Azzolini Riccardo 2020-11-23

Altri esempi di linguaggi non regolari

1 Linguaggio dei numeri primi

Si consideri il linguaggio

$$L_{pr} = \{w \in \{1\}^* \mid |w| \text{ è un numero primo}\} = \{1^p \mid p \text{ è un numero primo}\}$$

cioè il linguaggio dei numeri primi in rappresentazione unaria. Esso non è regolare; per dimostrarlo, si applicano il pumping lemma e alcune proprietà dei numeri primi.

Innanzitutto, si suppone per assurdo che L_{pr} sia regolare, e che N sia la sua costante di pumping. Siccome i primi sono infiniti, si può sicuramente scegliere un numero primo $p \geq N + 2$, che corrisponde alla stringa $w = 1^p \in L_{pr}$. Per definizione, $|w| = p > N$, dunque secondo il pumping lemma deve esistere una scomposizione $w = xyz$ che verifichi le tre proprietà del lemma:

1. $y \neq \epsilon$;
2. $|xy| \leq N$;
3. $\forall k \geq 0 \quad xy^kz \in L_{pr}$.

Assumendo le proprietà 1 e 2, si dimostra che la 3 non può essere verificata.

Prima di procedere con la dimostrazione, bisogna studiare alcuni fatti relativi alla lunghezza delle stringhe in cui è scomposta w. La lunghezza di y deve essere $|y| \geq 1$ per la 1, e $|y| \leq |xy| \leq N$ per la 2, quindi complessivamente si ha $|y| = m$ con $1 \leq m \leq N$. Invece, la lunghezza delle altre parti della stringa è ottenuta semplicemente sottraendo la lunghezza di y da quella totale:

$$|xz| = |xyz| - |y| = |w| - |y| = p - m$$

Sapendo che $p \geq N + 2$ (per definizione) e $N \geq m$, si ricava che

$$p \geq N + 2 \geq m + 2$$

ovvero che

$$p - m \geq 2$$

Tornando al pumping lemma, secondo la proprietà 3 con $k = p - m$ la stringa $xy^{p-m}z$ dovrebbe appartenere a L_{pr}. Tuttavia, la sua lunghezza è

$$\begin{aligned} |xy^{p-m}z| &= |xz| + |y^{p-m}| \\ &= (p-m) + |y|(p-m) \\ &= (p-m) + m(p-m) \\ &= (p-m)(m+1) \end{aligned}$$

cioè un numero avente due fattori. Per definizione, questo potrebbe essere un numero primo solo se uno dei due fattori fosse uguale a 1, ma invece essi sono entrambi maggiori di 1, in base ad alcuni dei fatti visti prima:

$$\begin{gathered} p - m \geq 2 > 1 \\ m \geq 1 \implies m + 1 \geq 1 + 1 > 1 \end{gathered}$$

Segue che $xy^{p-m}z \notin L_{pr}$, contrariamente a quanto affermato dal pumping lemma. Essendo giunti a un assurdo, si deduce che l'ipotesi che L_{pr} fosse regolare è scorretta, dimostrando così che esso non è un linguaggio regolare.

Questo linguaggio è interessante perché corrisponde al problema della *primalità* (determinare se un numero dato sia primo o composto), che è molto importante nell'informatica. Il fatto che si tratti di un linguaggio non regolare indica che tale problema non è decidibile da un automa a stati finiti (qui lo si è dimostrato per una rappresentazione unaria dei numeri, ma lo stesso vale per qualunque altra rappresentazione).

2 Linguaggio dei palindromi

Un palindromo è una stringa che risulta uguale leggendola da sinistra a destra e da destra a sinistra, cioè, formalmente, una stringa w tale che $w = w^R$. Il linguaggio dei palindromi sull'alfabeto $\{0, 1\}$ è allora

$$L_{pal} = \{w \in \{0,1\}^* \mid w = w^R\} = \{\epsilon, 0, 1, 00, 11, 00100, 01110, \dots\}$$

e con una semplice applicazione del pumping lemma si dimostra che esso non è regolare.

Come al solito, si inizia supponendo per assurdo che L_{pal} sia regolare, e che N sia la sua costante di pumping. Si considera poi la stringa

$$w = 0^N 1 0^N = \underbrace{0 \dots 0}_{N} 1 \underbrace{0 \dots 0}_{N} \in L_{pal}$$

Poiché $|w| = 2N + 1 > N$, dovrebbe esistere una scomposizione $w = xyz$ che verifichi le tre proprietà del pumping lemma, ma assumendo la 1 e la 2 si dimostra che non può essere verificata la 3.

Dalla proprietà 2 ($|xy| \leq N$) si deduce che la stringa y “rientra” interamente nella sequenza di zeri prima dell'1 centrale, quindi deve essere composta interamente di zeri, cioè avere la forma $y = 0^h$, e per la 1 ($y \neq \epsilon$) si deve inoltre avere $h \geq 1$. Adesso, si costruisce la stringa $xy^k z$ scegliendo, ad esempio, $k = 2$:

$$xy^2z = xyyz = \underbrace{0\ldots\overbrace{0\ldots0}^{yy}\ldots0}_{N+h}\,1\,\underbrace{0\ldots0}_{N}$$

A causa della ripetizione di y, a sinistra dell'1 ci sono $h \geq 1$ zeri in più di quelli a destra, dunque la stringa risultante non è un palindromo, $xy^2z \notin L_{pal}$; ciò contraddice la proprietà 3 del pumping lemma, dimostrando che L_{pal} non è appunto regolare.