

Formális nyelvek - 9.

Csuhaj Varjú Erzsébet

Algoritmusok és Alkalmazásaik Tanszék
Informatikai Kar
Eötvös Loránd Tudományegyetem
H-1117 Budapest
Pázmány Péter sétány 1/c
E-mail: csuhaj@inf.elte.hu

Véges automata - folytatás

Ismeretes, hogy minden A nemdeterminisztikus véges automatához meg tudunk konstruálni egy G reguláris grammatikát úgy, hogy $L(G) = L(A)$ teljesül. Az is tudott, hogy minden G reguláris grammatikához meg tudunk konstruálni egy A véges automatát úgy, hogy $L(A) = L(G)$.

Kérdés: Létezik-e olyan reguláris nyelv, amely nemdeterminisztikus véges automatával felismerhető, de nem ismerhető fel determinisztikus véges automatával?

Válasz: Nem létezik.

Determinisztikus versus nemdeterminisztikus véges automata

Tétel

Minden $A = (Q, T, \delta, Q_0, F)$ nemdeterminisztikus véges automatához meg tudunk konstruálni egy $A' = (Q', T, \delta', q'_0, F')$ determinisztikus véges automatát úgy, hogy $L(A) = L(A')$ teljesül.

Bizonyításvázlat

Legyen Q' a Q halmaz összes részhalmazainak halmaza.

Definiáljuk a $\delta' : Q' \times T \rightarrow Q'$ függvényt a következőképpen:

$$\delta'(q', a) = \bigcup_{q \in q'} \delta(q, a).$$

Továbbá legyen $q'_0 = Q_0$ és $F' = \{q' \in Q' \mid q' \cap F \neq \emptyset\}$

Először megmutatjuk, hogy $L(A) \subseteq L(A')$. Az állítás bizonyításához megmutatjuk, hogy a következő lemma állítása teljesül.

Bizonyításvázlat - folytatás

Lemma 1

Minden $p, q \in Q$, $q' \in Q'$ és $u, v \in T^*$ esetén, ha

$$qu \Longrightarrow_A^* pv \text{ és } q \in q',$$

akkor van olyan $p' \in Q'$, hogy

$$q'u \Longrightarrow_{A'}^* p'v \text{ és } p \in p'.$$

Lemma 1 bizonyításának vázlata:

Az állítás a $qu \Longrightarrow_A^* pv$ redukcióban szereplő lépések száma szerinti indukcióval bizonyítható.

Nulla számú lépés esetében az állítás triviálisan fennáll. Tegyük fel, hogy teljesül valamely n lépésre, ahol $n \geq 0$.

Lemma 1 bizonyításának vázlata - folytatás

(**Lemma 1** : Minden $p, q \in Q$, $q' \in Q'$ és $u, v \in T^*$ esetén, ha

$$qu \Longrightarrow_A^* pv \text{ és } q \in q',$$

akkor van olyan $p' \in Q'$, hogy

$$q'u \Longrightarrow_{A'}^* p'v \text{ és } p \in p'.)$$

Álljon a $qu \Longrightarrow_A^* pv$ redukció $n + 1$ lépésből. Akkor valamely $q_1 \in Q$ és $u_1 \in T^*$ -re

$$qu \Longrightarrow_A q_1u_1 \Longrightarrow_A^* pv$$

teljesül. Így van olyan $a \in T$, amelyre $u = au_1$ és $q_1 \in \delta(q, a)$. De $\delta(q, a) \subseteq \delta'(q', a)$, hacsak $q \in q'$, így választhatjuk q'_1 -et $q'_1 = \delta'(q', a)$ -nek, ahonnan

$$q'u \Longrightarrow_{A'} q'_1u_1$$

következik $q_1 \in q'_1$ -re. Az indukciós feltevés alapján valamely $p' \in Q'$ -re

$$q'_1u_1 \Longrightarrow_{A'}^* p'v \text{ és } p \in p'.$$

Azaz, Lemma 1 fennáll.

Bizonyításvázlat - folytatás

Legyen $u \in L(A)$, azaz teljesüljön $q_0 u \Longrightarrow_A^* p$ valamely $q_0 \in Q_0$ -ra és legyen $p \in F$.

Ekkor Lemma 1 alapján valamely p' -re fennáll, hogy $q'_0 u \Longrightarrow_{A'}^* p'$ és $p \in p'$.

Az F' definíciója alapján $p \in p'$ -ből és $p \in F$ -ből az következik, hogy $p' \in F'$, ahonnan $L(A) \subseteq L(A')$.

Bizonyításvázlat - folytatás

Ezután megmutatjuk, hogy $L(A') \subseteq L(A)$.

Az állítást Lemma 2 segítségével bizonyítjuk.

Lemma 2

Minden $p', q' \in Q'$, $p \in Q$ és $u, v \in T^*$ esetén, ha

$$q'u \Longrightarrow_{A'}^* p'v \text{ és } p \in p',$$

akkor van olyan $q \in Q$, hogy

$$qu \Longrightarrow_A^* pv \text{ és } q \in q'.$$

Bizonyításvázlat - folytatás

Lemma 2 bizonyításának vázlata: A lépések száma szerinti indukcióval történik. Nulla számú lépés esetén az állítás triviális.

Tegyük fel, hogy az állítás igaz n lépésre, ahol $n \geq 0$, és álljon a $q'u \Longrightarrow_{A'}^* p'v$ redukció $n + 1$ lépésből.

Akkor

$$q'u \Longrightarrow_{A'}^* p'_1 v_1 \Longrightarrow_{A'} p'v,$$

ahol $v_1 = av$ valamely $a \in T$ -re és $p'_1 \in Q'$.

Ekkor

$$p \in p' = \delta'(p'_1, a) = \bigcup_{p \in p'_1} \delta(p_1, a),$$

azaz, lennie kell olyan $p_1 \in p'_1$ -nek, amelyre $p \in \delta(p_1, a)$.

Ekkor p_1 -re teljesül, hogy $p_1 v_1 \Longrightarrow_A p'v$, és az indukciós hipotézis alapján $q \in q'$, ahol $qu \Longrightarrow_A^* p_1 v_1$. Azaz, Lemma 2 állítása teljesül.

Bizonyításvázlat - folytatás

Végezetül legyen $q'_0 u \Longrightarrow_{A'}^* p'$ és $p' \in F$. Az F' definíciója alapján van olyan $p \in p'$, ahol $p \in F$, és így Lemma 2 alapján valamely $q_0 \in q'_0$ -ra teljesül $q_0 u \Longrightarrow_A^* p$. Így $L(A') \subseteq L(A)$.

Ezzel a tétel bizonyítását befejeztük.

Példa

Legyen $A = (Q, T, \delta, Q_0, F)$ nemdeterminisztikus véges automata, ahol

$$Q = \{0, 1, 2\}, T = \{a, b\}, Q_0 = \{0\}, F = \{2\},$$

$$\delta(0, a) = \{0, 1\}, \quad \delta(0, b) = \{1\},$$

$$\delta(1, a) = \emptyset, \quad \delta(1, b) = \{2\},$$

$$\delta(2, a) = \{0, 1, 2\}, \quad \delta(2, b) = \{1\}.$$

Példa - folytatás

A vele ekvivalens determinisztikus véges automata állapotai a következők lesznek:

$$\emptyset, \{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}.$$

A további komponensek pedig $F = \{\{2\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}$, valamint δ' , ahol

$$\begin{array}{ll} \delta'(\emptyset, a) = \emptyset, & \delta'(\emptyset, b) = \emptyset, \\ \delta'(\{0\}, a) = \{0, 1\}, & \delta'(\{0\}, b) = \{1\}, \\ \delta'(\{1\}, a) = \emptyset, & \delta'(\{1\}, b) = \{2\}, \\ \delta'(\{2\}, a) = \{0, 1, 2\}, & \delta'(\{2\}, b) = \{1\}, \\ \delta'(\{0, 1\}, a) = \{0, 1\}, & \delta'(\{0, 1\}, b) = \{1, 2\}, \\ \delta'(\{0, 2\}, a) = \{0, 1, 2\}, & \delta'(\{0, 2\}, b) = \{1\}, \\ \delta'(\{1, 2\}, a) = \{0, 1, 2\}, & \delta'(\{1, 2\}, b) = \{1, 2\}, \\ \delta'(\{0, 1, 2\}, a) = \{0, 1, 2\}, & \delta'(\{0, 1, 2\}, b) = \{1, 2\}. \end{array}$$

Következmények

Korollárium

A reguláris nyelvek osztálya (\mathcal{L}_3) zárt a komplement műveletére nézve.

A bizonyítás vázlata :

Legyen L egy $A = (Q, T, \delta, q_0, F)$ véges automata által felismert nyelv. Akkor $\bar{L} = T^* - L$ felismerhető a $\bar{A} = (Q, T, \delta, q_0, Q - F)$ véges automatával.

Következmények

Korollárium

A reguláris nyelvek osztálya zárt a metszet műveletére nézve.

Bizonyításvázlat:

Ismeretes, hogy a reguláris nyelvek osztálya zárt az unió műveletére nézve. Mivel

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}},$$

az állítás következik.

Következmények

Korollárium

Eldönthető, hogy két reguláris grammatika azonos nyelvet generál-e vagy sem.

Bizonyításvázlat:

Legyenek G_1 és G_2 reguláris grammatikák, amelyek rendre az L_1 és L_2 nyelvet generálják. Az $L_3 = (L_1 \cap \bar{L}_2) \cup (\bar{L}_1 \cap L_2)$ nyelv szintén reguláris, így van olyan G_3 reguláris grammatika, amely L_3 -at generálja. Ekkor azonban $L_1 = L_2$ akkor és csak akkor, ha $L_3 = \emptyset$, amely minden G_3 környezetfüggetlen grammatika esetében eldönthető.

A reguláris nyelvosztály egy algebrai karakterizációja

Legyen L egy T ábécé feletti nyelv.

Az L nyelv által indukált E_L reláció alatt egy olyan bináris relációt értünk a T feletti szavak halmazán, amelyre a következő teljesül: bármely $u, v \in T^*$ -ra uE_Lv akkor és csak akkor, ha nincs olyan $w \in T^*$, hogy az uw és a vw szavak közül pontosan egy eleme L -nek.

Megjegyzés: E_L ekvivalencia reláció és jobb-invariáns. (Jobb-invariáns: ha uE_Lv , akkor uwE_Lvw is fennáll minden $w \in T^*$ szóra).

Az E_L reláció indexén ekvivalencia osztályainak számát értjük.

Tétel

$L \subseteq T^*$ akkor és csak akkor ismerhető fel determinisztikus véges automatával, ha E_L véges indexű.

Bizonyításvázlat

Tegyük fel, hogy $L \in T^*$ és E_L véges indexű. Legyenek $u_0 = \varepsilon, u_1, \dots, u_n$ az egyes ekvivalencia osztályokat reprezentáló szavak és jelöljük $[u_i]$ -vel, $0 \leq i \leq n$, az ekvivalencia osztályt.

Legyen $A = (Q, T, \delta, s_0, F)$ determinisztikus véges automata, ahol

$Q = \{[u_i] \mid 0 \leq i \leq n\}$, $[u_0]$ a kezdőállapot és $F = \{[u_j] \mid u_j \in L, 1 \leq j \leq n\}$.

Definiáljuk a δ állapot-átmenet függvényt a következő módon: $\delta([u_i], a) = [u_i a]$, $0 \leq i \leq n$.

Mivel E_L jobb-invariáns, ezért δ értékei nem függenek az $[u_i]$ osztályokat reprezentáló elemek választásától, továbbá, az elfogadó állapotok halmaza sem függ reprezentánsainak választásától.

A δ definíciójából közvetlenül adódik, hogy bármely $u \in T^*$ szóra $[q_0]u \Rightarrow^* [u]$. Ebből adódik, hogy L felismerhető egy determinisztikus véges automatával.

Bizonyításvázlat - folytatás

Tegyük most fel, hogy $L \subseteq T^*$ felismerhető egy $A = (Q, T, \delta, q_0, F)$ determinisztikus véges automatával.

Definiáljuk az E ekvivalencia relációt T^* -on a következőképpen: az u és v T feletti szavakra uEv akkor és csak akkor, ha van az A automatának olyan q állapota, hogy

$$q_0u \Longrightarrow_A^* q \text{ és } q_0v \Longrightarrow_A^* q$$

teljesül. (Nyilvánvaló, hogy E ekvivalencia reláció.)

Megmutatjuk, hogy uEv -ből uE_Lv következik. Tegyük fel az ellenkezőjét, vagyis, hogy valamely u', v', w' szavakra fennáll az, hogy $u'E_Lv'$, de csak egyike az $u'w'$ és a $v'w'$ szavaknak eleme az L nyelvnek.

Ekkor $q_0u'w' \Longrightarrow^* q_1$, ahol $q_1 \in F$ és $q_0v'w' \Longrightarrow^* q_2$, ahol $q_2 \notin F$. Mivel azonban $u'E_Lv'$ fennáll, ezért léteznie kell olyan $q \in Q$ állapotnak, hogy $q_0u \Longrightarrow_A^* q$ és $q_0v \Longrightarrow_A^* q$ teljesül.

Ez ellentmondás, így állításunk fennáll, ami azt jelenti, hogy E_L indexe nem nagyobb, mint E indexe, amely nem nagyobb az A automata állapotainak számánál, vagyis véges.

Példa a tétel használatára

Legyen $L = \{a^k b^k \mid 0 \leq k\}$. Megmutatjuk, hogy L nem reguláris nyelv. Tegyük fel az ellenkezőjét. Akkor létezik olyan determinisztikus véges automata, amely elfogadja L -et. Ebben az esetben E_L véges indexű. Ha E_L véges indexű, akkor $a^m E_L a^n$ fennáll valamely $m > n$ számpárra. De ez nem lehetséges, mivel csak az $a^m b^m$ és az $a^n b^m$ szavak egyike eleme L -nek. Ebből következően E_L nem lehet véges indexű, azaz, L nem reguláris nyelv.

Ez egy újabb megerősítése annak, hogy $\mathcal{L}_3 \subset \mathcal{L}_2$.

L generálható a $G = (\{S\}, \{a, b\}, \{S \rightarrow \varepsilon, S \rightarrow aSb\}, S)$ lineáris grammatikával.

Minimális számú állapottal rendelkező véges automata

Az A determinisztikus véges automata **minimális állapotszámú (minimális)**, ha nincs olyan A' determinisztikus véges automata, amely ugyanazt a nyelvet ismeri fel, mint A , de A' állapotainak száma kisebb, mint A állapotainak száma.

Tétel

Az L reguláris nyelvet felismerő minimális (állapotszámú) determinisztikus véges automata az izomorfizmus erejéig egyértelmű.

A bizonyítás az előző tétel alapján könnyen belátható.

Véges automata megkülönböztethető állapotai

Legyen $A = (Q, T, \delta, q_0, F)$ determinisztikus véges automata. Definiáljunk egy R relációt a Q állapotthalmazon úgy, hogy pRq , ha minden egyes $x \in T^*$ input szóra fennáll, hogy $px \Rightarrow_A^* r$ akkor és csak akkor, ha $qx \Rightarrow_A^* r'$ valamely $r, r' \in F$ állapotokra. ($r = r'$ lehetséges).

Azt mondjuk, hogy p és q **megkülönböztethetők**, ha van olyan $x \in T^*$, amelyre vagy $px \Rightarrow_A^* r$, $r \in F$, vagy $qx \Rightarrow_A^* r'$, $r' \in F$, de mindkét redukció egyszerre nem áll fenn. Egyébként p és q nem megkülönböztethetők.

Ha p és q **nem megkülönböztethetők**, akkor $\delta(p, a) = s$ és $\delta(q, a) = t$ sem megkülönböztethetők egyetlen $a \in T$ -re sem.

Összefüggő determinisztikus véges automata

Legyen $A = (Q, T, \delta, q_0, F)$ determinisztikus véges automata. A q állapotot a kezdő-állapotból **elérhetőnek** mondjuk, ha létezik $q_0x \Longrightarrow^* q$ redukció, ahol x valamely T feletti szó.

Az $A = (Q, T, \delta, q_0, F)$ determinisztikus véges automatát **összefüggőnek** mondjuk, ha minden állapota elérhető a kezdőállapotból.

Összefüggő determinisztikus véges automata

Definiáljuk a H halmazt a következőképpen:

Legyen $H_0 = \{q_0\}$, $H_{i+1} = H_i \cup \{r \mid \delta(q, a) = r, q \in H_i, a \in T\}$, $i = 1, 2, \dots$. Akkor létezik olyan $k \geq 0$, amelyre $H_k = H_l$, ahol $l \geq k$. Legyen $H = H_k$. Könnyen látható, hogy H azoknak az állapotoknak a halmaza, amelyek a kezdőállapotból elérhetők.

Ezután definiáljuk az $A' = (Q', T, \delta', q_0, F')$ determinisztikus véges automatát a következőképpen:

$Q' = H$, $F' = F \cap H$ és $\delta' : H \times T \rightarrow H$ úgy, hogy $\delta'(q, a) = \delta(q, a)$, ha csak $q \in H$.

Könnyen megmutatható, hogy A' összefüggő és ugyanazt a nyelvet fogadja el, amelyet A . Továbbá, A' az A legnagyobb összefüggő részautomatája.

Algoritmus véges determinisztikus automata minimalizálására

Tegyük fel, hogy az automata összefüggő. Eután partícionáljuk (bontsuk ekvivalencia osztályokra) a Q halmazt a következőképpen:

1.: Tekintsük a Q állapothalmazt. Osszuk két ekvivalenciaosztályra: F -re és $Q - F$ -re. Az F -beli állapotok megkülönböztethetők a $Q - F$ -beli állapotoktól. (Bármely elfogadó állapot megkülönböztethető egy nem elfogadó állapottól az üres szóval.) Ismételjük meg az ekvivalenciaosztályok további ekvivalenciaosztályokra való szétbontását addig, amíg a partícionálás változatlan marad a következőképpen:

2.: Ha adott egy partícionálás, $\mathcal{B} = B_1 \cup \dots \cup B_s$ akkor minden egyes B_i ekvivalenciaosztály bármely két p és q állapota pontosan akkor marad együtt, ha minden $a \in T$ -re a $\delta(p, a)$ és a $\delta(q, a)$ egyazon ekvivalenciaosztályba tartozik, egyébként a B_i ekvivalenciaosztályt szétbontjuk. Ezt az eljárást addig ismételjük, ameddig változás van, egyébként az eljárás leáll.

Algoritmus - folytatás

3.: Minden egyes B_i ekvivalenciaosztályt reprezentáljunk egy b_i szimbólummal. Konstruáljuk meg a determinisztikus véges automatát a következőképpen: legyen $A = (Q', T, \delta', q'_0, F')$, ahol

$$Q' = \{b_i \mid B_i \text{ a 2. eljárás során nyert ekvivalenciaosztály}\}.$$

$q'_0 = b_0$ a q_0 -t tartalmazó ekvivalenciaosztály reprezentánsa. $\delta'(b_i, a) = b_j$, ha $\delta(q, a) = p$, $q \in B_i$, $p \in B_j$. $F' = \{b_f\}$ azon ekvivalenciaosztály reprezentánsa, amely F elemeit tartalmazza.