Separating Words Problem

Csányi Dávid

Eötvös Loránd Tudományegyetem

Témavezető: Pálvölgyi Dömötör

2023. június

Bevezetés

Csányi Dávid

Egy egyszerű számítási modellel milyen nehéz megoldani a legkönnyebb feladatot, két szó megkülönböztetését?

- $\mathrm{sep}_l(a_1, \dots a_l)$ az a legkisebb k szám, amire létezik egy k állapotú véges automata, ami az $a_1, a_2, \dots a_l$ szavak esetén különböző végállapotokba kerül
- $S_l(n) = \max\{\mathrm{sep}_l(a_1, \dots a_l) : a_1, \dots a_l$ a Σ abc feletti maximum n hosszú páronként különböző szavak$\}$

Történet

Felső korlátok:

- 1986 Goralcik, Koubek: kérdés megfogalmazása, $S_2(n) = o(n)$
- 1989 Robson: $S_2(n) = \mathcal{O}(n^{\frac{2}{5}}(\log n)^{\frac{3}{5}})$
- 2020 Chase: $S_2(n) = \mathcal{O}(n^{\frac{1}{3}}(\log n)^7)$

A legjobb ismert alsó korlát $S_2(n) = \Omega(\log(n))$

Segédállítás

Az alábbi prímszámokkal kapcsolatos állítást a későbbiekben többször használni fogjuk:

Lemma ([2])

Minden $n \geqslant 2$ természetes számra létezik egy $q \leqslant 4,4 \log(n)$ prím, amelyre igaz, hogy $q \nmid n$.

Következmény ([2])

Ha $0 \leqslant i, j \leqslant n$, $n \geqslant 2$ és $i \neq j$, akkor létezik egy q prím, amelyre $q \leqslant 4,4 \cdot \log(n)$ és $i \not\equiv j \pmod q$.

Részszó kezdőpozícióinak halmaza

Definíció

Legyen $a = a_0 \dots a_{n-1} \in \{0,1\}^n$ és $w = w_0 w_1 \dots w_{l-1} \in \{0,1\}^l$. Ekkor $\mathbf{pos}_w(a)$-val jelöljük a w részszó kezdőpozícióinak halmazát az a szóban. Formálisan $\mathbf{pos}_w(a) = \{j : 0 \leqslant j \leqslant n-l, a_j = w_0, \dots a_{j+l-1} = w_{l-1}\}$.

Definíció

Legyen $m \in \mathbb{Z}^+$ és i egy maradékosztály modulo m. Ekkor $\mathbf{pos}_w(a)_{m,i}$-vel jelöljük a w részszó azon j kezdőpozícióinak halmazát, amelyekre $j \equiv i \mod m$.

Gyökös felső becslés

$a \neq b \Rightarrow \mathrm{pos}_1(a) \neq pos_1(b)$

Állítás (Chase [1])

Ha $A \neq B \subseteq \{0, 1 \dots n-1\}$, akkor létezik egy $p = \mathcal{O}(\sqrt{n \log(n)})$ prím és i maradékosztály, amelyekre $|A_{p,i}| \neq |B_{p,i}|$.

$$|\mathrm{pos}_1(a)_{p,i}| \neq |\mathrm{pos}_1(b)_{p,i}|$$
$$|\mathrm{pos}_1(a)_{p,i}| \not\equiv |\mathrm{pos}_1(b)_{p,i}| \mod q$$

Tétel (Chase [1])

$S_2(n) = \mathcal{O}(\sqrt{n} \cdot \log(n)^{1.5})$

Gyökös felső becslés

Lemma ([3])

Tekintsük az $A \neq B \subseteq \{0, \ldots n-1\}$ halmazokat és az $m \in \mathbb{Z}^+$ számot. Ha minden $i \in \{0, \ldots m-1\}$ maradékosztályra $|A_{m,i}| = |B_{m,i}|$, akkor $\Phi_m(x)$ osztja az $(A(x) - B(x))$ polinomot.

$$\Phi_m(x) | (x^m - 1) | (x^{i+k_1 m} - x^{i+k_2 m})$$

Gyökös felső becslés

Állítás (Chase [1])

Ha $A \neq B \subseteq \{0, 1 \ldots n-1\}$, akkor létezik egy $p = \mathcal{O}(\sqrt{n \log(n)})$ prím és i maradékosztály, amelyekre $|A_{p,i}| \neq |B_{p,i}|$.

$$n \geqslant \deg(A(x) - B(x))$$

$$\geqslant \sum_{p \leqslant k} \deg(\Phi_p(x)) = \sum_{p \leqslant k} (p-1) \sim \frac{k^2}{\log k}$$

A gyökös becslés általánosítás

Lemma

Tekintsük az $A_1, A_2, \ldots A_l \subseteq \{0, \ldots n-1\}$ páronként különböző halmazokat és az $m \in \mathbb{Z}^+$ számot. Ha minden $i_{1,2}, \ldots i_{l-1,l} \in \{0, \ldots m-1\}$ maradékosztály $\binom{l}{2}$-esre valamelyik h, j-re igaz, hogy $|(A_h)_{m,i_{h,j}}| = |(A_j)_{m,i_{h,j}}|$, akkor $\Phi_m(x)$ osztja az $(A_1)(x), \ldots (A_l)(x)$ polinomok közül valamelyik kettő különbségét.

A gyökös becslés általánosítás

Lemma

Ha $A_1, A_2, \ldots A_l \subseteq \{0, 1, \ldots n-1\}$ és páronként eltérőek, akkor létezik egy $p = \mathcal{O}(l\sqrt{n \log(n)})$ prím és egy $i_{1,2}, \ldots i_{l-1,l}$ maradékosztály $\binom{l}{2}$-es, hogy minden $0 \leqslant h \neq j \leqslant l-1$-re $|(A_h)_{p,i_{h,j}}| \neq |(A_j)_{p,i_{h,j}}|$.

$$\binom{l}{2} n \geqslant \deg(\prod_{h \neq j}((A_h)(x) - (A_j)(x)))$$

$$\geqslant \sum_{p \leqslant k} \deg(\Phi_p(x)) = \sum_{p \leqslant k} (p-1) \sim \frac{k^2}{\log k}$$

A gyökös becslés általánosítás

Tétel

$S_l(n) = \mathcal{O}(l^5\sqrt{n}\log(n)^{1.5})$

$$V(a) = |\mathrm{pos}_1(a)_{p,i_{1,2}}| + n \cdot |\mathrm{pos}_1(a)_{p,i_{1,3}}| + \cdots + n^{\binom{l}{2}-1}|\mathrm{pos}_1(a)_{p,i_{l-1,l}}|$$

$\exists q = \mathcal{O}(l^4\log(n))$, hogy
$V(a_j) \not\equiv V(a_k) \mod q$, ha $j \neq k$

Hivatkozások

[1] Zachary Chase. „A new upper bound for separating words”. (2020). URL: https://arxiv.org/abs/2007.12097.

[2] J. Shallit és Y. Breitbart. „Automaticity I: Properties of a measure of descriptional complexity.”. *J. Comput. System Sci.* 53 (1996), 10–25. old.

[3] M. N. Vyalyi és R. A. Gimadeev. „Separating Words by Occurrences of Subwords”. (2014). URL: https://link.springer.com/content/pdf/10.1134/S1990478914020161.pdf.