[redacted]guaggi Formali e Traduttori

[redacted].6 Pumping lemma per i linguaggi liberi

Sommario

Per dimostrare che un linguaggio **è libero**, basta esibire un automa a pila (PDA) che lo riconosce, oppure una grammatica libera che lo genera. L'incapacità di trovare siffatto automa o siffatta grammatica non è una dimostrazione del fatto che il linguaggio non è libero.

In questa lezione rispondiamo alle seguenti domande:

1. Esistono linguaggi **non** liberi?
2. Se sì, come dimostro che un linguaggio **non** è libero?

Linguaggi **non** liberi

- Cerchiamo una proprietà P soddisfatta da tutti i linguaggi liberi:

$$L \text{ libero} \Rightarrow L \text{ soddisfa } P$$

- Se troviamo un linguaggio L che **non soddisfa** P, allora per **contrapposizione** possiamo concludere che L **non è libero**:

$$L \text{ non soddisfa } P \Rightarrow L \text{ non è libero}$$

Pumping lemma per linguaggi liberi

Teorema

Per ogni linguaggio libero L esiste $n \in \mathbb{N}$ tale che, per ogni $z \in L$ con $|z| \geq n$, esistono u, v, w, x e y tali che $z = uvwxy$ e inoltre:

(1) $vx \neq \varepsilon$
(2) $|vwx| \leq n$
(3) $uv^k wx^k y \in L$ per ogni $k \geq 0$.

In prosa

- Ogni stringa z "sufficientemente lunga" ($|z| \geq n$) di un linguaggio libero L ...
- ... contiene due sottostringhe <u>non entrambe vuote</u> (1) ...
- ... e "non troppo distanti" una dall'altra (2) ...
- ... che possono essere <u>entrambe eliminate</u> ($k = 0$) ...
- ... o <u>replicate</u> a piacere <u>lo stesso numero di volte</u> ($k > 0$) ...
- ... consentendoci di trovare altre stringhe di L (3).

Il linguaggio a^{k}b^{k}c^k non è libero

Dimostriamo che $L = \{a^k b^k c^k \mid k \geq 0\}$ non è libero facendo vedere che per L il pumping lemma non vale.

Il linguaggio $a^k b^k c^k$ non è libero

Dimostriamo che $L = \{a^k b^k c^k \mid k \geq 0\}$ non è libero facendo vedere che per L il pumping lemma non vale.

Supponiamo, per assurdo, che esista n con le proprietà enunciate nella slide 4.

Il linguaggio $a^k b^k c^k$ non è libero

Dimostriamo che $L = \{a^k b^k c^k \mid k \geq 0\}$ non è libero facendo vedere che per L il pumping lemma non vale.

Supponiamo, per assurdo, che esista n con le proprietà enunciate nella slide 4.

Consideriamo la stringa $z = a^n b^n c^n$, che è in L e ha la proprietà $|z| = 3n \geq n$.

Il linguaggio $a^k b^k c^k$ non è libero

Dimostriamo che $L = \{a^k b^k c^k \mid k \geq 0\}$ non è libero facendo vedere che per L il pumping lemma non vale.

Supponiamo, per assurdo, che esista n con le proprietà enunciate nella slide 4.

Consideriamo la stringa $z = a^n b^n c^n$, che è in L e ha la proprietà $|z| = 3n \geq n$.

Devono esistere u, v, w, x e y tali che $z = uvwxy$ e che soddisfano le condizioni 1–3 della slide 4.

Il linguaggio $a^k b^k c^k$ non è libero

Dimostriamo che $L = \{a^k b^k c^k \mid k \geq 0\}$ non è libero facendo vedere che per L il pumping lemma non vale.

Supponiamo, per assurdo, che esista n con le proprietà enunciate nella slide 4.

Consideriamo la stringa $z = a^n b^n c^n$, che è in L e ha la proprietà $|z| = 3n \geq n$.

Devono esistere u, v, w, x e y tali che $z = uvwxy$ e che soddisfano le condizioni 1–3 della slide 4.

Dalla condizione 1 sappiamo che $vx \neq \varepsilon$.

Il linguaggio $a^k b^k c^k$ non è libero

Dimostriamo che $L = \{a^k b^k c^k \mid k \geq 0\}$ non è libero facendo vedere che per L il pumping lemma non vale.

Supponiamo, per assurdo, che esista n con le proprietà enunciate nella slide 4.

Consideriamo la stringa $z = a^n b^n c^n$, che è in L e ha la proprietà $|z| = 3n \geq n$.

Devono esistere u, v, w, x e y tali che $z = uvwxy$ e che soddisfano le condizioni 1–3 della slide 4.

Dalla condizione 1 sappiamo che $vx \neq \varepsilon$.

Dalla condizione 2 sappiamo che vwx non può contenere sia a che c, in quanto in z la a più a destra è separata dalla c più a sinistra da n b. Dunque i casi (non esclusivi, ma che coprono tutte le possibilità) sono due: o vx non contiene a oppure vx non contiene c.

Il linguaggio $a^k b^k c^k$ non è libero

Dimostriamo che $L = \{a^k b^k c^k \mid k \geq 0\}$ non è libero facendo vedere che per L il pumping lemma non vale.

Supponiamo, per assurdo, che esista n con le proprietà enunciate nella slide 4.

Consideriamo la stringa $z = a^n b^n c^n$, che è in L e ha la proprietà $|z| = 3n \geq n$.

Devono esistere u, v, w, x e y tali che $z = uvwxy$ e che soddisfano le condizioni 1–3 della slide 4.

Dalla condizione 1 sappiamo che $vx \neq \varepsilon$.

Dalla condizione 2 sappiamo che vwx non può contenere sia a che c, in quanto in z la a più a destra è separata dalla c più a sinistra da n b. Dunque i casi (non esclusivi, ma che coprono tutte le possibilità) sono due: o vx non contiene a oppure vx non contiene c.

Dalla condizione 3 sappiamo che $uwy \in L$.

Il linguaggio $a^k b^k c^k$ non è libero

Dimostriamo che $L = \{a^k b^k c^k \mid k \geq 0\}$ non è libero facendo vedere che per L il pumping lemma non vale.

b

Supponiamo, per assurdo, che esista n con le proprietà enunciate nella slide 4.

Consideriamo la stringa $z = a^n b^n c^n$, che è in L e ha la proprietà $|z| = 3n \geq n$.

Devono esistere u, v, w, x e y tali che $z = uvwxy$ e che soddisfano le condizioni 1–3 della slide 4.

Dalla condizione 1 sappiamo che $vx \neq \varepsilon$.

e

Dalla condizione 2 sappiamo che vwx non può contenere sia a che c, in quanto in z la a più a destra è separata dalla c più a sinistra da n b. Dunque i casi (non esclusivi, ma che coprono tutte le possibilità) sono due: o vx non contiene a oppure vx non contiene c.

Dalla condizione 3 sappiamo che $uwy \in L$. (elimino v e x

c

Ora, se vx non contiene a, in uwy il numero di a è rimasto n, mentre il numero di b e/o c è diminuito. Se vx non contiene c, in uwy il numero di c è rimasto n, mentre il numero di a e/o di b è diminuito. In entrambi i casi abbiamo raggiunto una contraddizione.

Programma per dimostrare il pumping lemma

1. Argomentiamo che ogni grammatica libera G può essere trasformata in una forma – detta in **forma normale di Chomsky** – che è "quasi equivalente" a G e in cui le produzioni sono particolarmente semplici.

2. L'esistenza della forma normale di Chomsky di una grammatica è conseguenza di una serie di trasformazioni (non difficili, ma complessivamente tediose) della grammatica dettagliate nel libro di testo (Sezioni 7.1.1 – 7.1.4, lettura facoltativa con caffè).

3. Per ogni grammatica in forma normale di Chomsky, dimostriamo una relazione forte tra la profondità di un albero sintattico della grammatica e la lunghezza del suo prodotto.

4. Dimostriamo il pumping lemma per i linguaggi liberi.

Forma normale di Chomsky

Definizione

Diciamo che una grammatica è in **forma normale di Chomsky** (**CNF**, da **C**homsky **N**ormal **F**orm) se ogni sua produzione è della forma

- $A \to BC$ dove A, B e C sono variabili, oppure
- $A \to a$ dove A è una variabile e a un terminale.

Osservazione

È evidente che nessuna variabile (inclusa quella iniziale) è annullabile in una grammatica CNF. Infatti, ogni derivazione $A \Rightarrow \cdots$ aumenta o lascia invariata la lunghezza della stringa derivata da A, la quale è una stringa lunga 1.

Teorema

Se G è una grammatica che genera almeno una stringa non vuota, allora esiste una grammatica G' in forma normale di Chomsky tale che $L(G') = L(G) - \{\varepsilon\}$.

Dimostrazione (facoltativa)

Si veda la Sezione 7.1 del libro.

Alberi sintattici di grammatiche CNF

Teorema

Sia G una grammatica in forma normale di Chomsky e w il prodotto di un albero sintattico di G avente profondità $n \geq 1$. Allora $|w| \leq 2^{n-1}$.

Dimostrazione

Si procede per induzione sulla profondità n dell'albero, ricordando che ogni foglia dell'albero deve essere etichettata con un terminale.

(Caso base $n = 1$) Allora l'albero ha una radice A e un'unica foglia a che coincide con w. Concludiamo $1 = |w| \leq 2^{n-1} = 1$.

(Caso induttivo $n > 1$) Allora l'albero ha una radice A con esattamente due figli etichettati B e C alla radice di due sottoalberi la cui profondità è <u>non superiore</u> a $n - 1$.

Detti w_1 e w_2 i prodotti di questi due sottoalberi, abbiamo che $w = w_1 w_2$.

Usando l'ipotesi induttiva, deduciamo che $|w_1| \leq 2^{n-2}$ e $|w_2| \leq 2^{n-2}$.

Concludiamo $|w| = |w_1| + |w_2| \leq 2^{n-2} + 2^{n-2} = 2^{n-1}$.

Pumping lemma: dimostrazione (1/5)

Sia L un linguaggio libero.

Se $L = \emptyset$ oppure $L = \{\varepsilon\}$, allora è sufficiente prendere $n = 1$ e l'enunciato del pumping lemma vale banalmente, dal momento che non ci sono stringhe di L di lunghezza maggiore o uguale a 1.

Se L contiene almeno una stringa diversa da ε, sia $G = (V, T, P, S)$ una grammatica in forma normale di Chomsky che genera $L - \{\varepsilon\}$.

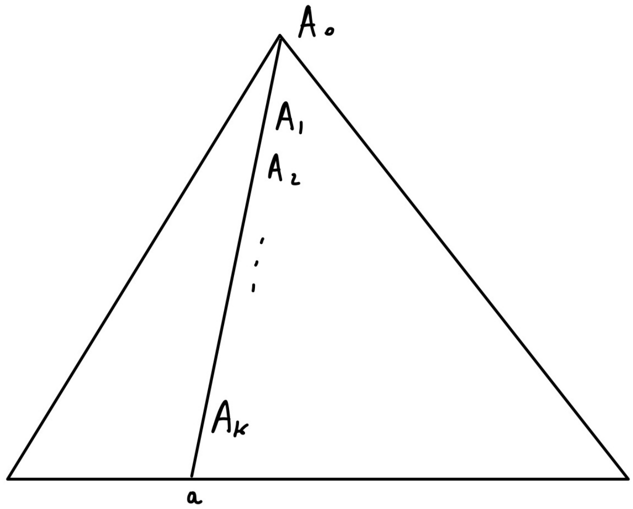
Poniamo $n = 2^m$ dove $m = |V|$. (1) (lenght = $2^{\text{num variabile}}$)

Prendiamo $z \in L$ tale che $|z| \geq n$. (2) (stringa base)

Per il teorema della slide precedente, ogni albero sintattico di G di profondità m ha come prodotto stringhe lunghe al massimo $2^{m-1} = n/2$. Deduciamo che ogni albero sintattico che ha come radice S e come prodotto z deve avere una profondità maggiore o uguale a $m + 1$, in quando z è lunga almeno il doppio di 2^{m-1}.

Pumping lemma: dimostrazione (2/5)

Deduciamo che questo albero sintattico avrà la forma



in cui è presente un cammino lungo $k \geq m + 1$ che tocca almeno $m + 2$ nodi, dei quali almeno $m + 1$ sono <u>nodi interni</u> e dunque etichettati con variabili A_i, mentre uno (l'ultimo) è una foglia etichettata con un terminale a.

In particolare, le almeno $m + 1$ variabili toccate dal cammino non possono essere tutte distinte, poiché la grammatica ne ha solo m.

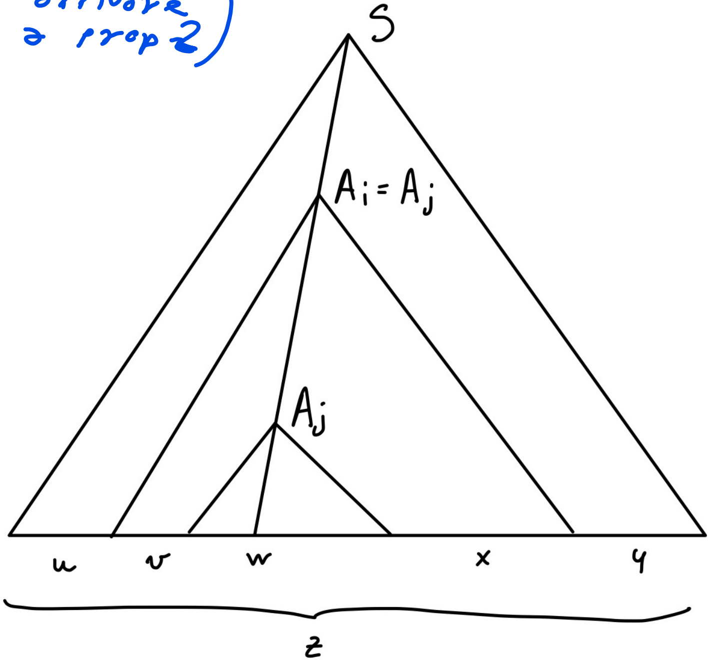
Pumping lemma: dimostrazione (3/5)

Deduciamo che almeno due delle ultime $m+1$ variabili del cammino (da A_{k-m} a A_k incluse) devono essere uguali.

(per arrivare a prop 2)

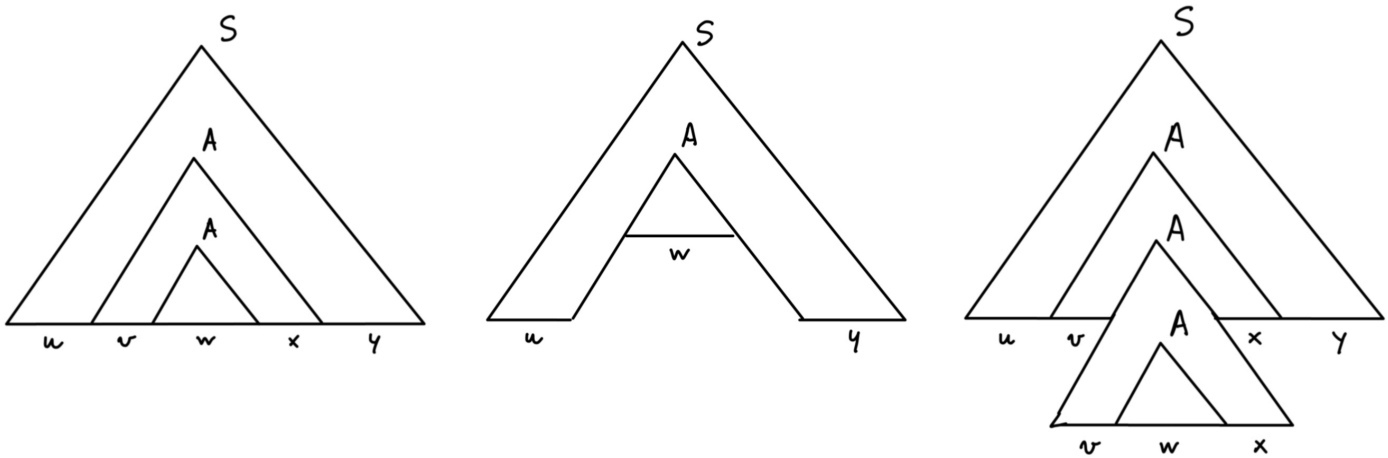
Supponiamo dunque $A_i = A_j = A$ con $k - m \leq i < j \leq k$. Il cammino fatto da $A_0 = S$ ad a può essere rappresentato come nella figura a destra, in cui il prodotto dell'albero è stato così scomposto:

- w è il prodotto del sottoalbero radicato in A_j;
- v e x sono le stringhe rispettivamente a sinistra e a destra di w nel prodotto del sottoalbero radicato in A_i;
- u e y sono le stringhe rispettivamente a sinistra e a destra del prodotto del sottoalbero radicato in A_i.



Evidentemente $z = uvwxy$.

Pumping lemma: dimostrazione (4/5)



Dall'albero sintattico iniziale (riprodotto a sinistra) è ora possibile costruirne altri:

- rimpiazzando il sottoalbero radicato in A_i con quello radicato in A_j si ottiene un albero sintattico con prodotto uwy;
- rimpiazzando il sottoalbero radicato in A_j con (una copia di) quello radicato in A_i si ottiene un albero sintattico con prodotto $uvvwxxyy = uv^2wx^2y$;
- in generale, iterando il rimpiazzamento del punto precedente, è possibile ottenere alberi sintattici con prodotto uv^kwx^ky per ogni $k \geq 1$.

Pumping lemma: dimostrazione (5/5)

Per concludere la dimostrazione osserviamo che:

- $vx \neq \varepsilon$

  Infatti, il cammino da A_i ad A_j deve contenere almeno una diramazione che produce almeno un simbolo in quanto la grammatica è in forma normale di Chomsky e non contiene produzioni ε;

- $|vwx| \leq n$

  Infatti il sottoalbero radicato in A_i ha profondità non superiore a $m + 1$ (vi è una sola variabile che si ripete nel suo cammino più lungo), dunque per il teorema in slide 8 il suo prodotto vwx ha una lunghezza non superiore a $2^m = n$.

  × teorema

Esercizi

Dimostrare che i seguenti linguaggi non sono liberi:

1. $\{ww \mid w \in \{0,1\}^*\}$
2. $\{0^i 1^j 2^i 3^j \mid i, j \geq 1\}$
3. $\{0^i 1^j 2^k \mid i < j < k\}$