teorema PUMPING LEMMA

Per ogni linguaggio regolare L esiste una costante n tale che, se $z \in L$ e $|z| \geq n$, allora possiamo scrivere $z = uvw$ con $|uv| \leq n$, $|v| \geq 1$ e ottenere che $uv^iw \in L$ per ogni $i \geq 0$

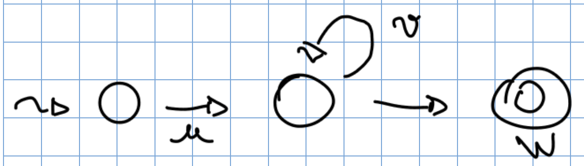
→ volte che si ripete v

Osservazione

Questo teorema ci da una condizione necessaria ma non sufficiente per i linguaggi regolari
Non posso utilizzarlo per dimostrare che un linguaggio L è regolare ma posso usarlo per dimostrare che un linguaggio L' non è regolare

Informalmente

Se L è regolare esiste un certo numero n che ha queste caratteristiche, se ho una stringa z che appartiene e $L \geq n$ la posso dividere in 3 parti, ognuno lunga "u" oppure "w" e la parte v la posso ripetere quante volte voglio ottenendo ancora stringhe $\in L$



$z = |\,u\,|\,v\,|\,w\,|$ $\qquad$ $z \in \mathcal{L}(A) = L$

Dimostrazione

Dato un linguaggio L regolare esiste ASFD che lo riconosca
sia $A = \langle \Sigma, Q, \delta, q_0, F \rangle$ un automa che riconosca L e sia $|Q| = n$

Per definizione

Sia $z \in L$ con $|z| = K \geq n \Rightarrow \bar{\delta}(q_0, z) \in F$
Supponiamo che $q_{i_0}, q_{i_1}, \dots, q_{i_K}$ sia la sequenza di stati attraversati da A durante la computazione z con $q_{i_0} = q_0$ e $q_{i_K} = \bar{\delta}(q_0, z) \in F$

↳ stringa

in altre parole q_{i_1} è lo stato in cui arriva l'automa A dopo aver letto il primo simbolo di z, q_{i_2} è lo stato in cui arriva A dopo aver letto il secondo simbolo di z

Indichiamo con z_h il prefisso di z di lunghezza h, allora avremo che q_{i_h} è lo stato in cui arriva l'automa A dopo aver letto z_h elementi cioè $q_{i_h} = \bar{\delta}(q_0, z_h)$

Dal momento che $k \geq n$, deve esistere almeno uno stato in cui l'automa si porta almeno 2 volte durante la computazione su z, cioè esistono 2 stati nella sequenza $q_{i_0}, q_{i_1}, \dots, q_{i_k}$ che coincidono. In realtà questi 2 stati che coincido si possono trovare già tra i primi $n+1$ elementi della sequenza $q_{i_0}, q_{i_1}, \dots q_{i_k}$. Possiamo affermare che esistono 2 indici r ed s con $0 \leq r < s \leq n$ tali che $q_{i_r} = q_{i_s}$ cioè lo stato in cui arriva l'automa leggendo il prefisso di z di lunghezza r (leggendo z_r) è esattamente lo stesso stato in cui arriva l'automa leggendo il prefisso di z di lunghezza s (leggendo z_s) cioè accade:

$$\bar{\delta}(q_0, z_r) = q_{i_r} = q_{i_s} = \bar{\delta}(q_0, z_s)$$

questa è ≥ 1 perché tra z_r e z_s ci sarà sicuramente un elemento dato che $r<s$

Per semplicità poniamo $u = z_r$ $\quad uv = z_s \quad uvw = z$

Chiaramente $|uv| = |z_s| = s \leq n$ e $|v| \geq 1$

perché $|u| = |z_r| = r < s = |z_s| = |uv|$

$$\cancel{|u|} < \cancel{|u|}|v| \to 0 < |v|$$

$$\underbrace{\overbrace{eeee}^{uv}}_{u}$$

$$|v| > 0 \implies |v| \geq 1$$

linguaggio

Inoltre dobbiamo dimostrare che $uv^iw \in L \quad \forall i \geq 0$

procediamo per induzione

$$\bar{\delta}(q_0, z_r) = q_{i_r} = q_{i_s} = \bar{\delta}(q_0, z_s)$$

Passo Base: $i = 0$

$$\bar{\delta}(q_0, uv^0w) = \bar{\delta}(q_0, u\varepsilon w) = \bar{\delta}(q_0, uw) = \bar{\delta}(\bar{\delta}(q_0, u), w)$$

$$= \bar\delta(\bar\delta(q_0, \overset{z_r}{\overset{\parallel}{z_s}}), w) = \bar\delta(\bar\delta(q_0, uv), w) = \bar\delta(q_0, uvw) =$$

$$\bar\delta(q_0, z) \in F$$

Cioè $uv^0w = uw \in L$

Passo induttivo $i > 0$

Per ipotesi induttiva $uv^{i-1} \in L$ cioè $\bar\delta(q_0, uv^{i-1}w) \in F$

$$\bar\delta(q_0, uv^iw) = \bar\delta(q_0, uvv^{i-1}w) = \bar\delta(\bar\delta(q_0, uv), v^{i-1}w)$$

$$= \bar\delta(\bar\delta(q_0, z_r), v^{i-1}w) = \bar\delta(\bar\delta(q_0, u), v^{i-1}w) =$$

$$= \underbrace{\bar\delta(q_0, uv^{i-1}w) \in F}_{\text{Per definizione del passo induttivo}} \Rightarrow \bar\delta(q_0, uv^iw) \in L \Rightarrow uv^iw \in L$$

linguaggi formali ⇒ grammatiche

Problema di andare a calcolare una funzione $f: N \to N$ può innescare un nuovo problema quello di calcolare e riconoscere il linguaggio
$L = \{ a^n b^{f(n)} \mid n \geq 0 \}$

- Possiamo usare le espressioni regolari per definire un linguaggio formale
- Un'altro approccio è quello generativo dove si usano strumenti formali che sono appunto le grammatiche formali che consentono di costruire le stringhe di un linguaggio tramite un insieme prefissato di regole, dette regole di produzione

APPROCCIO RICONOSCITIVO che consiste nell'utilizzare macchine astratte dette automi riconoscitori che definiscono algoritmi di riconoscimento dei linguaggi stessi, vale a dire algoritmi che per un dato linguaggio $L \subseteq \Sigma^*$ stabiliscono se una stringa $x \in \Sigma^*$ appartiene a L o meno

GRAMMATICHE DI CHOMSKY

La grammatica è un formalismo che permette di definire un'insieme di stringhe mediante l'imposizione di un particolare metodo per la loro costruzione. Deriva dallo studio di Noam Chomsky per la costruzione di frasi in lingua inglese

Definizione una grammatica formale G è una quartupla

$G = \langle V_T, V_N, P, S \rangle$

V_T: è l'insieme non vuoto di simboli dell'alfabeto terminale, i cui elementi sono detti caratteri terminali

V_N: insieme non vuoto di simboli dell'alfabeto non terminali, i cui elementi sono detti caratteri non terminali oppure variabili/categorie sintattiche

P: è un relazione binaria di cardinalità finita su:

$$(V_T \cup V_N)^* \circ V_N \circ (V_T \cup V_N)^* \times (V_T \cup V_N)^*$$

P è detta insieme delle regole di produzione o delle regole sintattiche

Una coppia $\langle \alpha, \beta \rangle \in P$, indichiamo con $\alpha \to \beta$ la relazione tra α e β

S: $S \in V_N$ è detto assioma ed è il simbolo non terminale di inizio ossia la categoria sintattica più generale

NOTAZIONE

$V = V_T \cup V_N \Rightarrow [V^* \circ V_N \circ V^*] \times V^*$

$a, b, c \in V_T$

$A, B, X, Y \in V_N$

$x, y, z, w \in V_T^*$

$\alpha, \beta, \gamma \in (V_T \cup V_N)^* = V^*$

* stella di Kleene

V_T^* è l'insieme di tutte le stringhe che possono essere formate usando i simboli presenti nell'alfabeto V_T

Se $V_T = \{a, b\}$ $V_T^* = \{\varepsilon, a, b, aa, \dots\}$

Osservazione

Le produzioni di una grammatica rappresentano le regole mediante le quali una stringa composta tutta di caratteri terminali può essere trasformata (riscritta) in un'altro modo

Osservazione

Il linguaggio generato dalla grammatica è l'insieme delle stringhe costituite solo dai caratteri terminali ai quali si può arrivare partendo dall'assioma S e applicando una sequenza, arbitrariamente lunga di passi di riscrittura

Una grammatica è data da un insieme di regole che applicate generano tutte le stringhe del linguaggio (metodo enumerativo)