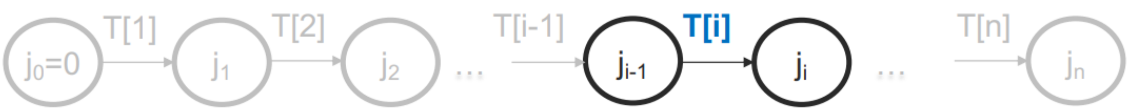
Lezione 13 23/11/2023

Esercizi sulla ricerca esatta negli automi

gli esercizi in esame saranno più semplici (minimi) di questi.

Durante la ricerca esatta del pattern acbdccbd con automa a stati finiti, si arriva allo stato 6 dopo avere letto un certo simbolo del testo. Specificare tale simbolo.



$$\delta(j_{i-1},\mathbf{T[i]}) = j_i \quad \Leftrightarrow \quad P[1,j_i] = T[i- j_i+1, i]$$

$$\delta(j_{i-1},T[i]) = j_i \quad \textbf{and} \quad j_i \le j_{i-1} \quad \Rightarrow \quad j_i = |B(P[1, j_{i-1}]T[i])|$$

Dopo avere letto T[i], si arriva allo stato j_i <u>se e solo se</u> il più lungo prefisso di P che ha *match* esatto con una sottostringa di T che finisce in posizione i (e inizia in posizione i - j_i + 1) è P[1, j_i]. Ne deriva che P[j_i] è uguale a T[i]

In altre parole, lo stato j_i di arrivo dopo avere letto T[i] fornisce la lunghezza del più lungo prefisso di P che ha *match* esatto con una sottostringa di T che finisce in posizione i

Viceversa, se si sa che il più lungo prefisso di P, che ha *match* esatto con una sottostringa di T in posizione i, ha lunghezza j_i, allora si arriverà allo stato j_i dopo avere letto T[i]

2. Se a partire da uno stato j_{i-1} si arriva a uno stato <u>successivo</u> $j_i = j_{i-1}+1$ dopo avere letto il simbolo T[i], significa (per definizione della funzione di transizione δ) che il simbolo P[j_i] è uguale a T[i].

3. Se a partire da uno stato j_{i-1} si arriva a uno stato <u>precedente</u> $j_i \le j_{i-1}$ dopo avere letto il simbolo T[i], significa (per definizione della funzione di transizione δ) che la lunghezza del bordo di P[1, j_{i-1}] concatenato a T[i] è pari a j_i

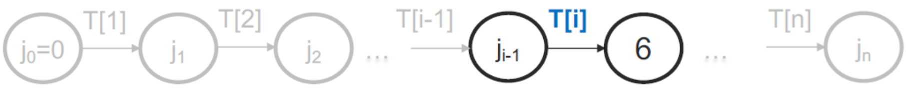
Esercizio 1

Durante la ricerca esatta del pattern acbdccbd con automa a stati finiti, si arriva allo stato 6 dopo avere letto un certo simbolo del testo. Specificare tale simbolo.

T è la stringa, quindi estraggo la sottostringa.

P è il pattern.

P = acbdc<u>c</u>bd



Il fatto di arrivare allo stato 6 dopo avere letto **T[i]** implica che P[1,6] ha *match* esatto con una sottostringa di T che finisce in posizione i. Di conseguenza l'ultimo simbolo della sottostringa coincide con l'ultimo simbolo del prefisso. Cioè:

T[i] = P[6] = **c**

Esercizio 2

L'esecuzione per la ricerca esatta del pattern acbdacad con automa si trova allo stato 6. Che simbolo del testo viene letto dopo, se l'algoritmo passa allo stato successivo?

P = acbdac<u>a</u>d

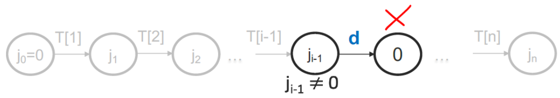
Si sa che dallo stato 6 si arriva allo stato <u>successivo</u>, cioè 7, dopo avere letto **T[i]**.
Il fatto di arrivare allo stato 7 dopo avere letto **T[i]** implica che P[1,7] ha *match* esatto con una sottostringa di T che finisce in posizione i. Di conseguenza l'ultimo simbolo della sottostringa coincide con l'ultimo simbolo del prefisso. Cioè:

T[i] = P[7] = **a**

Esercizio 3

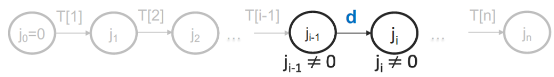
Dato il pattern dccdbcd, si può dire che l'esecuzione dell'automa su un determinato testo non arriva mai allo stato 0, dopo avere letto il simbolo **d** a partire da uno stato diverso da 0? In caso affermativo, specificare quali sono i possibili stati di arrivo.

P = dccdbcd



Supponiamo di arrivare a $j_i = 0$ a partire da $j_{i-1} \neq 0$ dopo avere letto un simbolo **d** in posizione i di T. Sicuramente j_i non può essere stato <u>successivo</u> a j_{i-1}. Di conseguenza (per definizione di δ) ne deriva che $|B(P[1, j_{i-1}]d)| = 0$, cioè $B(P[1, j_{i-1}]d) = \varepsilon$ per qualsiasi prefisso non nullo $P[1, j_{i-1}]$ (visto che $j_{i-1} \neq 0$).
Però, il fatto che P[1] è uguale al simbolo **d** implica che la concatenazione di qualsiasi prefisso non nullo di P con il simbolo **d** produce un bordo non nullo. Quindi non è vero che $B(P[1, j_{i-1}]d) = \varepsilon$ per qualsiasi prefisso non nullo $P[1, j_{i-1}]$ e perciò non si arriva mai allo stato 0 a partire da uno stato $\neq 0$ dopo avere letto il simbolo **d**.

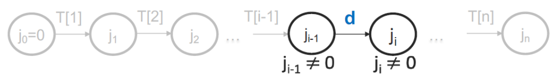
P = dccdbcd



Quali sono i possibili stati di arrivo j_i?
Il fatto di arrivare a $j_i \neq 0$ dopo avere letto un simbolo **d** in posizione i di T implica che il prefisso non nullo $P[1,j_i]$ ha *match* esatto con una sottostringa di T che finisce in posizione i e $P[j_i]$ è uguale a **d**.

Osserviamo che ogni prefisso non nullo di P che finisce con il simbolo **d** e che ha *match* esatto con una sottostringa di T che finisce in i, produce quindi uno stato di arrivo j_i pari alla sua lunghezza. E' quindi sufficiente trovare tutti i prefissi non nulli di P che finiscono con **d**.

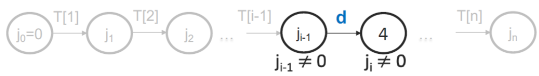
P = <u>dccdbcd</u>



Quali sono i possibili stati di arrivo j_i?

$j_i = 1$ prodotto da P[1,1]
$j_i = 4$ prodotto da P[1,4]
$j_i = 7$ prodotto da P[1,7] = P

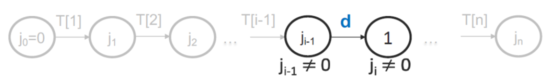
P = dccdbcd



Allo stato $j_i = 4$ posso arrivare da uno stato $j_{i-1} \geq 4$?

Il fatto di arrivare a 4 da uno stato $j_{i-1} \geq 4$ dopo avere letto un simbolo **d** implica (per definzione di δ) che $|B(P[1,j_{i-1}]d)|$ è uguale a 4.
Di conseguenza, si può arrivare allo stato 4 da $j_{i-1} \geq 4$ solo se esiste un prefisso di P di lunghezza ≥ 4 che concatenato a **d** produce un bordo lungo 4. Nel pattern in questione un tale prefisso non esiste e pertanto non si può arrivare allo stato 4 a partire da $j_{i-1} \geq 4$.
Allo stato 4 si arriva (dopo avere letto **d**) solo a partire dallo stato 3.

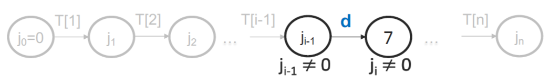
P = dccdbcd



Allo stato $j_i = 1$ posso arrivare da uno stato $j_{i-1} \geq 1$?

Il fatto di arrivare a 1 da uno stato $j_{i-1} \geq 1$ dopo avere letto un simbolo **d** implica (per definzione di δ) che $|B(P[1,j_{i-1}]\mathbf{d})|$ è uguale a 1.
Di conseguenza, si può arrivare allo stato 1 da $j_{i-1} \geq 1$ solo se esiste un prefisso di P di lunghezza ≥ 1 che concatenato a **d** produce un bordo lungo 1. Nel pattern in questione tutti i prefissi non nulli concatenati a **d** producono un bordo lungo 1.
In conclusione, allo stato $j_i = 1$ posso arrivare da uno stato $j_{i-1} \geq 1$.

P = dccdbcd

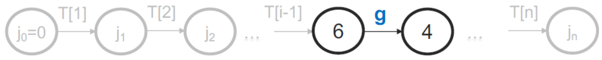


Allo stato $j_i = 7$ posso arrivare da uno stato $j_{i-1} \geq 7$?

Allo stato 7 si può solo arrivare a partire da uno stato 6, leggendo un simbolo **d**.
Infatti non sono possibili stati $j_{i-1} > 7$ in quanto P è lungo 7 e non è nemmeo possibile uno stato $j_{i-1} = 7$ in quanto $|B(P[1,7]\mathbf{d})| \neq 7$.

Esercizio 4

Durante la ricerca esatta di un pattern P con automa a stati finiti, si passa dallo stato 6 allo stato 4 dopo avere letto il simbolo g nel testo. Dimostrare che P[1] = g.



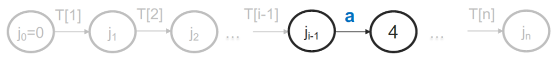
Il fatto di arrivare allo stato 4 dopo avere letto **g** in posizione i di T implica che il prefisso P[1,4] ha *match* esatto con una sottostringa di T che finisce in posizione i. Questa sottostringa inizia quindi in posizione i-4+1=i-3 ed è T[i-3, i]. Come conseguenza si ha P[1] = T[i-3] e P[4]=T[i] = **g**.
Inoltre il fatto di essere arrivati allo stato 6 dopo avere letto il simbolo T[i-1] implica che il prefisso P[1,6] ha *match* esatto con una sottostringa di T che finisce in posizione i-1. Questa sottostringa inizia quindi in posizione (i-1)-6+1=i-6 ed è T[i-6, i-1]. Come conseguenza si ha P[4] = T[i-3].

Ne deriva che P[1] = T[i-3] = P[4] = **g**.

Esercizio 5

Dato il pattern accabaccba, l'automa arriva allo stato 4 dopo avere letto il simbolo a. Determinare i possibili stati di partenza.

P = accabaccba



Sicuramente sarà (per definizione di δ) $j_{i-1} \geq 3$:

- $j_{i-1} = 3$ implica una transizione allo stato <u>successivo</u> 4 dopo avere letto un simbolo **a**
- $j_{i-1} \geq 4$ implica una transizione allo stato <u>precedente</u> 4 dopo avere letto un simbolo **a**

P = <u>acca</u>baccba

Il caso $j_{i-1} = 3$ è possibile?

Il fatto di essere arrivati a 3 dopo avere letto T[i-1] implica che il prefisso P[1,3] ha *match* esatto con una sottostringa di T che finisce in posizione i-1. Di conseguenza si può passare allo stato successivo 4 solo se viene letto un simbolo **T[i]** = P[4]. Siccome il simbolo letto dal testo in posizione i è **a** e P[4] è uguale ad **a**, si ha che $j_{i-1} = 3$ è possibile.

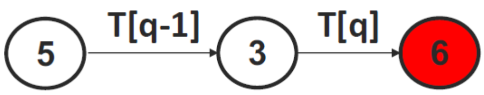
P = accabaccba

Il caso $j_{i-1} \geq 4$ è possibile?

Il fatto di passare da $j_{i-1} \geq 4$ a 4 leggendo un simbolo **a**, implica (per definizione di δ) che $|B(P[1, j_{i-1}]\mathbf{a})| = 4$.
Quindi, tutti i prefissi di P di lunghezza ≥ 4 che, concatenati al simbolo **a** forniscono un bordo di lunghezza 4, producono una transizione da $j_{i-1} \geq 4$ a 4 con il simbolo **a**.
Nel pattern in questione un solo prefisso di lunghezza ≥ 4 ha questa proprietà ed è P[1,8]. Di conseguenza, l'unico stato $j_{i-1} \geq 4$ che permette di arrivare a 4 leggendo un simbolo **a** è 8.

Esercizio 6

Dire, motivando la risposta, se la seguente catena si può verificare durante l'esecuzione di un automa a stati finiti per la ricerca esatta di un pattern.



NO, sono possibili solo transizioni da uno stato iniziale j_{i-1} a uno stato finale $j_{i-1}+1$ oppure $\leq j_{i-1}$

Esercizio 7

Dire, motivando la risposta, se la seguente catena si può verificare durante l'esecuzione di un automa a stati finiti per la ricerca esatta di un pattern. Se non è possibile, renderla «plausibile» correggendo uno dei tre simboli di transizione. Supporre che l'ultimo simbolo letto sia in posizione i del testo

Arrivo allo stato 3 leggendo T[i] = **b**

⇒ P[1,3] = T[i-2, i] = **abb**

Arrivo allo stato 4 leggendo T[i-1] = **b**

⇒ P[1,4] = T[i-4, i-1] = ****ab**

Arrivo allo stato 5 leggendo T[i-2] = **a**

⇒ P[1,5] = T[i-6 ,i-2] = ******a**

Da P[1,3] = **abb** deduco che i primi tre simboli di P sono **a**, **b** e **b**
Da P[1,4] = ****ab** deduco che il terzo e il quarto simbolo di P sono **a** e **b**
Da P[1,5] = ******a** deduco che il quinto simbolo di P è **a**

Ho una contraddizione tra P[1,3] che mi dice che il terzo simbolo è **b** e P[1,4] che dice che il terzo simbolo è **a**

Devo quindi correggere la **b** in posizione i in un simbolo a oppure la **a** in posizione i-2 in un simbolo b.

Proviamo a correggere **b** in posizione i in un simbolo a.

⇒ P[1,5] = ababa

Il fatto di passare dallo stato 5 allo stato 4 leggendo T[i-1] = **b** implica che |B(P[1,5]**b**)| = 4, cioè |B(ababa**b**)| = 4. Questa affermazione è vera in quanto il bordo di ababa**b** è abab ed è lungo 4.
Di conseguenza lo stato 4 dopo avere letto **b** in posizione i-1 è corretto.

A questo punto, la transizione dallo stato 4 allo stato 3 dopo avere letto T[i] = **a** è corretta?
La transizione da 4 a 3 **non** è corretta, perché dallo stato 4 leggendo un simbolo **a** si deve arrivare allo stato successivo 5, in quanto P[5] = **a**.

Torniamo all'inizio e proviamo ora a correggere **a** in posizione i-2 in un simbolo b

⇒ P[1,5] = bbbbb

Il fatto di passare dallo stato 5 allo stato 4 leggendo T[i-1] = **b** implica che |B(P[1,5]**b**)| = 4, cioè |B(bbbbb**b**)| = 4.
Questa affermazione **non** è vera in quanto il bordo di bbbbb**b** è bbbbb ed è lungo 5.
Di conseguenza lo stato 4 dopo avere letto **b** in posizione i-1 deve essere corretto in 5.

Il fatto poi di passare dallo stato 5 allo stato 4 leggendo T[i] = **b** implica che |B(P[1,5]**b**)| = 3, cioè |B(bbbbb**b**)| = 3.
Questa affermazione **non** è vera in quanto il bordo di bbbbb**b** è bbbbb ed è lungo 5.
Di conseguenza lo stato 3 dopo avere letto **b** in posizione i deve essere corretto in 5.