

Longest run & “Runs” conjecture

Riccardo Lo Iacono

25 settembre 2024

Considerata s una stringa qualsiasi di lunghezza n , qual è il numero massimo di ripetizioni $\rho(n)$ in essa?

Congettura: $\rho(n) < n$?

- Compressione di testi
- Indicizzazione di testi
- Ricerca di pattern genomici¹

¹Si è dimostrato che alcuni pattern genomici sono indicatori di alcune malattie.

Kolpakov e Kuchеров in [?], dimostrano come $\rho(n)$ sia limitato superiormente da una funzione $\mathcal{O}(n)$.

Segue la “Runs” conjecture.

- Dimostrazione della runs conjecture.
- Soluzione algoritmica per il calcolo delle ripetizioni massimali in $\mathcal{O}(n)$.

- Σ è un alfabeto² finito di simboli
- $s \in \Sigma^*$ è una stringa, la cui lunghezza è $|s|$
- $s[i]$ è l' i -esimo carattere di s , $s[i, j]$ è la sotto-stringa compresa tra gli indici i, j inclusivamente, $i, j \in (1, |s|)$
- $p \in \mathbb{N}$ periodo di $s \iff s[i] == s[i + p], 1 \leq i \leq |s| - p$
- \mathcal{I} insieme di intervalli, $Beg(\mathcal{I})$ posizioni iniziali degli intervalli in \mathcal{I}
- \prec ordine totale su Σ e ordine lessicografico indotto su esso

²Si assume Σ non unario.

Concetto di ripetizione ed esempio

Definizione: una terna $r = (i, j, p)$ è una ripetizione (o *run*) di una qualche stringa ω , se il più piccolo periodo p di $\omega[i, j]$ è tale che $|\omega[i, j]| \geq 2p$.

Sia $Runs(\omega)$ l'insieme delle runs in ω .

Esempio: sia $\omega = babbabbab$. Si osserva facilmente che le ripetizioni in essa sono quelle in *Figura 1*.

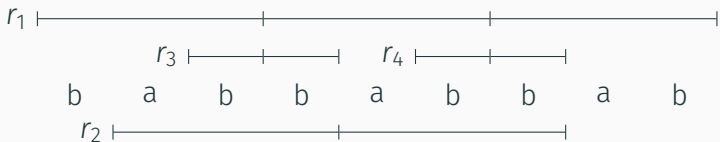


Figura 1: Esempio di ripetizioni.

Segue che

$$\text{Runs}(babbabbab) = \{(1, 9, 3), (2, 7, 3), (3, 4, 1), (6, 7, 1)\}$$

Definizione (Lyndon word): una stringa non vuota $\omega \in \Sigma$ è detta essere una *Lyndon word*, rispetto \prec , se $\omega \prec u$, per ogni u suffisso proprio di ω .

Definizione (L-root): data $r = (i, j, p)$ una run per una qualche stringa $\omega \in \Sigma^*$, un intervallo $\lambda = [i_\lambda, j_\lambda]$ è detto essere *L-root* di r rispetto \prec se $i \leq i_\lambda \leq j_\lambda \leq j$ e $\omega[i_\lambda, j_\lambda]$ è una Lyndon word.

“Runs” Theorem

Lemma: per ogni stringa ω e posizione i , sia $\ell \in \{0, 1\}$, tale che $\hat{\omega}[k] \prec_{\ell} \hat{\omega}[i]$, per $k = \min\{k' \mid \hat{\omega}[k'] \neq \hat{\omega}[i], k' > i\}$. Allora $l_{\ell}(i) = [i, i]$ e $l_{\bar{\ell}}(i) = [i, j]$, per qualche $j > i$.