Martin Pernecký, 2i3

Zadanie

Majme danú kontextovú gramatiku G. Zostrojte lineárne ohraničený automat (LBA) A, taký že L(G) = L(A). Pri simulácií použite spätné odvodenie (teda postupne na páske robíte spätné kroky až kým sa nedostanete ku σ).
Pri dôkaze použite matematickú indukciu.

Riešenie

Ku kontextovej gramatike
G = (N; T; σ; P)
vytvoríme automat
A = (K; T; N ∪ T ∪ {ε}; δ; q; {q_k})
$Q = \bigcup_{a_1a_2...a_n \to b_1b_2...b_m \in P} q_{[a_1a_2...a_n;b_1b_2...b_m]}, q'_{[a_1a_2...a_n;b_1b_2...b_m]}; q'_{[a_2...a_n;b_2...b_m]}; \ldots q'_{[\varepsilon;\varepsilon]}$
K = {q; q_k; q_r; q_{c0}; q_{c1}} ∪ Q

Definícia delta funkcie

1. Nedeterministicky sa rozhodne pre nejaké pravidlo
   $\delta(q; w) = \{ (q_{[a; b]}; w; 0) \mid a \to b \in P \}$ (∀w ∈ Γ)
2. Rozhodne sa pre nejakú pozíciu na páske
   $\delta(q_{[a; b]}; w) = \{ (q_{[a; b]}; w; 1) \}$ (∀w ∈ Γ) ($\forall q_{[a; b]} \in Q$)
3. Prepisovanie pravidla;
   $\delta(q_{[a_1a_2\ldots a_n; b_1b_2\ldots b_m]}; w) = \{ (q'_{[a_1a_2\ldots a_n; b_1b_2\ldots b_m]}; w; 0) \}$ (∀w ∈ Γ) ($\forall q_{[a_1a_2\ldots a_n; b_1b_2\ldots b_m]} \in Q$)[1]
   $\delta(q'_{[a_1a_2\ldots a_n; b_1b_2\ldots b_m]}; b_1) = \{ (q'_{[a_2\ldots a_n; b_2\ldots b_m]}; a_1; 1) \}$ ($\forall q'_{[a_1a_2\ldots a_n; b_1b_2\ldots b_m]} \in Q$)
   .
   .
   .
   $\delta(q'_{[a_n; b_n\ldots b_m]}; b_n) = \{ (q'_{[\varepsilon; b_{n+1}\ldots b_m]}; a_n; 1) \}$ ($\forall q'_{[a_1a_2\ldots a_n; b_1b_2\ldots b_m]} \in Q$)
   $\delta(q'_{[\varepsilon; b_{n+1}\ldots b_m]}; b_{n+1}) = \{ (q'_{[\varepsilon; b_{n+2}\ldots b_m]}; \varepsilon; 1) \}$ ($\forall q'_{[a_1a_2\ldots a_n; b_1b_2\ldots b_m]} \in Q$)
   .
   .
   .
   $\delta(q'_{[\varepsilon; b_m]}; b_m) = \{ (q'_{[\varepsilon; \varepsilon]}; \varepsilon; 1) \}$ ($\forall q'_{[a_1a_2\ldots a_n; b_1b_2\ldots b_m]} \in Q$)
4. Presun na začiatok pásky
   $\delta(q'_{[\varepsilon; \varepsilon]}; w) = \{ (q\mathrm{r}; w; -1) \}$ (∀w ∈ Γ − {¢}))
   $\delta(q_r; w) = \{ (q_r; w; -1) \}$ (∀w ∈ Γ − {¢}))
5. Kontrola, či na páske neostalo len σ
   $\delta(q_r; ¢) = \{ (q_{c0}; ¢; 0) \}$
   $\delta(q_{c0}; \sigma) = \{ (q_{c1}; \sigma; 1) \}$
   $\delta(q_{c0}; \varepsilon) = \{ (q_{c0}; \varepsilon; 1) \}$
   $\delta(q_{c1}; \varepsilon) = \{ (q_{c1}; \varepsilon; 1) \}$
   $\delta(q_{c1}; \$) = \{ (q_k; \$; 0) \}$
6. Návrat do počiatočného stavu
   $\delta(q_r; ¢) = \{ (q; ¢; 1) \}$

   $\delta(q'_{[a; b]}; \varepsilon) = \{ (q'_{[a; b]}; \varepsilon; 1) \}$ ($\forall q'_{[a; b]} \in Q$)

Popis funkcie automatu

Automat simuluje odvodenie v G odzadu, takže keď máme napríklad v G odvodenie slova σ ⇒ Aσ ⇒ AA ⇒ aA ⇒ aa, tak v automate A by overenie slova (zjednodušene) vyzeralo nasledovne: aa |– aA |– AA |– Aσ |– σ.
Prepisovanie slove prebieha v automate v týchto 6 etapách:

1. automat sa nedeterministicky rozhodne, ktoré pravidlo bolo použité ako posledné pri odvodení vetnej formy zapísanej na páske

[1] Prechod do „prepisovacieho“ módu, stavy $q_{[a; b]}$ určujú pravidlo, podľa ktorého sa prepisuje a stavy $q'_{[a; b]}$ sú už prepisovacie stavy – v tomto stave sa prepíše prvé písmeno pravej strany za prvé písmeno ľavej strany.

2. automat si ďalej vyberie počiatočnú pozíciu pravej strany pravidla
3. následne sa automat dostane do prespisovacieho stavu a začne postupne prepisovať pravú stranu pravidla ľavou stranou (simulácia odzadu). Ak automat zistí, že symboly na páske sa nezhodujú zo symbolmi na pravej strane pravidla, tak sa odvodzovanie zablokuje.
   Ak je ľavá strana pravidla kratšia, tak sa na pásku dopíšu symboly ε (napríklad pri použití pravidla A → AAA sa páska prepíše nasledovne: CCCAAAC |– CCCAεεC).
   Z definície kontextovej gramatiky vyplýva, že ľavá strana nemôže byť dlhšia ako pravá strana pravidla, teda sa netreba zaoberať vecami, ako je pretečenie pásky alebo prepísanie symbolov, ktoré sa nemajú prepísať. Jediné čo treba ošetriť je preskakovanie ε-ov pri prepisovaní pravidiel
4. Po prepísaní pravidla automat presunie hlavu na počiatočnú pozíciu a nedeterministicky sa rozhodne či skončil (to znamená či na páske ostali už len ε a jedna σ)
5. Ak áno tak overí, či naozaj je na páske $¢ε^*σε^*$$, ak áno tak prejde do akceptačnej konfigurácie, ak nie tak sa zablokuje
6. Ak nie tak prejde do počiatočnej konfigurácie a pokračuje s prepisovaním s novým pravidlom (krok 1)

Dôkaz

1. L(G) ⊆ L(A)

Dôkaz matematickou indukciou na počet krokov odvodenia w ∈ L(G)

1° k = 1

$\sigma \Rightarrow_G w \Rightarrow (\sigma \rightarrow w) \in P$

odvodenie v A vyzerá nasledovne: $q¢w\$ \vdash q_{[\sigma;\ w]}¢w\$ \vdash ¢q_{[\sigma;\ w]}w\$ \vdash ¢q'_{[\sigma;\ w]}w\$ \vdash^* ¢σε^*q'_{[ε;\ ε]}\$ \vdash ¢σε^*q_r\$ \vdash^* ¢q_rσε^*\$ \vdash ¢q_{c0}σε^*\$ \vdash ¢σq_{c1}ε^*\$ \vdash^* ¢σ\ ε^*q_{c2}\$ \vdash ¢σ\ ε^*q_k\$$

2° k= n+ 1

Nech w je vetná forma odvodená v G na n krokov, podľa indukčného predpokladu sa dá odvodiť v automate A. Nech pre w' platí: $\sigma \Rightarrow^* w \Rightarrow_{(a \rightarrow b)} w'$

odvodenie w' vyzerá nasledovne:

$q¢w'\$ \vdash q_{[a;\ b]}¢w'\$ \vdash ¢q_{[a;\ b]}w'\$ \vdash ¢q'_{[a;\ b]}w\$ \vdash^* ¢wq'_{[ε;\ ε]}\$ \vdash ¢wq_r\$ \vdash^* ¢q_{rW}\$ \vdash q¢w\$ \vdash^*_{(IP)} ¢q_{c0}σε^*\$ \vdash ¢σq_{c1}ε^*\$ \vdash^* ¢σ\ ε^*q_{c2}\$ \vdash ¢σ\ ε^*q_k\$$

$\Rightarrow L(G) \subseteq L(A)$

2. L(G) ⊇ L(A)

Nech w ∈ L(A), z toho vyplýva, že existuje postup odvodenia w v A. Keď rozdelíme tento postup podľa konfigurácií $¢q$w\$, tak každá táto časť predstavuje použitie nejakého pravidla z G

Dôkaz matematickou indukciou na počet konfigurácií typu $¢q$w\$ (w ∈ Γ^*)

1° k = 1, aby sa tam vyskytlo $¢q$w\$ (w ∈ Γ^*) len raz musí automat akceptovať slovo hneď po prvom prechode všetkých etáp

$q¢w\$ \vdash q_{[\sigma;\ w]}¢w\$ \vdash ¢q_{[\sigma;\ w]}w\$ \vdash ¢q'_{[\sigma;\ w]}w\$ \vdash^* ¢σε^*q'_{[ε;\ ε]}\$ \vdash ¢σε^*q_r\$ \vdash^* ¢q_rσε^*\$ \vdash ¢q_{c0}σε^*\$ \vdash ¢σq_{c1}ε^*\$ \vdash^* ¢σ\ ε^*q_{c2}\$ \vdash ¢σε^*q_k\$$

toto je ekvivalentné odvodeniu v G:

$\sigma \Rightarrow_G w$

2° k= n+ 1

Nech pre každú vetnú formu, ktorá je akceptovaná automatom A a jej akceptačný výpočet obsahuje konfiguráciu $¢q$w\$ (w ∈ Γ^*) práve n-krát existuje odvodenie v gramatike G (IP), tak potom pre každú vetnú formu ktorá je akceptovaná automatom A a jej akceptačný výpočet obsahuje konfiguráciu $¢q$w\$ (w ∈ Γ^*) práve n + 1-krát existuje odvodenie v gramatike G.

Nech w' obsahuje n + 1-krát konfiguráciu $¢q$w\$ (w ∈ Γ^*) a w obsahuje konfiguráciu $¢q$w\$ (w ∈ Γ^*) n-krát vo svojm odvodení. Nech ďalej platí: $q¢w'\$ \vdash^* q¢w\$$

Kedže w obsahuje konfiguráciu $¢q$w\$ (w ∈ Γ^*) práve n-krát tak existuje odvodenie v G, také že $\sigma \Rightarrow^* w$

V gramatike G potom platí: $\sigma \Rightarrow^* w \Rightarrow w'$

$\Rightarrow L(G) \supseteq L(A)$

L(A) = L(G)