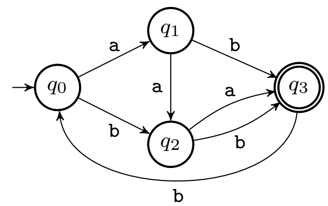
Automi e Linguaggi (M. Cesati)

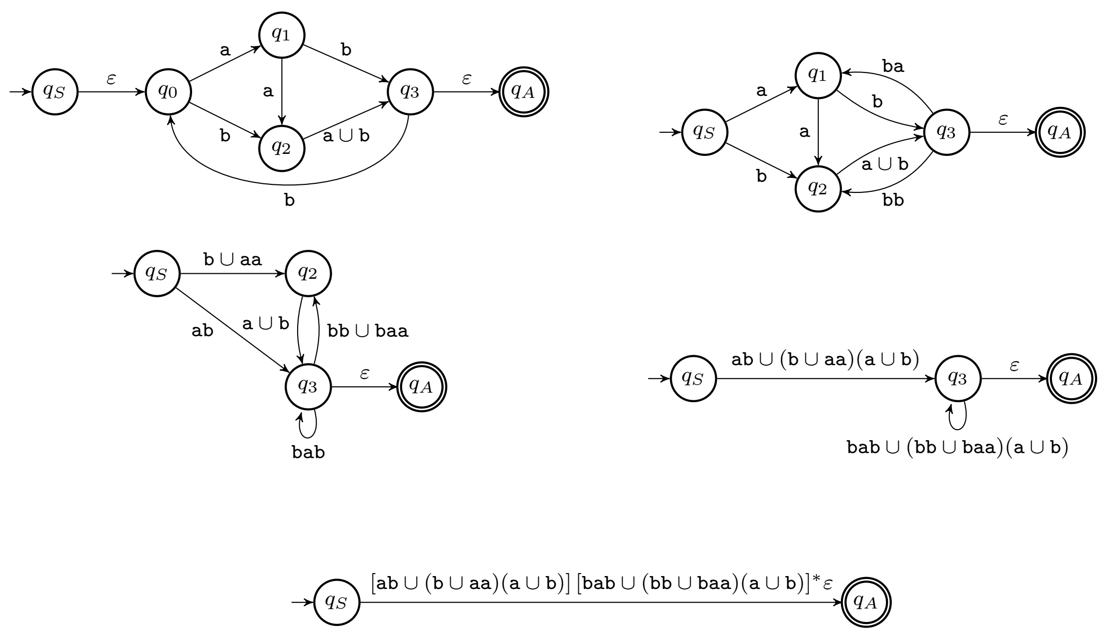
Facoltà di Ingegneria, Università degli Studi di Roma Tor Vergata

Compito scritto del 13 luglio 2022

Esercizio 1 [6] Determinare una espressione regolare per il linguaggio riconosciuto dal DFA:



Soluzione: Convertiamo il DFA in un GNFA e rimuoviamo nell'ordine i nodi q_0, q_1, q_2 e q_3. Si ottiene:

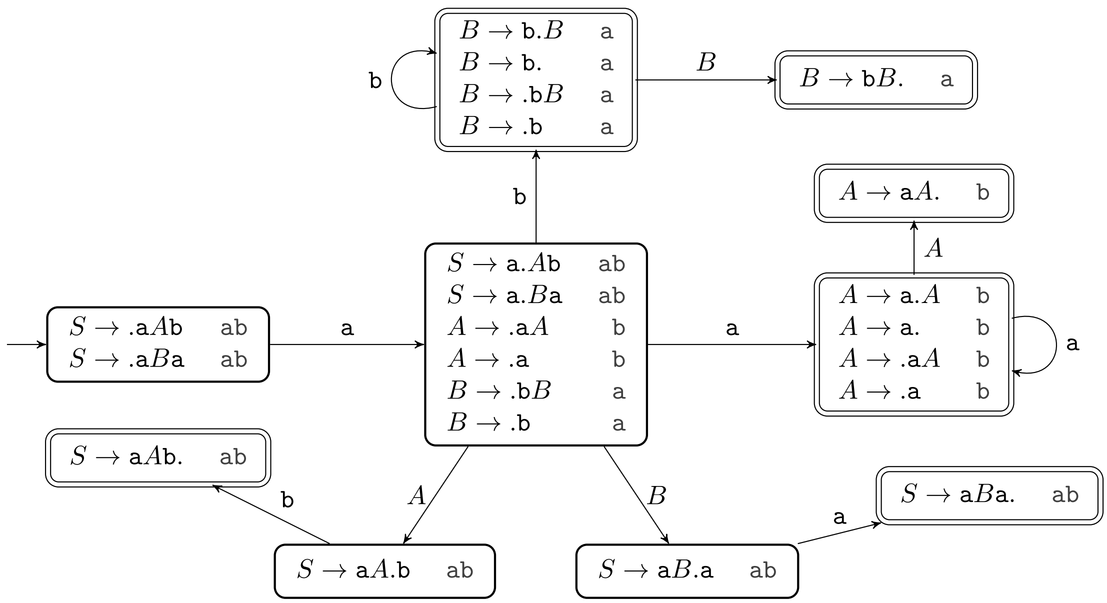


Semplificando l'espressione regolare, poiché $(\mathtt{bb} \cup \mathtt{baa}) = \mathtt{b}\,(\mathtt{b} \cup \mathtt{aa})$ e $\mathtt{bab} \cup \mathtt{b} \cdots = \mathtt{b}(\mathtt{ab} \cup \cdots)$, si ottiene:

$$[\mathtt{ab} \cup (\mathtt{b} \cup \mathtt{aa})(\mathtt{a} \cup \mathtt{b})]\,\{\mathtt{b}\ [\mathtt{ab} \cup (\mathtt{b} \cup \mathtt{aa})(\mathtt{a} \cup \mathtt{b})]\}^*$$

Esercizio 2 [6] Si consideri la grammatica G con variabile iniziale S e regole $S \to \mathtt{a}\,A\,\mathtt{b} \mid \mathtt{a}\,B\,\mathtt{a}$, $A \to \mathtt{a}\,A \mid \mathtt{a}$, $B \to \mathtt{b}\,B \mid \mathtt{b}$. Determinare se la grammatica G è LR(1) e se G è deterministica.

Soluzione: Per determinare se la grammatica G è LR(1) costruiamo il DFA DK_1; questo automa, scartando i simboli di lookahead, serve anche a determinare se G è LR(0) (deterministica).



La grammatica G è LR(1), in quanto nessuno degli stati di accettazione contiene due regole tra loro consistenti. Infatti, tutti gli stati di accettazione hanno una unica regola completata; inoltre, in ogni regola non completata in cui il punto è seguito da un simbolo terminale, questo simbolo non è incluso tra i simboli di lookhead della regola completata.

D'altra parte lo stesso automa testimonia che G non è deterministica, in quanto esistono stati di accettazione contenenti regole non completate in cui il punto è seguito da un simbolo terminale.

Esercizio 3 [6] Dimostrare che se A e B sono linguaggi liberi dal contesto (CFL), allora il linguaggio $C = \{x^n y^n \mid n \geq 0, x \in A, y \in B\}$ è libero dal contesto.

Soluzione: Non è possibile esibire la dimostrazione richiesta in quanto l'asserto è falso. Infatti si consideri il seguente contro-esempio. Sia $A = \{0^h\,1^h \mid h \geq 0\}$; sappiamo che A è CFL. Sia $B = \{2\}$, ossia l'insieme costituito dall'unico simbolo 2. Poiché B è finito, B è regolare e quindi anche CFL. Il linguaggio C definito come nel testo per questi A e B particolari è $C = \{(0^h 1^h)^n 2^n \mid n, h \geq 0\}$.

Se per assurdo C fosse CFL, allora dovrebbe valere per esso il pumping lemma. Sia dunque $p > 0$ la lunghezza associata a C, e sia $s = (0^p 1^p)^3 2^3 = 0^p 1^p 0^p 1^p 0^p 1^p 222$. Ovviamente $|s| = 6p + 3 > p$ e $s \in C$ ($n = 3$, $h = p$). Per il pumping lemma deve esistere una suddivisione $s = uvxyz$ con $|vxy| \leq p$, $|vy| > 0$ e $uv^i xy^i z \in C$ per ogni $i \geq 0$. Si considerino ora le seguenti due alternative. Se la suddivisione è tale che pompando non si modificano il numero di 2 nella stringa, allora per poter far parte del linguaggio la stringa pompata deve modificare tutte e tre le copie di $0^p 1^p$ in testa alla stringa. D'altra parte, la condizione $|vxy| \leq p$ consente di modificare soltanto una o due di tali copie, e quindi la stringa pompata non può far parte di C.

L'alternativa è che la suddivisione consenta di modificare il numero di 2 alla fine della stringa; in tal caso però la condizione $|vxy| \leq p$ consente di modificare solo l'ultima occorrenza di 1^p nella stringa, perciò la stringa pompata avrà la forma $0^p1^p0^p1^p0^p1^q2^r$, e qualunque siano i valori di q e r tale stringa non può far parte di C. Perciò il pumping lemma non vale e C non è CFL.

Esercizio 4 [7] Sia A un qualsiasi linguaggio libero dal contesto (CFL) contenente stringhe di lunghezza pari. Per ciascuna stringa $w = c_1c_2 \cdots c_{2m} \in A$, considerare la stringa $w^{\#} = c_1c_3 \cdots c_{2m-1}c_2c_4 \cdots c_{2m}$. Dimostrare che il linguaggio $A^{\#} = \{w^{\#} \mid w \in A\}$ non è necessariamente libero dal contesto.

Soluzione: La dimostrazione consiste semplicemente nell'esibire un contro-esempio. Consideriamo il linguaggio $D = \{(\mathtt{ab})^n\mathtt{c}^{2n} \mid n \geq 0\}$. È immediato verificare che D è CFL, in quanto ad esempio omomorfo al linguaggio CFL $\{0^n1^n \mid n \geq 0\}$, con $h(0) = \mathtt{ab}$ e $h(1) = \mathtt{cc}$.

Il linguaggio $D^{\#}$ è quindi $\{\mathtt{a}^n\mathtt{c}^n\mathtt{b}^n\mathtt{c}^n \mid n \geq 0\}$, che è non CFL. Supponiamo per assurdo che lo sia, dunque valga per esso il pumping lemma. Sia $p > 0$ tale che ogni stringa s di $D^{\#}$ di dimensione $\geq p$ è suddivisibile in modo da essere "pompata". Consideriamo quindi $s = \mathtt{a}^p\mathtt{c}^p\mathtt{b}^p\mathtt{c}^p \in D^{\#}$ con $|s| = 4p > p$. Per qualunque suddivisione $s = uvxyz$ con $|vxy| \leq p$:

1. Se ciascuna stringa v e y contiene al più solo tipo di simbolo, qualunque stringa uv^ixy^iz con $i \neq 0$ non può far parte di $D^{\#}$ in quanto due sottosequenze di simboli uguali conservano certamente lunghezza p mentre almeno un'altra sottosequenza assume lunghezza diversa da p ($|vy| > 0$)
2. Se almeno una delle stringhe v e y contiene più di un tipo di simbolo, la stringa uv^ixy^iz con $i \neq 0$ non può far parte di $D^{\#}$ in quanto la stringa non farebbe parte di $\mathtt{a}^*\mathtt{c}^*\mathtt{b}^*\mathtt{c}^*$.

Poiché s non può essere suddivisa in modo conforme al pumping lemma, il pumping lemma non è valido, e quindi $D^{\#}$ non può essere CFL.

Esercizio 5 [7] Sia A un linguaggio decidibile prefissato, con $A \neq \Sigma^*$ e $A \neq \emptyset$. Dimostrare che un qualsiasi linguaggio L è decidibile se e solo se L riduce tramite Turing ad A (ossia se $L \leq_T A$).

Soluzione: Sia A un qualunque linguaggio decidibile prefissato. È richiesto di dimostrare che (1) ogni linguaggio decidibile riduce tramite Turing ad A, e che (2) se un linguaggio riduce tramite Turing ad A allora esso è decidibile.

Per la prima parte, sia L un linguaggio decidibile, e sia quindi M una TM che decide L. È possibile considerare M come una TM M^A con oracolo A che non fa alcuna domanda all'oracolo (ovvero ignora le risposte date dall'oracolo). Pertanto banalmente M^A decide L, e quindi $L \leq_T A$.

Per la seconda parte, consideriamo un linguaggio L che riduce tramite Turing ad A, ossia $L \leq_T A$. Pertanto, esiste una TM con oracolo M^A che decide L. Poiché A è decidibile, esiste anche una TM T che decide A. Consideriamo ora la TM N ottenuta da M^A sostituendo ogni interrogazione dell'oracolo per A con la simulazione della TM T che decide A. Le risposte fornite dalla simulazione di T sono esattamente le stesse di quelle fornite dall'oracolo per A, dunque su ogni istanza x, $M^A(x)$ fornisce lo stesso risultato di $N(x)$. Possiamo dunque concludere che N decide il linguaggio L.

Esercizio 6 [8] Si consideri un grafo non diretto $G = (V, E)$ con $|V| = n$ nodi e $|E| = m$ archi. Si dimostri che il problema di stabilire se il grafo G contiene un sottoinsieme W di nodi con $|W| \leq \frac{n}{2}$ tale che ogni arco di G contiene almeno un nodo in W è NP-completo.

Soluzione: Questo problema è generalmente noto con il nome HALF VERTEX COVER (HVC). Dimostrare che HVC è in NP è molto semplice. Data una istanza $\langle G \rangle$, un certificato per l'esistenza di una soluzione è una lista di nodi del grafo G. Il verificatore controlla che nella lista non vi siano nodi ripetuti, che la lista contenga non più di $|V(G)|/2$ nodi, e che ciascun arco di G sia collegato ad un nodo della lista. Tutti questi controlli possono naturalmente essere effettuati in tempo polinomiale nella dimensione del grafo.

Esibiamo ora una riduzione tra INDEPENDENT SET (IS) e HVC. Si presti attenzione che le istanze di IS e di HVC hanno una differente struttura: le istanze di IS codificano un grafo G ed un numero intero k, mentre le istanze di HVC codificano soltanto un grafo. Dunque la riduzione deve "eliminare" il parametro k dall'istanza di IS.

La riduzione considera il valore del parametro k ed opera differentemente nei casi $k = n/2$, $k < n/2$, e $k > n/2$. Si ricordi che una riduzione è una funzione calcolabile in tempo polinomiale, e confrontare due valori interi è una operazione eseguibile in tempo polinomiale nella lunghezza delle codifiche dei valori.

1. $\underline{k = n/2}$: data l'istanza di IS $\langle G, k \rangle$, la riduzione costruisce l'istanza $\langle G \rangle$ (si elimina semplicemente il valore k dall'istanza). Infatti, una istanza-sì di IS è tale per cui esiste un sottoinsieme $W \subseteq V$ con $|W| \geq k = n/2$ tale che non esiste alcun arco tra i nodi in W; l'insieme $U = V \setminus W$ è un ricoprimento tramite vertici con $|U| \leq n - k = n/2$; il grafo è dunque anche una istanza-sì di HVC. Ovviamente vale anche il viceversa.

2. $\underline{k < n/2}$: data l'istanza di IS $\langle G, k \rangle$, la riduzione costruisce l'istanza $\langle G' \rangle$ in cui il grafo G' è costituito dal grafo G al quale sono aggiunti $n - 2k$ nodi isolati. Sia $\langle G, k \rangle$ una istanza-sì di IS, dunque esista un insieme indipendente $W \subseteq V$ con $|W| \geq k$. Nel grafo G', W unito ai nuovi nodi costituisce un insieme indipendente W' di dimensione $k+(n-2k) = n-k$ nodi. Pertanto $U = V(G') \setminus W' = V \setminus W$ ha $|U| = n-k$ nodi ed è un

ricoprimento tramite vertici per G'. D'altra parte, G' contiene $n + (n - 2k) = 2(n - k)$ nodi, quindi G' è una istanza-sì di HVC.

Se invece $\langle G, k \rangle$ è una istanza-no di IS, non esiste alcun insieme indipendente di dimensione maggiore o uguale a k in G. Nel grafo G' pertanto non esiste alcun insieme indipendente di dimensione maggiore o uguale a $n-k$, perché G' è ottenuto da G aggiungendo solo $n-2k$ nodi e non rimuovendo alcun arco. Dunque in G' nessun ricoprimento tramite vertici può essere di dimensione minore o uguale a $2(n - k) - (n - k) = n - k$, e quindi $\langle G' \rangle$ è una istanza-no di HVC.

3. $\underline{k > n/2}$: data l'istanza di IS $\langle G, k \rangle$, la riduzione costruisce l'istanza $\langle G' \rangle$ in cui il grafo G' è costituito dal grafo G al quale sono aggiunti $2k - n$ nodi ed i seguenti archi: tutti i nuovi nodi sono collegati tra loro (quindi costituiscono un insieme completo), inoltre ciascuno dei nuovi nodi è collegato a ciascuno dei nodi del grafo G. Ovviamente, una istanza-sì di IS è anche una istanza-sì di G'. Infatti G' ha $n + (2k - n) = 2k$ nodi, e l'insieme indipendente W con $|W| \geq k$ è anche un insieme indipendente di G' con almeno la metà dei nodi di G'; pertanto il complemento di questo insieme è un ricoprimento tramite vertici con al più la metà dei nodi di G'.

   Consideriamo ora una istanza-no di IS, dunque supponiamo che nel grafo G non esista alcun insieme indipendente di dimensione maggiore o uguale a k. Il grafo G' è ottenuto da G aggiungendo un sottografo completo ed aggiungendo tutti gli archi tra i vecchi ed i nuovi nodi. Pertanto, nessuno dei nuovi nodi può essere aggiunto ad un insieme indipendente di G. Si conclude dunque che non può esistere in G' un insieme indipendente avente la metà dei nodi di G'. Pertanto, ogni ricoprimento tramite vertici in G' deve contenere più della metà dei nodi di G', dunque $\langle G' \rangle$ è una istanza-no di HVC.

È facile verificare che la dimensione dell'istanza $\langle G' \rangle$ è polinomialmente limitata dalla dimensione dell'istanza $\langle G, k \rangle$ (infatti, $k \leq n$, quindi in ogni caso $|V(G')| \leq 2n$). Inoltre la riduzione è calcolabile in tempo polinomiale. Quindi HVC è NP-hard, e di conseguenza NP-completo.