Prezentácia z Formálnych jazykov a automatov

Ivan Piliš, 2i3, príklad č. 8.3

17.4.2001

Zadanie:

a. Dokážte, že neexistuje algoritmus, ktorý pre ľubovoľný TS rozhodne, či tento stroj akceptuje nejaké slovo obsahujúce podslovo `aba`.
b. Daný je TS A. L_{3A} je množina vsetkých slov nad abecedou Σ, ktoré obsahujú podslovo u dĺžky 3, ktoré je podslovom nejakého slova z L(A). Dokážte, ze L_{3A} je regulárny jazyk.

Riešenie:

a. Označme problém v zadaní ako P. Prevedieme problém P na problém $L(G) = \emptyset$, ktorý je pre rekurzívne vyčísliteľné (frázové) jazyky nerozhodnuteľný, teda neexistuje algoritmus, ktorý pre danú frázovú gramatiku (alebo teda Turingov stroj) rozhodne, či $L(G) = \emptyset$. Predpokladajme teda, že existuje algoritmus ALG, ktorý pre pre danú frázovú gramatiku rozhodne, či $L(G) = \emptyset$ a vezmime si ľubovoľný Turingov stroj (ako vstup zo zadania) $A=(K, \Sigma, \Gamma, \delta, q_o, F)$. Zostrojme Turingov stroj A' takto:

$K' = K \cup \{ q_s , q_{sB} , q_{sBa} , q_{sBab} , q_{sBaba} , q_h , q_{h_left} , q_{h_shift1} , q_{h_shift2a} , q_{h_shift2ab} , q_{h_shift2aba} , q_{h_shift2abaB} ,\}$

$\Sigma' = \Sigma$

$\Gamma' = \Gamma \cup \{B'\}$

$q_o' = q_s$

$F' = F$

Funkciu δ' dodefinujeme (pôvodné hodnoty ponecháme) o tieto hodnoty:

(1) $\delta' (q_s, c) = (q_s, c, -1)$, $c \in \Sigma$ - posun pred vstupné slovo
(2) $\delta' (q_s, B) = (q_{sB}, B', -1)$ - zápis falošného blanku
(3) $\delta' (q_{sB}, B) = (q_{sBa}, a, -1)$
(4) $\delta' (q_{sBa}, B) = (q_{sBab}, b, -1)$
(5) $\delta' (q_{sBab}, B) = (q_{sBaba}, a, 1)$ - zápis 'aba' pred vstupné slovo a B'
(6) $\delta' (q_{sBaba}, b) = (q_{sBaba}, b, 1)$
(7) $\delta' (q_{sBaba}, a) = (q_{sBaba}, a, 1)$
(8) $\delta' (q_{sBaba}, B') = (q_o, B', 1)$ - posun hlavy späť pred vstupné slovo

Ešte pridáme pre každé $(q_j, c, offset) \in \delta (q_i, B)$, kde $q_i , q_j \in K , c \in \Sigma$, $offset \in \{-1, 0, 1\}$, takýto blok:

(9) $\delta' (q_i, B') = (q_h, B', 1)$

(10) $\delta'(q_h, B) = (q_h, B, 0)$

(11) $\delta'(q_h, c) = (q_{h_left}, c, -1) \quad c \in \Sigma$ - zisíme, či je to Blank vľavo, alebo vpravo

(12) $\delta'(q_{h_left}, B') = (q_{h_shift1}, B', -1)$ - ak je vľavo, je to náš falošný, preto treba

(13) $\delta'(q_{h_shift1}, a) = (q_{h_shift1}, a, -1)$ pridať nový

(14) $\delta'(q_{h_shift1}, b) = (q_{h_shift1}, b, -1)$ - cez stav h_shift1 sa presunieme na úplný začiatok vstupného slova

(15) $\delta'(q_{h_shift1}, B) = (q_{h_shift2a}, a, 1)$ - cez stav h_shift2 shiftneme začiatok o 1 doľava a pridáme jeden ďalsí falošný Blank

(16) $\delta'(q_{h_shift2a}, a) = (q_{h_shift2ab}, b, 1)$

(17) $\delta'(q_{h_shift2ab}, b) = (q_{h_shift2aba}, a, 1)$

(18) $\delta'(q_{h_shift2aba}, a) = (q_{h_shift2abaB}, B', 1)$ - tu pridáme falošný Blank

(19) $\delta'(q_{h_shift2abaB}, B') = (q_j, c, offset)$ - na falošnom Blanku prevedieme to isté, ako pôvodný automat na pravom

Skonštruovali sme nový automat A' taký, že $L(A')=\{abaB'w \mid w \in L(A)\}$.

Dôkaz:

Nech najprv $w \in L(A)$. Ukážeme, že $abaB'w \in L(A')$. Teda existuje postupnosť konfigurácií $C_1, C_2, \ldots C_n$ taká , že C_n je akceptačná konfigurácia a C_1 je počiatočná konfigurácia so vstupným slovom w automatu A. Ukážeme, že existuje postupnosť konfigurácií $K_1, K_2, \ldots K_m$ automatu A' taká, K_m je akceptačná konfigurácia a K_1 má na páske rovnaké slovo ako C_1. Zaveďme ešste predtým pojem podobnosť konfigurácií '~' takto: $C_i \sim K_j \Leftrightarrow$ stav v C_i je rovnaký ako stav v K_j, slovo na páske v K_j je zreťazením "abaB' " a slova na páske v C_i a hlava v K_i sa nachádza o 4 pozície viac vpravo ako hlava v C_i (uvažujúc na oboch páskach iba slovo z pásky v C_i, tak sa hlava nachádza na rovnakej pozícii). Označme ešte W(K) , slovo na páske , S(K) stav v konfigurácii K. Máme teda $W(C_1) = W(K_1)$, $S(C_1) = q_o$, $S(K_1) = q_s$. Použitím (1) až (8) (resp. (2) až (8) v príade prázdneho slova w) sa dostaneme z konfigurácie K_1 do konfigurácie K_{i1} (teda $i1 \in \{8,9\}$) takej, že $C_1 \sim K_{i1}$. Uvažujme konfiguráciu C_2. Pri prechode z C_1 na C_2 muselo byť použitá nejaká hodnota δ funkcie. Uvažujme 2 možnosti:

1) pri prechode z C_1 na C_2 bol pod čítacou hlavou symbol rôzny od Blanku. Potom ale v automate A' je rovnaká hodnota δ funkcie, a teda aj $C_2 \sim K_{i1+1}$
2) pri prechode z C_1 na C_2 bol pod čítacou hlavou Blank. V tom prípade sú dve možnosti.
   2.1) tento Blank je za vstupným slovom (zrejme pri čítaní prvého znaku vstupného slova by to bolo možné len v prípade, že w je prázdne, ale uvažujme tento prípad, aby sme tieto uváhy mohli použiť aj pri prechode neskoršími konfiguráciami). Potom (9) a (10) by spôsobilo zacyklenie A'. Nedeterministický automat ale akceptuje slovo vždy, keď existuje (nejaký) akceptačný výpočet. Preto použitím hodnty δ funkcie prevzatej z automatu A' (a práve tej, ktorú sme použili pri prechode z C_1 na C_2) dostávame $C_2 \sim K_{i1+1}$.
   2.2) tento Blank je pred vstupným slovom. Potom v K_{i1} je pod čítacou hlavou symbol B' (náš falošný Blank), kdežto v C_1 je pod hlavou B. Keďže δ funkcia nie je v A pre B' definovaná, tak sa musia použiť dodofinované hodnoty (9) až (19). Nech $(q_j, c, offset) \in \delta(q_i, B)$, kde $q_i, q_j \in K, c \in \Sigma$, $offset \in \{-1, 0, 1\}$ je hodnota, ktorú sme použili v automate A v prechode z C_1 na C_2 (teda stav q_i sa zmenil na q_j, namiesto B sme zapísali c, a posun je offset). Podľa definície δ' máme pre

$(q_j, c, offset) \in \delta(q_i, B)$ definované hodnoty (9) až (19). Teda použijeme (9). Keďže na pásku nemožno zapisovať Blank, tak prázdne slovo môže byť na páske len v 1 okamihu – na začiatku, teda na vstupe. No a ak je prázdne slovo z jayzyka L(A), tak potom $q_o \in F$, a teda z (1) až (8) dostávame, že "abaB' " $\in$ L(A'). Uvažujem teda prípad, že w je neprázdne. Potom ale ani počas výpočtu sa nemôže na páske vyskytnúť prázdne slovo, a teda ďalej použijeme (11). Tým si v stave (q_{h_left}) zapamätáme, že sme na falošnom Blanku. Pomocou (12) zmeníme stav na q_{h_shift1}, ktorý nám indikuje, že budume posúvať prefix doľava a v tomto stave presúvame hlavu na začiatok – (13), (14). Ak je už hlava na začiatku, zapíšeme "abaB' " o jeden znak vľavo ako bolo predtým a pridáme ďalší falošný Blank – (15), (16), (17), (18). Teda z pôvodného "abaB'w" sme dostali "abaB'B'w". Hodnotou (18) sme nastavili hlavu opäť na B', na ktorom bola na začiatku. Teda môžeme použiť modifikované pravidlo pravidla, ktoré sme použili pri prechode z C_1 na C_2 - namiesto B prečítame B'. Po použití (19) na páske zostane opäť iba 1 falošný Blank. Navyše dostávame, že existuje také j , že $C_2 \sim K_{i1+j}$ (K_{i1+1}, K_{i1+2}, … K_{i1+j} sú konfigurácie, ktoré sme dostali použitím hodnôt (9), (11), (12), (13), (14), (15), (16), (17), (18), (19)). Teda existuje také i2 (= i1+j), že $C_2 \sim K_{i2}$.

Rovnakým postupom dostaneme, že existuje i3, i4, ... in také, že $C_3 \sim K_{i3}$, $C_4 \sim K_{i4}$, … $C_n \sim K_{in}$, a teda existuje také m, že K_m je akceptačná konfigurácia a zároveň je to posledná konfigurácia výpočtu vstupného slova w. Stačí totiž položiť m=in, pretože nakoľko $C_n \sim K_m$, tak $S(C_n) = S(K_m)$, a teda $S(K_m)$ je akceptačný stav. No a postupnosť $K_1, K_2, ... K_{i1}, K_{i1+1}, K_{i1+2}, \dots K_{i2}, K_{i2+1}, .. K_{i3}, \; . \; . \; . \; K_{in} = K_m$ je výpočet slova w automate A'. Teda abaB'w$\in$L(A') (pretože $C_n \sim K_m$).

Naopak, nech abaB'w$\in$L(A'). Ukážeme, že w$\in$L(A). Vezmime postupnosť konfigurácií $K_1, K_2, ... K_m$ vo výpočte w v automate A'. Ak sme pri prechode z K_i na K_{i+1} použili hodnotu, ktorá je definovaná aj v δ funkcii, tak ju aplikujme na pásku automatu A, kde robíme výpočet slova w. Ak sme použili ľubovoľnú dodefinovanú hodnotu okrem (19), tak neurobíme nič. Inak (teda ak išlo o (19)), aplikujme hodnotu $(q_j, c, offset) \in (q_i, B)$. Takto dostaneme výpočet slova w v automate L(A). Teda w $\in$ L(A).

Dokázali sme, že w$\in$L(A) $\Leftrightarrow$ abaB'w $\in$L(A').

Teraz spustíme algoritmus ALG z predpokladu na náš automat A'. Ten nám rozhodne, či existuje slovo w$\in$L(A') také, ktoré obsahuje podslovo aba. Keďže ale všetky slová z L(A) začínajú podslovom aba, tak ALG zistí, či L(A') = $\emptyset$. Keďže sme dokázali, že w$\in$L(A) $\Leftrightarrow$ abaB'w $\in$L(A'), tak L(A') = $\emptyset$ $\Leftrightarrow$ L(A) = $\emptyset$. Ak teda ALG odpovie ÁNO, tak L(A) = $\emptyset$. Ak odpovie NIE, tak L(A) $\neq \emptyset$. Dostávame tvrdenie:
Ak je problém P rozhodnuteľný, tak je aj problém L(A) = $\emptyset$ pre rekurzívne vyčísliteľné jazyky rozhodnuteľný. Posledne spomínaný problém ale rozhodnuteľný nie je, a preto nie je ani rozhodnuteľný problém P.

b. Nech A je ľubovoľný Turingov stroj. Nech L_3 je množina všetkých slov dĺžky 3 takých, ktoré sú podslovom nejakého slova z L(A), teda nech $L_3=\{u_1,u_2,\dots u_n\}$. To, že ich je konečný počet vyplýva z toho, že Σ je konečná množina symbolov, a teda aj počet trojprvkových podmnožin tejto množiny je konečný, a teda počet všetkých trojpísmenových slov nad abecedou Σ je konečný (presne šesťnásobok počtu týchto podmnožín). Uvažujme takúto gramatiku G:

G = (N,T,P, σ), kde N = {σ, α} , T = Σ (z automatu A), P je množina týchto pravidiel:

(1) $\sigma \rightarrow c\sigma \quad \forall c \in \Sigma$

(2) $\sigma \rightarrow u_1\alpha \mid u_2\alpha \mid u_3\alpha \mid \ldots \mid u_n\alpha \mid u_1 \mid u_2 \mid u_3 \mid \ldots \mid u_n$

(3) $\alpha \rightarrow c\alpha \mid c \quad \forall c \in \Sigma$

Zrejme G je regulárna. Ukážeme ešte, že L(G) = L_{3A}.

L(G) $\subseteq$ L_{3A}

Nech w$\in$ L(G). Potom existuje odvodenie slova w $\sigma \Rightarrow w_1 \Rightarrow w_2 \Rightarrow \ldots \Rightarrow w_n = w$ v gramatike G. Keďže zrejme slovo w neobsahuje neterminály, tak v poslednom kroku odvodenia muselo byť použité jedno z pravidiel (2), (3). Ak to bolo (2), skončili sme, pretože potom pre nejaké i platí $u_i \in w$. Ak to bolo (3), potom $\alpha \in w_{n-1}$. Keďže $\sigma \neq \alpha$, tak musí existovať j také, že $\alpha \notin w_j$ & $\alpha \in w_{j+1}$. Potom ale v kroku $w_j \Rightarrow w_{j+1}$ muselo byť použité pravidlo (2), a teda dostávame, že $u_i \in w$ pre nejaké i.

L(G) $\supseteq$ L_{3A}

Nech $w=c_1c_2\ldots c_n \in L_{3A}$. Potom existuje i také, že $c_ic_{i+1}c_{i+2} \in L_3$. Potom odvodenie v G slova w vyzerá takto:Najprv i-1 krát aplikujeme pravidlo 1 (podľa $c_1,c_2,\ldots c_{i-1}$). Potom pravidlo (2) podľa toho, pre ktoré k je $u_k = c_ic_{i+1}c_{i+2}$. Ak i+2 = n, tak pravidlom (2) skončíme. Ak n>i+2, tak n-(i+2)+1 krát použijeme pravidlo (3) a dostaneme slovo w.

Dokázali sme, že L(G) = L_{3A}. Teda regulárna gramatika vyhovuje zadaniu. Ešte si uvedomme, že sme nenašli algoritmus, ktorý z daného TS A zostrojí regulárnu gramatiku pre L_{3A} , pretože množinu L_3 sme nezostrojovali, iba sme povedali, že taká existuje. Na druhej strane našou úlohou bolo iba ukázať, že L_{3A} je regulárny jazyk, a nie zostrojiť ho z daného automatu. Preto sme úlohu splnili.