

Come calcolare FOLLOW

Luca Padovani
padovani@sti.uniurb.it

1 Introduzione

Abbiamo visto che per la costruzione delle tabelle di parsing LL(1) e SLR a partire da una grammatica fissata G abbiamo bisogno di due funzioni ausiliarie FIRST e FOLLOW. FIRST determina l'insieme di simboli terminali che possono *cominciare* una stringa di simboli, FOLLOW determina l'insieme di simboli terminali che possono *seguire* un simbolo non terminale in una derivazione in G .

Consideriamo la grammatica

$$\begin{aligned} E &\rightarrow TE' \\ E' &\rightarrow +TE' \mid \varepsilon \\ T &\rightarrow FT' \\ T' &\rightarrow \times FT' \mid \varepsilon \\ F &\rightarrow a \mid (E) \end{aligned}$$

Il calcolo di FIRST porta alla seguente tabella:

$$\begin{aligned} \text{FIRST}(E) &= \{a, (\} \\ \text{FIRST}(E') &= \{+, \varepsilon\} \\ \text{FIRST}(T) &= \{a, (\} \\ \text{FIRST}(T') &= \{\times, \varepsilon\} \\ \text{FIRST}(F) &= \{a, (\} \end{aligned}$$

2 Definizione di FOLLOW

Secondo la definizione il FOLLOW di un simbolo non terminale si calcola così:

1. se S è il simbolo iniziale della grammatica, allora $\$$ sta in $\text{FOLLOW}(S)$;
2. se $A \rightarrow \alpha B \beta$ è una produzione della grammatica, tutti i simboli in $\text{FIRST}(\beta) - \{\varepsilon\}$ stanno in $\text{FOLLOW}(B)$;
3. se $A \rightarrow \alpha B$ è una produzione della grammatica, oppure se $A \rightarrow \alpha B \beta$ è una produzione della grammatica e $\varepsilon \in \text{FIRST}(\beta)$, tutti i simboli in $\text{FOLLOW}(A)$ stanno in $\text{FOLLOW}(B)$.

3 Procedimento

Notiamo che la definizione non consente di determinare il FOLLOW di un simbolo non terminale “in un colpo solo”, in quanto lo stesso simbolo non terminale può comparire in diverse produzioni della grammatica e dunque i casi 2 e 3 potrebbero dover essere applicati più volte. Notiamo infatti che la definizione precisa ad esempio

tutti i simboli in $\text{FIRST}(\beta) - \{\varepsilon\}$
stanno in $\text{FOLLOW}(B)$

che è ben diverso da dire

$$\text{FOLLOW}(B) = \text{FIRST}(\beta) - \{\varepsilon\}$$

Dunque, il calcolo di $\text{FOLLOW}(B)$ (dove B è un simbolo non terminale arbitrario) procede in due fasi. Nella prima fase si fa una passata su tutte le produzioni della grammatica in cui B compare *nel corpo della produzione* e si annotano a parte i simboli che scopriamo appartenere al $\text{FOLLOW}(B)$. Per ogni occorrenza di B incontrata durante la passata ci sono due possibilità:

1. se la produzione ha la forma $A \rightarrow \alpha B \beta$ dove β non è la stringa vuota, allora scriviamo l'annotazione

$$\text{FIRST}(\beta) - \{\varepsilon\} \subseteq \text{FOLLOW}(B)$$

(questo deriva dalla regola 2). Se, in aggiunta, abbiamo che $\varepsilon \in \text{FIRST}(\beta)$, allora scriviamo *anche* l'annotazione

$$\text{FOLLOW}(A) \subseteq \text{FOLLOW}(B)$$

(questo deriva dalla regola 3)

2. se la produzione ha la forma $A \rightarrow \alpha B$, allora scriviamo l'annotazione

$$\text{FOLLOW}(A) \subseteq \text{FOLLOW}(B)$$

(questo deriva dalla regola 3)

Notare che lo stesso simbolo non terminale può comparire *più volte* nel corpo della stessa produzione. In questo caso le annotazioni di cui sopra vanno prese *per ogni occorrenza*.

Prendendo ad esempio la grammatica G descritta prima, le annotazioni ottenute dopo questa passata sono le seguenti:

$$\begin{aligned}
\$ &\in \text{FOLLOW}(E) \\
+ &\in \text{FOLLOW}(T) \\
\text{FOLLOW}(E) &\subseteq \text{FOLLOW}(T) \\
\text{FOLLOW}(E) &\subseteq \text{FOLLOW}(E') \\
\text{FOLLOW}(E') &\subseteq \text{FOLLOW}(T) \\
\times &\in \text{FOLLOW}(F) \\
\text{FOLLOW}(T) &\subseteq \text{FOLLOW}(F) \\
\text{FOLLOW}(T) &\subseteq \text{FOLLOW}(T') \\
\text{FOLLOW}(T') &\subseteq \text{FOLLOW}(F) \\
) &\in \text{FOLLOW}(E)
\end{aligned}$$

Abbiamo omesso le relazioni ovvie (come $\text{FOLLOW}(E') \subseteq \text{FOLLOW}(E')$) e abbiamo scritto $+ \in \text{FOLLOW}(T)$ anziché $\{+\} \subseteq \text{FOLLOW}(T)$.

Una volta completata la scrittura delle annotazioni, si procede con la seconda fase in cui riempiamo i FOLLOW dei simboli non terminali seguendo le catene di inclusione. Per esempio, sappiamo che $+ \in \text{FOLLOW}(T)$. Quindi mettiamo $+$ nel $\text{FOLLOW}(T)$. Poi osserviamo che $\text{FOLLOW}(T) \subseteq \text{FOLLOW}(F)$ e quindi mettiamo $+$ anche nel $\text{FOLLOW}(F)$. Iterando il procedimento per tutti i simboli terminali si cui abbiamo scoperto l'appartenenza a qualche FOLLOW otteniamo infine la seguente tabella, che definisce il FOLLOW di tutti i simboli non terminali della grammatica:

| X | $\text{FOLLOW}(X)$ |
|------|------------------------|
| E | $\{\$,)\}$ |
| E' | $\{\$,)\}$ |
| T | $\{\$,), +\}$ |
| T' | $\{\$,), +\}$ |
| F | $\{\$,), +, \times\}$ |