Maximal repetition & “Runs” conjecture

Riccardo Lo Iacono

1 ottobre 2024

Definizioni preliminari

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

Esempio: sia considerata la stringa $s = aababaababb$. Si ha che gli intervalli $[1, 2] = a^2, [6, 7] = a^2, [10, 11] = b^2, [2, 6] = (ab)^{5/2}, [7, 10] = (ab)^2, [4, 9] = (aba)^2, [1, 10] = (aabab)^2$ hanno rispettivamente periodo, 1, 2, 3, 5.

Sia posto $\rho(n)$ il numero di ripetizioni massimali. Inoltre, sia $Runs(s)$ l'insieme delle run in s.

Background storico

Kolpakov e Kucherov[1] dimostrano $\rho(n) = \mathcal{O}(n)$.

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

[1] R. M. KUCHEROV AND G. KOLPAKOV, *Finding maximal repetition in a word in linear time.*

Punti chiave della discussione

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

Notazione

- Σ è un alfabeto finito di simboli
- $s \in \Sigma^*$ è una stringa, la cui lunghezza è $|s|$
- $s[i]$ è l'iesimo 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 intevalli, $Beg(\mathcal{I})$ posizioni iniziali degli intervalli in $\mathcal{I}$
- $\prec_0$ ordine totale su Σ e ordine lessicografico indotto su Σ^*, $\prec_1$ ordine rovesciato.

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

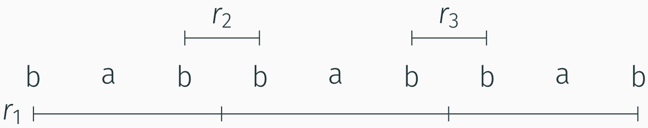

Figura 1: Esempio di ripetizioni massimali

Segue che

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

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

Definizione equivalente è la seguente.

Definizione: data una stringa s di lunghezza n e $i \in [1, n]$, la stringa $s_i s[1, i-1]$ è detta *shift ciclico* di s, non banale se $i > 1$. Una Lyndon word è una stringa non vuota che è lessicograficamente minore di ogni suo shift ciclico non banale.

Definizione (L-root): data $r = (i, j, p)$ una run per una qualche stringa $s \in \Sigma^*$, 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 $s[i_\lambda, j_\lambda]$ è una Lyndon word.

Sia $\hat{s} = s\$, \$ \notin \Sigma$.

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

Lemma 2: sia $r = (i, j, p)$ una ripetizione massimale in una stringa s, sia inoltre $\ell_r \in \{0, 1\}$ tale che $\hat{s}[j + 1] \prec_{\ell_r} \hat{s}[j + 1 - p]$. Allora, ogni L-root $\lambda = [i_\lambda, j_\lambda]$ di r rispetto $\prec_{\ell_r}$ è uguale a $l_{\ell_r}(i_\lambda)$.

Sia B_r l'insieme delle L-root tali da soddisfare *Lemma 2*, allora vale quanto segue.

Lemma 3: per ogni coppia di ripetizioni massimali r, r', con $r \neq r'$, $Beg(B_r) \cap Beg(B_{r'}) = \varnothing$.

Da ciò, poiché

$$|Beg(B_r)| = |B_r| \geq \lfloor e_e - 1 \rfloor \geq 1$$

e inoltre, dato che $1 \notin Beg(B_r)$ per ogni run r, regge $\sum_{r \in Runs(s)} |B_r| = \sum_{r \in Runs(s)} |Beg(B_r)| \leq |s| - 1$.

Teorema: $\rho(s) < n$.