Veremautomaták és környezetfüggetlen nyelvtanok ekvivalenciája

Varga Richárd

Tétel:

Minden környezetfüggetlen nyelvtanhoz meg lehet adni verem automatát úgy, hogy a verem automata (üres veremmel vagy végállapottal) ugyanazt a nyelvet ismeri fel, amit a környezetfüggetlen nyelvtan generál.

Bizonyítás:

Vegyünk egy G=(N ,Σ, R ,S) környezetfüggetlen nyelvtant, amelyhez megadunk egy P verem automatát úgy, hogy az üres veremmel ugyanazt a nyelvet ismeri fel, amit a G nyelvtan generál. Legyen ez $P = (\{q\}, \Sigma, \Gamma, \delta, q, Z_0, \emptyset)$. Itt a veremautomata 1 állapotból áll, és erre is igaz lesz, hogy fel tudja ismerni a környezetfüggetlen nyelvtant. Mivel most a végállapotoknak nincs szerepe itt, ezért a végállapot halmaz lehet üres. A verem szimbólumai(Γ) a nyelvtan terminálisai(Σ) és nemterminálisai(N) lesznek. Kezdőszimbóluma(Z_0) a nyelvtan kezdőszimbólumával lesz egyenlő(S). Az átmenetfüggvényt definiáljuk a következőképpen:

- minden $A \in N$-re($\in \Gamma$) $\delta(q, \varepsilon, A) = \{(q, \alpha) \mid A \to \alpha \in R\}$,
  tehát ha q állapotból nem olvasunk inputot és A van a verem tetején, akkor ugyanúgy q állapotban marad, és A helyére beírja valamelyik $A \to \alpha$ szabályt(A bal oldalú szabály jobb oldalát). Az összes ilyen (q, α) alakú párok halmaz lesz ebben a halmazban (ezért látszik, hogy nemdeterminisztikus).
- minden $a \in \Sigma$-ra($\in \Gamma$) $\delta(q, a, a) = \{(q, \varepsilon)\}$, vagyis ha ugyanolyan terminális van a verem tetején, mint ami az input, akkor elfogadja, és törli a veremből.

A veremautomatánk a nyelvtannak a levezetéseit fogja szimulálni.
Elég igazolni, hogy minden $X \in (N \cup \Sigma)$ és $w \in \Sigma^*$ esetén a nyelvtanban X-ből levezethető a w akkor és csakis akkor, ha $(q,w,X) \vdash^* (q, \varepsilon, \varepsilon)$, vagyis ha a (q,w,X) konfigurációból valamennyi lépésből el lehet jutni $(q, \varepsilon, \varepsilon)$-ba. Ha X helyére a verem kezdőszimbólumát(S) írunk, akkor S-ből levezethető a w a nyelvtanban akkor és csakis akkor, ha $(q,w,S) \vdash^* (q, \varepsilon, \varepsilon)$, és mivel S a kezdőszimbólum, ezért az azt jelenti, hogy a w szót elfogadta a veremautomata üres veremmel.
Tegyük fel, hogy X-ből valamennyi n lépésből levezethető a *w.*

- Ha n = 0, akkor az csak úgy lehet, ha $X = w$, ami csak akkor lehet, ha mind a kettő eleme a Σ-nak. Ekkor a $(q,w,X) = (q,w,w)$, w egyetlen betűje a Σ-nak, amiből el tudunk jutni a $(q, \varepsilon, \varepsilon)$-ba, mivel az elején ilyen átmenetek letttek definiálva.
- Ha n-re teljesül, akkor az indukciós lépés szerint n+1-re is teljesülni fog.
  $X \Rightarrow X_1 \ldots X_k \Rightarrow^n w_1 \ldots w_k = w$. Az első lépésben alkamazzuk a nyelvtan egyik szabályát, majd lesz még n lépés, és abból levezetjük a w-t. X_1-ből levezetjük a w_1-t, X_k-ból a w_k-t, és ezek kokatenációj a *w.* Megállapítjuk, hogy
    - $X \to X_1 \ldots X_k \in R$, vagyis ezek szabályai a nyelvtannak, és
    - $X_i \Rightarrow^{n_i} w_i$ minden $1 \leq i \leq k$-ra, ahol $n_i \leq n$(mivel $X_1 \ldots X_k$ együttes lépés n darab). Ez az állítás már hasonló ahhoz, amit feltettünk, de ez legfeljebb n hosszúságra. Használjuk erre az indukciós feltevést:

Minden $1 \leq i \leq k$ esetén $(q,w_i,X_i) \vdash^* (q, \varepsilon, \varepsilon)$. Innen kapjuk, hogy
$(q,w,X) = (q,w_1 \ldots w_k ,X)$, előzőek alapján ($w = w_1 \ldots w_k$)

$\vdash (q,w_1 \ldots w_k ,X_1 \ldots X_k)$, X nemterminális helyére beírom a szabály jobb oldalát

$\vdash^* (q,w_2 \ldots w_k ,X_2 \ldots X_k)$

$\ldots$

$\vdash^*(q,wk ,Xk)$

$\vdash^* (q, \varepsilon, \varepsilon)$.

Ezzel végeztünk a bizonyítás egyik oldalával.

*PÉLDA

Tétel: Minden veremautomatával felismert nyelv környezetfüggetlen.

Bizonyítás:

Legyen $P = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$ veremautomata. Most ehhez adunk meg egy olyan G környezetfüggetlen nyelvtant, ami ugyanazt a nyelvet generálja, mint amit a P veremautomata üres veremmel felismer ($L(G) = L_\emptyset(P)$). Legyen $G=(N ,\Sigma, R ,S)$,ahol S egy új szimbólum, a terminálisok megyegyeznek, a nemterminálisok pedig:
$N = \{S\} \cup \{[qZr] \mid q,r \in Q, Z \in \Gamma\}$, vagyis az S és még a [qZr] halamazok uniója. Ebből már látható, hogy sok nemterminálisa lesz a nyelvtannak. Egy [qZr] hármas jelentése az, hogy ha a veremautomata q állapotban van, a verem legfelső szimbóluma Z, akkor az r állapotba jutva tudja kivenni (törölni) Z-t a veremből.
Most adjuk meg az R szabályok halmazát, amire teljesülnek, hogy:

- minden $q \in Q$-ra legyen $S \rightarrow [q_0Z_0q]$ szabály R-ben, vagyis $[q_0Z_0q]$ egy nemterminális lényegében.
- veremautomata átmeneteiből késztjük el a nyelvtan szabályait:
  kétféle átmenet van, az egyik mikor a verembe írunk valamit a Z helyére:
  minden $q \in Q$, $a \in (\Sigma \cup \{\varepsilon\})$, $Z \in \Gamma$-ra,
  ha $(s_0, Z_1 \ldots Z_k) \in \delta(q, a, Z)$, (ahol $k \geq 1$, $Z_1, \ldots , Z_k \in \Gamma$) akkor
  minden $s_1, \ldots ,s_k \in Q$ sorozatra legyen
  $[qZs_k] \rightarrow a[s_0Z_1s_1] \ldots [s_{k-1}Z_ks_k]$ szabály R-ben. Tehát amit az átmenetfüggvény elolvasott, azt írja ki a nyelvtan, és végén a Z el fog tűnni a veremből, mert üres veremmel felismeri. Ez úgy lehetséges, hogy először az s_0 állapotba megy át, és Z-k bekerülnek a verembe, majd sorban törlődnek onnan. Nem tudjuk, hogy milyen állapotokba kerül közben, ezért minden lehetséges $s_1, \ldots ,s_k \in Q$ sorozatra fel kell venni $[qZs_k] \rightarrow a[s_0Z_1s_1] \ldots [s_{k-1}Z_ks_k]$ szabályt($[qZs_k]$ lesz az utolsó, amibe kerül, ahonnan a Z kitörlődik). Itt látjuk, hogy Z_1-t kitörli valamilyen s_1 állapottal, Z_2-t s_2 állapottal, és így tovább.
- a másik mikor a Z-t kitöröljük:
  Ez az előzőnek egy speciális esete, mikor a $(s_0, Z_1 \ldots Z_k) \in \delta(q, a, Z)$-ben a $k = 0$.
  Minden $q \in Q$, $a \in (\Sigma \cup \{\varepsilon\})$, $Z \in \Gamma$-ra, ha $(s_0, \varepsilon) \in \delta(q, a, Z)$,
  akkor legyen $[qZs_0] \rightarrow a$ szabály R-ben.

Elegendő megmutatni, hogy minden $q,r \in Q$, $Z \in \Gamma$ ´es $w \in \Sigma^*$ esetén
$(q,w, Z) \vdash^* (r, \varepsilon, \varepsilon)$ akkor és csak akkor, ha $[qZr] \Rightarrow^* w$. (Azért választottunk ilyen 3-asokat nemterminálisoknak, mert itt (q,w, Z) q-ból indul, Z van a verem tetején, és mire a Z eltűnik, r állapotba kerül). Ilyenkor igaz arra is, amikor $q = q_0$ és $Z = Z_0$.

$w \in L_\emptyset(P)$
$\Leftrightarrow (q_0,w, Z_0) \vdash^* (r, \varepsilon, \varepsilon)$
$\Leftrightarrow [q_0Z_0r] \Rightarrow^* w$
$\Leftrightarrow S \Rightarrow [q_0Z_0r] \Rightarrow^* w$
$\Leftrightarrow w \in L(G)$
Ebben az esetben azt jelenti, hogy (q,w, Z)-ból levezethetjük (r, ε, ε), vagyis ha q a kezdőállapot Z a kezdőszimbóluma, az azt jelenti hogy w-t felismeri üres veremmel és $[q_0Z_0r] \Rightarrow^* w$. Ha ebből levezethető, akkor S-ből is levezethető, tehát w-t generálja a nyelvtan.
Tegyük fel, hogy $[qZr] \Rightarrow^n w$ valamilyen $n \geq 1$-re.

- $n = 1$
  $[qZr] \Rightarrow w$, ez csak úgy lehet, ha $w = a \in (\Sigma \cup \{\varepsilon\})$ és $(r, \varepsilon) \in \delta(q, a, Z)$. Tehát $(q,w, Z) = (q, a, Z) \vdash (r, \varepsilon, \varepsilon)$. 1 Lépésben csak olyat vezethetek le, ami terminális
- $n = n + 1$ (indukciós lépés).
  Legelső lépés: $a[s_0Z_1s_1] \ldots [s_{k-1}Z_ks_k]$ minden ilyen alakú. mivel csak ilyen szabályokat definiáltunk.
  $[qZr] \Rightarrow a[s_0Z_1s_1] \ldots [s_{k-1}Z_ks_k] \Rightarrow^n aw_1 \ldots w_k = w$
  Ebből következik, hogy:
  - $[qZr] \to a[s_0Z_1s_1] \ldots [s_{k-1}Z_ks_k] \in R$, $r = s_k$ (így adtuk meg a nyelvtanban), azaz $(s_0, Z_1 \ldots Z_k) \in \delta(q, a, Z)$. Ha ez az átmenet történik, akkor minden $s_1 \ldots s_k$ állapot esetén ezt a szabály betesszük az R-be.
  - Minden nemterminálisból levezetünk egy w szót, tehát
    Minden $1 \leq i \leq k$-ra $[s_{i-1}Z_is_i] \Rightarrow^{ni} w_i$, és $ni \leq n$. Erre már lehet Alkamazni az indukciós feltevést:

Minden $1 \leq i \leq k$-ra $(s_{i-1},w_i, Z_i) \vdash^* (s_i, \varepsilon, \varepsilon)$. Kapjuk, hogy:
$(q,w, Z) = (q, aw_1 \ldots w_k , Z)$
$\vdash (s_0,w_1 \ldots w_k , Z_1 \ldots Z_k)$
$\vdash^* (s_1,w_2 \ldots w_k , Z_2 \ldots Z_k)$
$\ldots$
$\vdash^* (s_{k-1},w_k , Z_k)$
$\vdash^* (s_k , \varepsilon, \varepsilon) = (r, \varepsilon, \varepsilon)$.
Legelső lépésben a-t beolvassuk és Z helyére írom a $Z_1 \ldots Z_k$-t és s_0 állapotba kerülünk, mert csak akkor van ilyen szabály benne, ha létezik ilyen átmenet. Innentől alkalmazva az indukciós feltevést eljutunk (r, ε, ε)-hoz.

Feladatok:

- Legyen $L= \{a^nb^mc^k \mid n = m \text{ vagy } m = k\}$
  Adjunk meg egy olyan veremautomatát, ami az L nyelvet ismeri fel.

- Vegyünk egy G nyelvtant:
  $S \to A \mid B \mid a$
  $A \to bBb \mid B$
  $B \to A \mid S \mid b$
  Konstruáljuk azt a $P=(\{q\}, \Sigma, \Gamma, \delta, q, K, \emptyset)$ veremautomatát, ami ugyanazt a nyelvet ismeri fel, mint amit a nyelvtan generál.

- Konstruáljuk meg az alábbi veremautomatához azt a nyelvtant, ami ugyanazt a nyelvtant generálja