Non contestualità

a.a. 2020-2021

Corso di Fondamenti di Informatica - 1 modulo

Corso di Laurea in Informatica

Università di Roma "Tor Vergata"

Prof. Giorgio Gambosi

Pumping lemma

Teorema

Sia $L \subseteq V_T^*$ *un linguaggio non contestuale. Esiste allora una costante* n *tale che se* $z \in L$ *e* $|z| \geq n$ *allora esistono 5 stringhe* $u, v, w, x, y \in V_T^*$ *tali che*

$$
\begin{array}{ll}
i) & uvwxy = z \\
ii) & |vx| \geq 1 \\
iii) & |vwx| \leq n \\
iv) & \forall i \geq 0 \; uv^iwx^iy \in L.
\end{array}
$$

Pumping lemma: interpretazione come gioco a due

Se L è context free, Alice vince sempre questo gioco con Bob:

1. Alice fissa un intero $n > 0$ opportuno
2. Bob sceglie una stringa $z \in L$ con $|z| > n$
3. Alice divide z in cinque parti $uvwxy$ con $|vwx| \leq n$ e $|vx| \geq 1$
4. Bob sceglie un intero $i \geq 0$
5. Alice mostra a Bob che $uv^iwx^iy \in L$

Dimostrazione

Grammatica $\mathcal{G} = \langle V_T, V_N, P, S \rangle$ in CNF che genera $L = L(\mathcal{G})$ e sia $k =\mid V_N \mid$ il numero di simboli non terminali in $\mathcal{G}$.

Qualunque albero sintattico $A(\sigma)$ relativo ad una stringa $\sigma \in V_T^*$ derivata in $\mathcal{G}$ sarà tale da avere tutti i nodi interni (corrispondenti a simboli non terminali) di grado 2, eccetto quelli aventi foglie dell'albero come figli, che hanno grado 1.

Dimostrazione

- Se h è l'altezza di A (numero massimo di archi, e anche numero massimo di nodi interni, in un cammino dalla radice ad una foglia), il massimo numero di foglie $l(h)$ è dato dal caso in cui l'albero è completo (i nodi interni hanno due figli, eccetto i padri di foglie, che ne hanno uno). Si può facilmente verificare che in tal caso abbiamo $l(h) = 2^{h-1}$, in quanto $l(1) = 1 = 2^0$ e $l(h+1) = 2 \cdot L(h) = 2 \cdot 2^{h-1} = 2^h$
- Se l'albero sintattico $A(\sigma)$ relativo alla stringa $\sigma \in L$ ha altezza $h(\sigma)$, la lunghezza di σ è allora $|\sigma| \leq 2^{h(\sigma)-1}$, e quindi $h(\sigma) \geq 1 + \log_2 |\sigma|$
- Se σ è una stringa sufficientemente lunga (in questo caso, $|\sigma| > 2^{|V_N|-1}$), ne risulta che $h(\sigma) \geq 1 + \log_2 |\sigma| > |V_N|$
- Quindi, se $|\sigma| > 2^{|V_N|-1}$ esiste almeno un cammino c dalla radice ad una foglia di $A(\sigma)$ che attraversa almeno $|V_N| + 1$ nodi interni

Dimostrazione

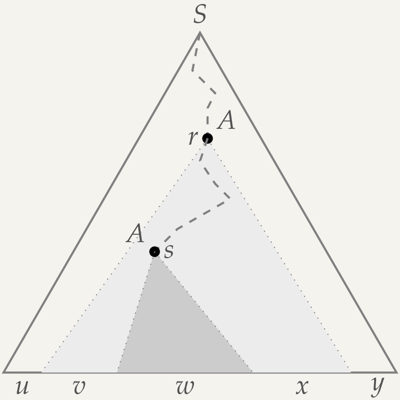
- I nodi interni di $A(\sigma)$ sono etichettati da simboli non terminali (le parti sinistre della produzioni nella derivazione di σ)
- Dato che i simboli non terminali sono $|V_N|$ mentre i nodi interni in c sono più di $|V_N|$, deve esistere (per il pigeonhole principle) un simbolo non terminale A che compare in due diversi nodi di c
- Di questi due nodi, indichiamo con r il nodo più vicino alla radice e con s il nodo associato ad A più vicino alla foglia
- Indichiamo con $r(\sigma)$ e $s(\sigma)$ le sottostringhe di σ corrispondenti alle foglie dei due sottoalberi $R(\sigma), S(\sigma)$ di $A(\sigma)$ aventi radice r e s
- Dato che s è un discendente di r, necessariamente $s(\sigma)$ è una sottostringa di $r(\sigma)$, per cui esistono due sottostringhe di v, x di σ tali che $r(\sigma) = v \cdot s(sigma) \cdot x$

Dimostrazione

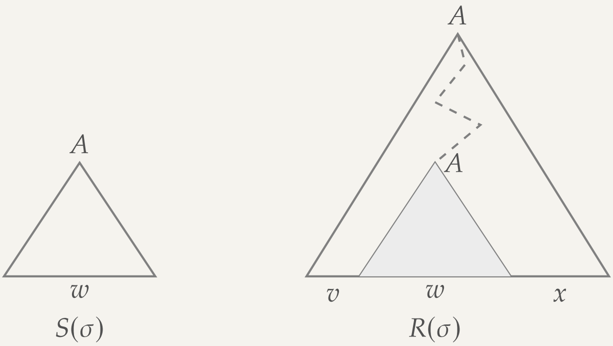
- La grammatica considerata è in CNF, per cui non sono presenti produzioni unitarie (a parte quelle relative alle foglie): di conseguenza, non può essere $s(\sigma) = r(\sigma)$, e quindi $|vx| > 1$
- Senza perdere generalità, possiamo assumere che $r(\sigma)$ sia il nodo in c più vicino alle foglie per il quale c'è un nodo sottostante $s(\sigma)$ associato allo stesso non terminale: quindi, il cammino più lungo da $r(\sigma)$ ad una foglia attraversa al più $|V_N| + 1$ nodi interni (esso stesso incluso) .
- Dalle osservazioni precedenti, ne deriva che $r(\sigma)$ ha lunghezza al più $2^{|V_N|+1-1} = 2^{|V_N|}$

Dimostrazione

Poniamo $s(\sigma) = w$ e quindi $r(\sigma) = vwx$.



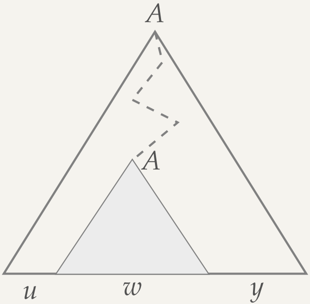
Dimostrazione



Dimostrazione

- Gli alberi $R(\sigma)$ e $S(\sigma)$ possono essere sostituiti (avendo radice corrispondente allo stesso non terminale) l'uno all'altro all'interno di un qualunque albero sintattico
- Quindi, anche la stringa uwy è generata dalla grammatica (sostituendo, in $A(\sigma)$, $S(\sigma)$ a $R(\sigma)$)
- Mediante la sostituzione opposta, anche la stringa $uvvwxxy$ risulta generabile.

Dimostrazione



Pumping lemma

La proprietà mostrata fornisce soltanto una condizione necessaria perché un linguaggio sia context free: non può essere utilizzata per mostrare la non contestualità di un linguaggio, ma solo per dimostrarne la contestualità.

$$L \text{ non contestuale} \implies \text{pumping lemma verificato}$$

$$\text{pumping lemma non verificato} \implies L \text{ non contestuale}$$

Pumping lemma: utilizzo come gioco a due

Se Alice vince sempre questo gioco con Bob, allora L non è CF

1. Bob sceglie un intero $n > 0$
2. Alice sceglie una stringa $z \in L$ con $|z| > n$
3. Bob divide z in cinque parti $uvwxy$ con $|vwx| \leq n$ e $|vx| \geq 1$
4. Alice sceglie un intero $i \geq 0$
5. Alice mostra a Bob che $uv^iwx^iy \notin L$

Esempio

$L = \{a^k b^k c^k | k > 0\}$ non è CF

1. Bob sceglie un intero $n > 0$
2. Alice sceglie la stringa $a^n b^n c^n \in L$
3. Bob divide z in cinque parti $uvwxy$ con $| vwx | \leq n$ e $| vx | \geq 1$. vwx o è una sequenza di occorrenze dello stesso simbolo (ad esempio a^h, $h > 0$) o è composta di due sottosequenze di stessi simboli (ad esempio $a^r b^s$, $r, s > 0$). Quindi, almeno uno dei simboli a, b, c non compare in vwx e quindi né in v né in x
4. Alice sceglie $i = 2$
5. Alice mostra a Bob che $uv^2wx^2y \notin L$ in quanto almeno un simbolo ha aumentato il numero di occorrenze ed almeno un altro simbolo ha un numero di occorrenze invariato