Pumping lemma per linguaggi regolari

Nicola Gulmini

Teorema 1 (Pumping Lemma)**.** Sia $L \in Reg$[1] (non finito). Esiste un numero n naturale $n \in \mathbb{N}$ (nonnullo, $n \neq 0$) dipendente dal linguaggio stesso $n(L)$ detta *pumping length*, tale che: $\forall w \in L$ tale che $\|w\| \geq n$, essa può essere scomposta in $w = xyz$ che soddisfano tutte e tre queste proprietà:

1. $y \neq \varepsilon$, in altre parole $\|y\| > 0$;
2. $\|xy\| \leq n$;
3. $\forall k \geq 0$, $xy^k z \in L$.

Il pumping lemma serve per dimostrare che un linguaggio non è regolare. Infatti esso asserisce che se un linguaggio è regolare, esiste un n che soddisfa la proprietà enunciata (guarda i punti) per ogni sua stringa. Per dimostrare che un linguaggio non è regolare, pertanto, è sufficiente considerare una sola stringa w, fissando n, per cui **ogni possibile "taglio"** (in riferimento alla suddivisione in $w = xyz$ tre parti) la fa *uscire* dal linguaggio anche solo per un k. Come primo caso, a questo proposito, è utile considerare da subito $k = 0$, ossia rimuovere completamente il pezzetto y considerato e verificare che la stringa appartenga ancora al linguaggio. Appena si trova un k adatto ci si ferma e si conclude che non si ha regolarità. La cosa risulta di semplice comprensione se si pensa che la negazione del $\forall$ è $\exists$ e la negazione del $\exists$ è il $\forall$, guardando l'enunciato del lemma.

0.1 esercizio 1

Sia $L = \{0^n 1^n | n \geq 0\}$. Supponiamo che sia regolare e che la pumping length sia $n(L) = \bar{n}$.

Prendendo la stringa $w_1 = 0^{\frac{\bar{n}}{2}} 1^{\frac{\bar{n}}{2}} \in L$ la dimostrazione diventa lunga perché è necessario vagliare ogni possibile taglio y, il che comporta considerare il caso in cui $y = 0^{\|y\|}$, il caso in cui $y = 0^{\|y\|-m} 1^m$ per un opportuno m e, infine, il caso in cui $y = 1^{\|y\|}$. Inoltre si rischia di cadere nel seguente errore:

[1] Scrivere $L \in Reg$ significa che il linguaggio L è regolare, appartiene alla classe dei linguaggi regolari.

Esempio 1 (Esempio SBAGLIATO!)**.** Sia $w = xyz$ tale che $y = 0^i 1^i$, così $x = 0^{\frac{\bar{n}}{2}-i}$ e $z = 1^{\frac{\bar{n}}{2}-i}$. Innanzitutto le condizioni sono rispettate poiché $\|xy\| = \frac{\bar{n}}{2} + i < \bar{n}$ e $y \neq \varepsilon$; poi si nota subito che $xy^k z \in L$ per ogni $k \geq 0$.

L'errore sta nella parte finale. Sia, per comodità, $i = 2$:

$$xy^k z = 0 \cdots 0 \cdot 001100110011 \cdots 111 \cdots 1 \notin L.$$

La dimostrazione si semplifica se si prende la stringa $w_2 = 0^{\bar{n}} 1^{\bar{n}} \in L$ ancora appartenente al linguaggio, il pumping lemma deve continuare a valere poiché deve valere $\forall w \in L$. In questo caso non si può più selezionare y esattamente a cavallo tra gli zeri e gli uni, poiché altrimenti la condizione 2. cadrebbe, dunque l'unico caso ammissibile è $xy = 00 \ldots 0$ e $\forall k \geq 0$ si ha $xy^k z \notin L$ poiché la stringa si "sbilancia".

0.2 esercizio 2

- $L_< = \{a^n b^m | n < m\}$. Sia N la pumping length, scelgo $w = a^N b^{N+1}$, in questo modo $y = a^{\|y\|}$. Per $k = 0$ $xz \in L$, e anche con $k = 1$ la stringa rimane nel linguaggio. É sufficiente prendere un qualsiasi $k > 1$ affinché la stringa esca, dato che si hanno più a che b: $L_<$ non è regolare.

- $L_> = \{a^n b^m | n > m\}$. Analogamente a prima, si prende $w = a^{N+1} b^N$. In un solo passaggio, si nota subito che per qualsiasi y, con $k = 0$ la stringa esce dal linguaggio, perciò $L_>$ è non regolare.

- $L_\geq = \{a^n b^m | n \geq m\}$ si dimostra non essere regolare prendendo $w = a^N b^N$ con N pumping length e mostrando che per qualsiasi scomposizione di w in xyz possibile, con $k = 0$ la stringa non appartiene più al linguaggio.

0.3 esercizio 3

Il linguaggio $L_{\neq} = \{a^n b^m | n \neq m, \quad n, m \geq 1\}$ è regolare? Proviamo a dimostrare che non lo è:

- adottando approcci simili a quelli usati nel precedente esercizio non si riesce a dimostrarlo, basti verificare che $a^N b^{N+1}$ con $N > 0$ pumping length non funziona.

- Un modo furbo per procedere può essere quello di considerare il linguaggio complementare $\bar{L}_{\neq} = \Sigma^* / L_{\neq}$: se $L_{\neq}$ è regolare, anche il suo complementare lo è. Dimostrando che il complementare non è regolare si conclude: il complementare di un linguaggio regolare non può essere non regolare, dunque $L_{\neq}$ risulta non regolare[2].

Si verifica che l'insieme Σ^* è partizionabile in:

- insieme $L_{\neq}$;
- insieme delle stringhe $a^n b^n$, chiamato ragionevolmente $L_{=}$;
- insieme delle stringhe formate da a e b in ordine arbitrario (non disposte in modo da avere prima solo a e poi solo b), per semplicità chiamato L_{ar}.

Il linguaggio $\bar{L}_{\neq}$ è formato proprio dall'unione degli ultimi due.

Si consideri dunque $L_f = \bar{L}_{\neq} \cap L(\mathbf{a}^*\mathbf{b}^*)$ regolare perché intersezione di $\bar{L}_{\neq}$ regolare per ipotesi e $L(\mathbf{a}^*\mathbf{b}^*)$ certamente regolare[3]. Si ha, però, che $L_f = L_{=}$, che non è regolare, il che permette di concludere.

Riassumendo lo schema di ragionamento che ha portato alla conclusione:

1. $L_{\neq} \in \mathit{Reg} \Rightarrow \bar{L}_{\neq} \in \mathit{Reg}$. Per il modus tollens, è sufficiente dimostrare che $\bar{L}_{\neq} \notin \mathit{Reg}$ per concludere che anche $L_{\neq} \notin \mathit{Reg}$;
2. per dimostrare che $\bar{L}_{\neq} \notin \mathit{Reg}$ si procede allo stesso modo. Se $\bar{L}_{\neq} \in \mathit{Reg}$ allora la sua intersezione con un linguaggio regolare fornisce un linguaggio regolare;
3. si interseca quindi $\bar{L}_{\neq}$ con $L(\mathbf{a}^*\mathbf{b}^*)$ sicuramente regolare e si ottiene $L_{=} \notin \mathit{Reg}$. Se l'intersezione tra un supposto regolare $\bar{L}_{\neq}$ e un sicuramente regolare $L(\mathbf{a}^*\mathbf{b}^*)$ dà un non regolare allora $\bar{L}_{\neq} \notin \mathit{Reg}$;
4. tornando al primo punto, si è ottenuto che $\bar{L}_{\neq} \notin \mathit{Reg}$, dunque per il modus tollens anche $L_{\neq} \notin \mathit{Reg}$.

La dimostrazione risulta delicata perché è un attimo cadere in errori logici, infatti bisogna ricordare che:

- l'intersezione di regolari è regolare, ma non è detto che l'intersezione di non regolari sia sempre non regolare: nessun teorema lo garantisce. Nella nostra dimostrazione, infatti, si dice solo che se l'intersezione di due regolari non può fornire un non regolare;

[2] Il complementare di un regolare è regolare. Se il complementare è non regolare allora il linguaggio di partenza è sicuramente non regolare.

[3] $\mathbf{a}^*\mathbf{b}^*$ è un'espressione regolare. Il linguaggio $L(\mathbf{a}^*\mathbf{b}^*)$ è il linguaggio generato da tale espressione.

- l'unione di non regolari può non essere non regolare, e via così per ogni proprietà di chiusura...

Per ciascuno di questi possibili punti è sufficiente trovare un esempio che lo provi. Particolarmente interessante risulta il caso in cui la concatenazione di regolare e non regolare dà un regolare: siano $\tilde{L} \notin Reg$ e $\varnothing \in Reg$. Per la proprietà di annichilazione si ha che $\varnothing \cdot \tilde{L} = \varnothing \in Reg$.

- Usando il pumping lemma, invece, conviene procedere nel seguente modo, prestando attenzione ad alcuni passaggi. Sia n la pumping length e
$$w = a^n b^{n+n!} \in L_{\neq}.$$
$w = xyz$ con $x, y, z \in \{a, b\}^*$ e $y = a^{\|y\|}$ dato che $\|xy\| \leq n$.
Sia $\|y\| = m \geq 1$ e $w_k = xy^k z$, così
$$\#_a(w_k) = (n - m) + km = n - m + km = n + (k - 1)m.$$
É sufficiente trovare un k per cui $w_k \notin L_{\neq}$ affinché la dimostrazione della non regolarità si conclusa, sia allora
$$k = \frac{n!}{m} + 1 = k',$$
scelta che giustifica la presenza del fattoriale dato che $n \geq m$,
$$n! = n(n - 1) \cdots m(m - 1) \cdots 1$$
quindi $n!/m \in \mathbb{N}$.
Si ha
$$\#_a(w_{k'}) = n + \left(\frac{n!}{m} + 1 - 1\right)m = n + n! = \#_b(w_{k'}),$$
quindi $w_{k'} \notin L_{\geq}$. Riassumendo: fissando n e la stringa data, per ogni y ammissibile esiste k' tale che $xy^{k'}z$ non è nel linguaggio, quindi non vale il pumping lemma e $L_{\geq} \notin Reg$ classe dei linguaggi regolari.

0.4 esercizio 4

Dimostrare che $L = \{a^m | m \text{ è un numero primo}\}$ non è regolare.

Sia n la pumping length e sia $w = a^n$. Per ogni scomposizione $w = xyz$ che rispetta le ipotesi del P.L. si ha che xy^kz deve appartenere al linguaggio, dunque che $\|xy^kz\|$ sia primo per ogni k. Esiste, tuttavia, almeno un k t.c. la $\|xy^kz\|$ non è primo: $k = n + 1$. Si ha infatti
$$\|xy^kz\| = \|xyz\| + (k - 1) = n + k - 1 = n + n = 2n$$
che non è primo perché pari.