Azzolini Riccardo 2020-11-20

Chiusura dei linguaggi regolari rispetto all'inversione di stringhe

1 Reverse (inverso) di una stringa e di un linguaggio

Si indica con w^R la stringa **reverse** (o **inversa**) di w, cioè la stringa w scritta "al contrario": se $w = a_1 \dots a_n$, allora $w^R = a_n \dots a_1$. Formalmente, il reverse w^R di $w \in \Sigma^*$ è definito induttivamente nel modo seguente:

$$w^R = \begin{cases} \epsilon & \text{se } w = \epsilon \\ ax^R & \text{se } w = xa, \text{ con } x \in \Sigma^* \text{ e } a \in \Sigma \end{cases}$$

Il **reverse** di un linguaggio L, indicato con L^R, è il linguaggio

$$L^R = \{w^R \mid w \in L\}$$

1.1 Inversione e concatenazione di stringhe

Utilizzando questa definizione induttiva di reverse di una stringa si dimostra che, se la stringa w è composta da due stringhe $\alpha, \beta \in \Sigma^*$, cioè $w = \alpha\beta$, allora

$$w^R = \beta^R \alpha^R$$

La dimostrazione è per induzione su $|\beta|$.

- *Caso base*: $|\beta| = 0$, cioè $\beta = \epsilon$.

$$\begin{aligned} (\alpha\beta)^R &= (\alpha\epsilon)^R && [\beta = \epsilon] \\ &= \alpha^R && [\text{concatenazione con la stringa vuota}] \\ &= \epsilon\alpha^R && [\text{concatenazione con la stringa vuota}] \\ &= \epsilon^R\alpha^R && [\text{caso base della definizione di reverse: } \epsilon^R = \epsilon] \\ &= \beta^R\alpha^R && [\beta = \epsilon] \end{aligned}$$

- *Caso induttivo*: $|\beta| > 0$, cioè $\beta = \gamma c$, con $\gamma \in \Sigma^*$ e $c \in \Sigma$. L'ipotesi induttiva è $(\alpha\gamma)^R = \gamma^R\alpha^R$.

$$
\begin{aligned}
(\alpha\beta)^R &= (\alpha\gamma c)^R && [\beta = \gamma c] \\
&= c(\alpha\gamma)^R && [\text{caso induttivo della definizione di reverse: } (xa)^R = ax^R] \\
&= c\gamma^R\alpha^R && [\text{ipotesi induttiva}] \\
&= (\gamma c)^R\alpha^R && [\text{caso induttivo della definizione di reverse}] \\
&= \beta^R\alpha^R && [\beta = \gamma c]
\end{aligned}
$$

1.2 Proprietà dell'inversione dei linguaggi

Siano L e M due linguaggi qualunque (non necessariamente regolari). Valgono le seguenti proprietà:

1. $(L \cdot M)^R = M^R \cdot L^R$
2. $\forall i \geq 0,\ (L^i)^R = (L^R)^i$
3. $(L^*)^R = (L^R)^*$

Le dimostrazioni sono relativamente semplici, e non verranno presentate; sostanzialmente, la 1 segue da $(\alpha\beta)^R = \beta^R\alpha^R$, la 2 segue dalla 1 e la 3 segue dalla 2.

2 Reverse di un'espressione regolare

Data un'espressione regolare E, l'espressione **reverse** di E, indicata con E^R, è definita induttivamente sulla struttura di E:

$$
E^R = \begin{cases}
E & \text{se } E = \varnothing,\ E = \epsilon \text{ o } E = a \in \Sigma \\
S^R + T^R & \text{se } E = S + T \\
T^R \cdot S^R & \text{se } E = S \cdot T \\
(S^R)^* & \text{se } E = S^* \\
(S^R) & \text{se } E = (S)
\end{cases}
$$

Teorema: Data un'espressione regolare E, il linguaggio generato dal reverse di E è uguale al reverse del linguaggio generato da E, cioè $L(E^R) = L(E)^R$.

La dimostrazione è per induzione sulla struttura di E.

- *Casi base*: $E = \varnothing$, $E = \epsilon$ o $E = a \in \Sigma$. In questi casi, per definizione $E^R = E$, e si può facilmente osservare che vale anche $L(E)^R = L(E)$, quindi $L(E)^R = L(E^R) = L(E)$.

- *Caso induttivo* $E = S \cdot T$: si assume, per ipotesi induttiva, che S^R e T^R siano le espressioni regolari tali che $L(S)^R = L(S^R)$ e $L(T)^R = L(T^R)$. Per definizione, $E^R = T^R \cdot S^R$, e si ha che:

$$
\begin{aligned}
L(S \cdot T)^R &= (L(S) \cdot L(T))^R && \text{[semantica di } \cdot \text{]} \\
&= L(T)^R \cdot L(S)^R && \text{[proprietà 1 dell'inversione]} \\
&= L(T^R) \cdot L(S)^R && \text{[ipotesi induttiva]} \\
&= L(T^R \cdot S^R) && \text{[semantica di } \cdot \text{]}
\end{aligned}
$$

- Gli altri casi induttivi si dimostrano in modo analogo.

3 Chiusura rispetto all'inversione

Teorema: Se L è un linguaggio regolare, allora anche il linguaggio L^R è regolare.

Dimostrazione: Dato che L è regolare, esiste un'espressione regolare E che lo genera, ovvero tale che $L = L(E)$. Considerando adesso l'espressione regolare E^R, per il teorema precedente si ha che $L(E)^R = L(E^R)$. Mettendo insieme queste due proprietà, si deduce che $L^R = L(E)^R = L(E^R)$: esiste un'espressione regolare (E^R) che genera L^R, dunque quest'ultimo è un linguaggio regolare.

Come verrà mostrato in seguito, questo stesso risultato può essere dimostrato utilizzando gli automi invece delle espressioni regolari.

4 Inversione tramite automi

La chiusura dei linguaggi regolari rispetto all'inversione può anche essere dimostrata utilizzando gli automi invece delle espressioni regolari.

Dato un ϵ-NFA $A = \langle Q, \Sigma, \delta, q_0, F \rangle$, l'idea è di costruire un altro ϵ-NFA, l'*automa reverse* A_R, in cui vengono sostanzialmente invertiti tutti i percorsi accettanti dell'automa A: se in A esiste un percorso che accetta la stringa $w = a_1 \dots a_n$, passando dallo stato iniziale q_0 a uno stato finale f,

$$q_0 \xrightarrow{a_1} \cdots \xrightarrow{a_n} f$$

allora in A_R deve esistere un percorso da f a q_0 che accetta la stringa $w^R = a_n \dots a_1$,

$$f \xrightarrow{a_n} \cdots \xrightarrow{a_1} q_0$$

Perché questo percorso sia accettante, in A_R lo stato q_0 deve essere finale, e f dovrebbe intuitivamente essere iniziale. C'è però un problema: in generale, A può avere più stati finali, ma A_R — come ogni automa stati finiti — deve avere un solo stato iniziale. La soluzione è introdurre un nuovo stato iniziale p_0 con delle ϵ-transizioni che portano da esso a tutti gli stati che erano finali in A, tra cui lo stato f del percorso considerato prima.

Formalmente,

$$A_R = \langle Q \cup \{p_0\}, \Sigma, \delta_R, p_0, \{q_0\}\rangle$$

, dove $p_0 \notin Q$ (p_0 è appunto uno stato nuovo), e la funzione di transizione $\delta_R : (Q \cup \{p_0\}) \times \Sigma_\epsilon \to 2^{Q\cup\{p_0\}}$ è definita come segue:

$$\delta_R(p_0, \epsilon) = F$$
$$\forall q \in Q,\ \forall a \in \Sigma \quad \delta_R(q, a) = \{p \in Q \mid q \in \delta(p, a)\}$$

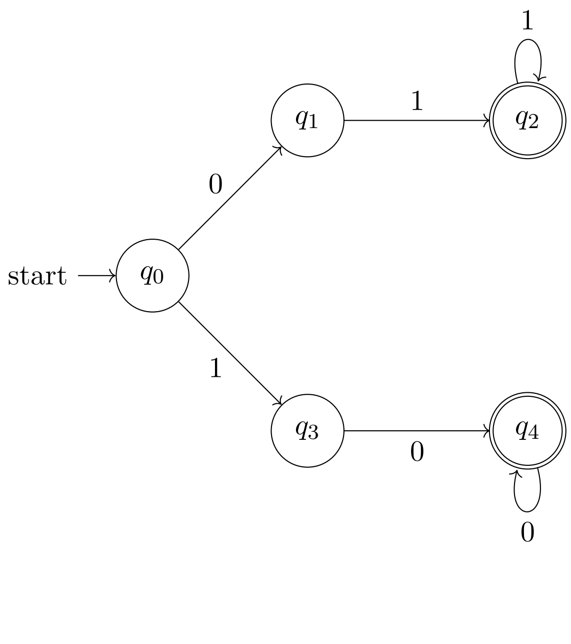
La prima parte di questa definizione, $\delta_R(p_0, \epsilon) = F$, specifica le ϵ-transizioni che portano dal nuovo stato iniziale p_0 a quelli che erano gli stati finali di A. Invece, la seconda parte della definizione inverte la direzione di tutte le transizioni di δ: se in A esistono ad esempio le transizioni $p_1 \xrightarrow{a} q$ e $p_2 \xrightarrow{a} q$, cioè $q \in \delta(p_1, a)$ e $q \in \delta(p_2, a)$, in A_R si hanno invece $q \xrightarrow{a} p_1$ e $q \xrightarrow{a} p_2$, ovvero $\delta_R(q, a) = \{p_1, p_2\}$.

Teorema: Dato un ϵ-NFA A, si ha che $L(A)^R = L(A_R)$.

La dimostrazione, che qui non verrà mostrata, può essere svolta ragionando induttivamente sulle stringhe riconosciute da A.

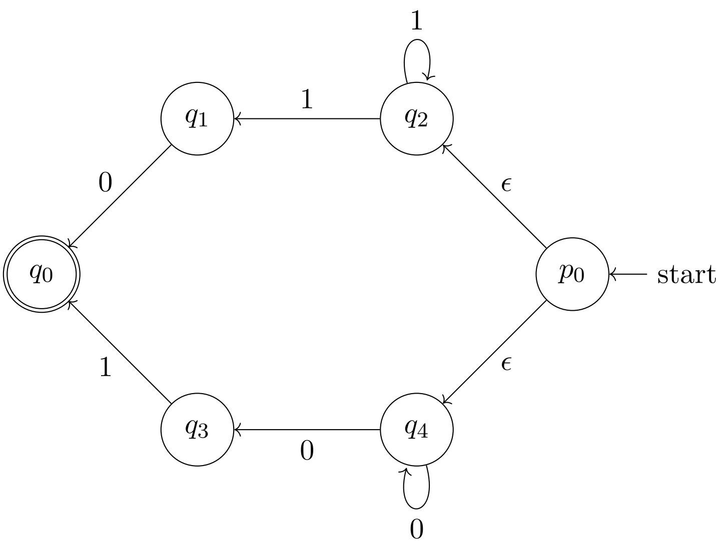
4.1 Esempio

Si consideri il seguente automa A:



Esso riconosce il linguaggio generato dall'espressione regolare $E = 011^* + 100^*$, cioè $L(A) = L(011^* + 100^*)$.

In base alla costruzione appena presentata, il reverse di A è il seguente automa A_R:



Per il teorema, si ha che $L(A_R) = L(A)^R$: il linguaggio riconosciuto da A_R è il reverse di quello riconosciuto da A. In termini di espressioni regolari, si può osservare intuitivamente che $L(A_R) = L(1^*10 + 0^*01)$, e quest'espressione è proprio il reverse dell'espressione E corrispondente al linguaggio di A:

$$\begin{aligned}(011^* + 100^*)^R &= (011^*)^R + (100^*)^R \\ &= (1^*)^R(01)^R + (0^*)^R(10)^R \\ &= (1^R)^*1^R0^R + (0^R)^*0^R1^R \\ &= 1^*10 + 0^*01\end{aligned}$$

4.2 Dimostrazione della chiusura rispetto all'inversione

Come anticipato, utilizzando questo risultato sull'automa reverse si può dare una dimostrazione alternativa del teorema di chiusura dei linguaggi regolari rispetto all'inversione, ricordando che l'enunciato del teorema è:

Teorema: Se L è un linguaggio regolare, allora anche il linguaggio L^R è regolare.

Dimostrazione: Dato che L è regolare, esiste un ϵ-NFA A tale che $L(A) = L$. Per il teorema precedente, l'automa reverse A_R è tale che $L(A_R) = L(A)^R = L^R$, quindi anche L^R è riconosciuto da un ϵ-NFA, e ciò dimostra che è un linguaggio regolare.