

Alcuni risultati su parole finite

Gabriele Fici

Dipartimento di Informatica ed Applicazioni – Università di Salerno

Varese - 17 Luglio 2006

I - Parole incompletabili minimali

II - Automa dei suffissi, parole proibite minimali e wildcard

I - Parole incompletabili minimali

Sia X un insieme finito di parole finite sull'alfabeto A .

X si dice **completo** se $A^* \subseteq \text{Fact}(X^*)$.

Sia X un insieme finito di parole finite sull'alfabeto A .

X si dice **completo** se $A^* \subseteq \text{Fact}(X^*)$.

Viceversa X si dirà **non completo**.

Sia X un insieme finito di parole finite sull'alfabeto A .

X si dice **completo** se $A^* \subseteq \text{Fact}(X^*)$.

Viceversa X si dirà **non completo**.

Example

A^n , con n intero positivo, è completo.

Parole incompletabili minimali

Consideriamo un insieme X non completo.

Una parola $w \in A^*$ si dice **incompletabile in X^*** se $w \notin \text{Fact}(X^*)$.

Consideriamo un insieme X non completo.

Una parola $w \in A^*$ si dice **incompletabile in X^*** se $w \notin \text{Fact}(X^*)$.

Problem

*Se X è non completo, qual è la lunghezza della **più corta** parola incompletabile in X^* ?*

Parole incompletabili minimali

Sia k la lunghezza della più lunga parola in X .

Sia m la lunghezza della più corta parola incompletabile in X^* .

Parole incompletabili minimali

Sia k la lunghezza della più lunga parola in X .

Sia m la lunghezza della più corta parola incompletabile in X^* .

Congettura (A. Restivo – 1980)

Se X è non completo, $m \leq 2k^2$.

Parole incompletabili minimali

Sia k la lunghezza della più lunga parola in X .

Sia m la lunghezza della più corta parola incompletabile in X^* .

Congettura (A. Restivo – 1980)

Se X è non completo, $m \leq 2k^2$.

Tale congettura è basata sul fatto che da risultati sperimentali la più corta parola incompletabile ha sempre la stessa struttura:

$$v u_1 v u_2 \dots u_{k-1} v$$

dove $|v| = k$ e $|u_i| = k - 1$.

Con J. Sakarovitch abbiamo adattato il programma **Vaucanson** per fare degli esperimenti.

Con J. Sakarovitch abbiamo adattato il programma **Vaucanson** per fare degli esperimenti.

Usando una classe di esempi trovata da E. Pribavkina siamo riusciti a trovare una **classe di controesempi** alla congettura (per $A = \{a, b\}$)

Parole incompletabili minimali

$$X(k) = \{A^k \cup b(\sum_{i=1}^{k-2} a^i)A \cup Aba^{k-4}A \cup Aba \cup (\sum_{i=1}^{k-3} a^i)b \cup b^4\} \setminus \{a^{k-2}bb\}$$

Parole incompletabili minimali

$$X(k) = \{A^k \cup b(\sum_{i=1}^{k-2} a^i)A \cup Aba^{k-4}A \cup Aba \cup (\sum_{i=1}^{k-3} a^i)b \cup b^4\} \setminus \{a^{k-2}bb\}$$

$$\begin{aligned} w(k) = & a^{k-2}bb \ a^{k-1} \ a^{k-2}bb \ ba^{k-4} \\ & (a^{k-2}bb \ ba \ a^{k-2}bb \ bba^{k-5})^{k-6} \\ & a^{k-2}bb \ ab \ a^{k-2}bb \ bba^{k-3} \ a^{k-2}bb \ ba^{k-3} \ a^{k-2}bb \end{aligned}$$

$$m(w) = 3k^2 - 9k + 1$$

Parole incompletabili minimali

$$X(k) = \{A^k \cup b(\sum_{i=1}^{k-2} a^i)A \cup Aba^{k-4}A \cup Aba \cup (\sum_{i=1}^{k-3} a^i)b \cup b^4\} \setminus \{a^{k-2}bb\}$$

$$\begin{aligned} w(k) = & a^{k-2}bb \ a^{k-1} \ a^{k-2}bb \ ba^{k-4} \\ & (a^{k-2}bb \ ba \ a^{k-2}bb \ bba^{k-5})^{k-6} \\ & a^{k-2}bb \ ab \ a^{k-2}bb \ bba^{k-3} \ a^{k-2}bb \ ba^{k-3} \ a^{k-2}bb \end{aligned}$$

$$m(w) = 3k^2 - 9k + 1$$

I - Parole incompletabili minimali

II - Automa dei suffissi, parole proibite minimali e wildcard

II - Automa dei suffissi, parole proibite minimali e wildcard

Sia A un alfabeto finito. La **wildcard** N rappresenta un simbolo qualsiasi di A .

Sia A un alfabeto finito. La **wildcard** N rappresenta un simbolo qualsiasi di A .

Una **parola con wildcard** è un elemento di $(A \cup \{N\})^*$.

Sia A un alfabeto finito. La **wildcard** N rappresenta un simbolo qualsiasi di A .

Una **parola con wildcard** è un elemento di $(A \cup \{N\})^*$.

Example

La parola con wildcard $w_N = aaNbN$ corrisponde all'insieme $P_{w_N} = \{aaaba, aaabb, aabba, aabbb\}$.

L'automa dei suffissi di w_N è l'automa dei suffissi di P_{w_N} .

L'automa dei suffissi di w_N è l'automa dei suffissi di P_{w_N} .

Esiste una procedura per calcolarlo (indiretta) in tempo $O(|P_{w_N}|) = O(|A|^{\# \text{ di } N \text{ in } w_N})$.

L'**automa dei suffissi di w_N** è l'automa dei suffissi di P_{w_N} .

Esiste una procedura per calcolarlo (indiretta) in tempo $O(|P_{w_N}|) = O(|A|^{\# \text{ di } N \text{ in } w_N})$.

Vogliamo dare una procedura di calcolo **diretta** (sequenziale) e che prenda un **tempo inferiore** (almeno in certi casi).

Una parola $v = a_0 a_1 \dots a_n$ si dice **proibita minimale** per un insieme di parole P se:

- $v \notin \text{Fact}(P)$
- $a_0 a_1 \dots a_{n-1}$ e $a_1 a_2 \dots a_n \in \text{Fact}(P)$

Una parola $v = a_0 a_1 \dots a_n$ si dice **proibita minimale** per un insieme di parole P se:

- $v \notin \text{Fact}(P)$
- $a_0 a_1 \dots a_{n-1}$ e $a_1 a_2 \dots a_n \in \text{Fact}(P)$

Example

Sia $P = \{abaa, aaba\}$. Allora $v = baab$ è una parola proibita minimale per P .

L'insieme $\mathcal{MF}(P)$ delle p.p.m. per P è antifattoriale $\Rightarrow$ può essere rappresentato in un trie.

L'insieme $\mathcal{MF}(P)$ delle p.p.m. per P è antifattoriale $\Rightarrow$ può essere rappresentato in un trie.

Vogliamo dare un algoritmo per calcolare il trie di $\mathcal{MF}(P)$ a partire dall'automa dei suffissi di P .

L'insieme $\mathcal{MF}(P)$ delle p.p.m. per P è antifattoriale $\Rightarrow$ può essere rappresentato in un trie.

Vogliamo dare un algoritmo per calcolare il trie di $\mathcal{MF}(P)$ a partire dall'automa dei suffissi di P .

Inoltre vogliamo che questo algoritmo sia lineare nella taglia del trie di output.

Perché tutto ciò?

Perché tutto ciò?

Le parole con wildcard si possono usare per rappresentare sequenze con **incertezze di lettura**.

Perché tutto ciò?

Le parole con wildcard si possono usare per rappresentare sequenze con **incertezze di lettura**.

Le parole proibite minimali si possono usare per ricostruire sequenze da un insieme di frammenti (**fragment assembly**).