popis hrán incidentných s i, k $-(2n - 2)$
 - + popis i $\log n$
 - bitový vektor o susedoch i $(n - 1)$
 - popis k + s ním incidentné hrany $\log n + (n - 2) - (c + 3) \log n$
- $$n(n - 1)/2 + 2 \log n + 2n - 3 - (c + 3) \log n - 2n + 2 \geq n(n - 1)/2 - c \log n$$
- $$c \log n \geq (c + 1) \log n + 1$$