Formális nyelvek - 8.

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

Automaták

Formális nyelvek megadása nemcsak generatív, hanem felismerő eszközökkel is lehetséges, azaz olyan számítási eszközök segítségével, amelyek szavak (sztringek) feldolgozására és azonosítására alkalmasak.

Ilyen eszköz például az automata, amely egy szó (sztring) mint input hatására kétféleképpen viselkedhet: vagy elfogadja, vagy elutasítja (igen vagy nem). Megjegyezzük, hogy olyan lehetőség is ésszerű, hogy az automata egyáltalán nem ad választ.

Az automaták analitikus eszközök, míg a grammatikák szintetizáló megközelítést alkalmaznak.

Véges automata

Definíció

A véges automata egy rendezett ötös,

$$A = (Q, T, \delta, q_0, F),$$

- ahol Q állapotok véges nemüres halmaza,
- T input szimbólumok ábécéje,
- $\delta : Q \times T \to Q$, az ún. állapot-átmeneti (röviden átmeneti) függvény,
- $q_0 \in Q$ az ún. kezdeti állapot (kezdőállapot),
- $F \subseteq Q$ az ún. elfogadó állapotok halmaza.

Szokás az elfogadó állapot helyett végállapotot is mondani.

Véges automata - folytatás

A δ leképezés kiterjeszthető $Q \times T^*$ halmazból Q halmazba való $\widehat{\delta}$ leképezéssé a következő módon:

- $\widehat{\delta}(q, \varepsilon) = q$,
- $\widehat{\delta}(q, xa) = \delta(\widehat{\delta}(q, x), a)$ minden $x \in T^*$-ra és minden $a \in T$-re.

Véges automata - működési mód

A véges automata diszkrét időintervallumokban végrehajtott lépések sorozata által működik.

Minden egyes lépés során az automata elolvassa a következő input szimbólumot és átmegy egy olyan állapotba, amelyet az állapot-átmeneti függvény meghatároz az adott aktuális input szimbólumra és azon állapotra vonatkozóan, amelyben az automata az adott pillanatban van.

Véges automata - működési mód

Az input szó (input sztring) - vagy röviden az input - megadását elképzelhetjük oly módon, hogy a szó egy ún. input szalagon van, amely cellákból áll, és minden egyes cella tartalmaz egy szimbólumot.

Az input szalagon van egy ún. olvasófej, amelynek a mozgását egy vele összekötött, az állapotok változását kontrolláló ún. véges kontroll, valamint az a szimbólum határozza meg, amelyet az a cella tartalmaz, ahol az adott pillanatban az olvasófej áll.

Az olvasófej kezdetben az első cellán áll, a véges kontroll pedig a kezdőállapotot jelöli meg. Minden egyes pillanatban az olvasófej egy cellával jobbra mozdul, a véges kontroll pedig állapotát az automata aktuális állapotától és a cellában található betűtől függően változtatja meg.

Véges automata - működési mód

Kezdetben az A véges automata a q_0 kezdőállapotban van és az olvasófej az input szalagon levő $u \in T^*$ szó első betűjét dolgozza fel (olvassa el, a megfelelő cellán áll).

Ezután a véges automata lépések sorozatát végrehajtva elolvassa az input u szót; betűről betűre haladva olvas és új állapotba kerül.

Véges automata - működési mód

Miután az u input szó utolsó betűjét is elolvasta a véges automata, vagy $q \in F$, azaz, elfogadó állapotba kerül, és akkor az u szót az automata elfogadja, vagy az új állapot nem lesz eleme F-nek, és ekkor az automata a szót nem fogadja el (visszautasítja).

Vegyük észre, hogy a fenti definíció alapján véges hosszúságú szót olvasva a véges automata sohasem fordul végtelen ciklusba, mivel minden egyes lépésben egy új szimbólumot kell elolvasnia. Ebből az adódik, hogy minden u szó esetén $|u|$ lépésben egyértelmű választ ad arra, hogy elfogadja-e a szót vagy visszautasítja.

Példa

Legyen

$$A = (Q, T, \delta, q_0, F)$$

véges automata, ahol

$$Q = \{q_0, q_1, q_2, q_3\},\ \ T = \{a, b\},\ \ F = \{q_0\}$$

és legyen

$$\begin{array}{ll} \delta(q_0, a) = q_2, & \delta(q_0, b) = q_1, \\ \delta(q_1, a) = q_3, & \delta(q_1, b) = q_0, \\ \delta(q_2, a) = q_0, & \delta(q_2, b) = q_3, \\ \delta(q_3, a) = q_1, & \delta(q_3, b) = q_2. \end{array}$$

Az A véges automata pontosan azokat a szavakat fogadja el, amelyek páros számú a betűt és páros számú b betűt tartalmaznak.

Példa - folytatás

$$\begin{array}{ll} \delta(q_0, a) = q_2, & \delta(q_0, b) = q_1, \\ \delta(q_1, a) = q_3, & \delta(q_1, b) = q_0, \\ \delta(q_2, a) = q_0, & \delta(q_2, b) = q_3, \\ \delta(q_3, a) = q_1, & \delta(q_3, b) = q_2. \end{array}$$

Tekintsük a $bbabab$ input szót.

Az állapotok sorozata akkor q_0, q_1, q_0, q_2, q_3, q_1, q_0.

Determinisztikus véges automata

A δ függvény egyértékű, ezért minden egyes (q, a) párra, ahol $(q, a) \in Q \times T$ egyetlen olyan p állapot létezik, amelyre $\delta(q, a) = p$ teljesül. Ezért ezt a véges automatát **determinisztikusnak** nevezzük.

Nemdeterminisztikus véges automata

Ha többértékű állapot-átmeneti függvényt is megengedünk, azaz δ a $Q \times T$ halmazból a 2^Q halmazba való leképezés, akkor **nemdeterminisztikus** véges automatáról beszélünk. Ebben az esetben aktuális állapotnak egy állapothalmaz valamely elemét, mintsem egyetlen állapotot tekinthetünk.

Ez azt jelenti, hogy a kezdeti állapot helyettesíthető egy $Q_0 \subseteq Q$ kezdőállapothalmazzal (kezdeti állapothalmazzal).

Az is előfordulhat, hogy egy a input szimbólum esetén $\delta(q, a)$ üres az aktuális állapotok mindegyikére.

Alternatív jelölés

Legyen $A = (Q, T, \delta, q_0, F)$ (determinisztikus) véges automata vagy nemdeterminsiztikus

Az A automata állapot-átmeneteit

$$qa \to p$$

alakú szabályok formájában is írhatjuk $p \in \delta(q, a)$ esetén.

Jelöljük M_δ-val az $A = (Q, T, \delta, Q_0, F)$ nemdeterminisztikus véges automata δ állapot-átmenet függvénye által az előbbi módon származó szabályok halmazát.

Ha minden egyes (q, a) párra egyetlen $qa \to p$ szabály van M_δ-ban, akkor a véges automata determinisztikus, egyébként nemdeterminisztikus.

Közvetlen redukció

Definíció

Legyen $A = (Q, T, \delta, q_0, F)$ egy véges automata és legyenek $u, v \in QT^*$ szavak (ahol QT^* a Q és T^* konkatenációját jelöli).

Azt mondjuk, hogy az A automata az u szót a v szóra redukálja egy lépésben vagy közvetlenül ($u \Longrightarrow_A v$), ha van olyan $qa \to p$ szabály M_δ-ban (azaz, $\delta(q, a) = p$) és van olyan $w \in T^*$ szó, amelyre $u = qaw$ és $v = pw$ teljesül.

Redukció

Definíció

Az $A = (Q, T, \delta, q_0, F)$ véges automata az $u \in QT^*$ szót a $v \in QT^*$ szóra redukálja ($u \Longrightarrow_A^* v$), ha vagy $u = v$, vagy van olyan $z \in QT^*$, amelyre $u \Longrightarrow_A^* z$ és $z \Longrightarrow_A v$ teljesül.

A $\Longrightarrow_A$ reláció, valamint tranzitív és reflexív lezártja, $\Longrightarrow_A^*$, lényegében a grammatikák elméletéből ismert levezetésnek felel meg.

Az automata által elfogadott nyelv

Az $A = (Q, T, \delta, q_0, F)$ véges automata által elfogadott (vagy másképpen felismert) nyelv alatt az

$$L(A) = \{u \in T^* | q_0 u \Longrightarrow_A^* p,\ q_0 \in Q_0 \text{ és } p \in F\}$$

szavak halmazát értjük.

Az üres szó, ε, akkor és csak akkor van benne az automata által elfogadott $L(A)$ nyelvben, ha $Q_0 \cap F \neq \emptyset$.

Minden $qa \rightarrow p \in M_\delta$ $(q, p \in Q, a \in T, \delta(q, a) = p)$ állapot-átmeneti szabály hossz-csökkentő, de könnyen belátható, hogy a redukció megfelel egy bal-lineáris grammatikabeli levezetés megfordítottjának.

Tétel

Minden A nemdeterminisztikus véges automatához meg tudunk adni egy 3-típusú G grammatikát úgy, hogy $L(G) = L(A)$ teljesül.

Bizonyításvázlat

Először megmutatjuk, hogy $L(A) \subseteq L(G)$.

Legyen $A = (Q, T, \delta, Q_0, F)$ egy nemdeterminisztikus véges automata. Definiáljuk a $G = (N, T, P, S)$ grammatikát úgy, hogy $N = Q \cup \{S\}$ és legyen

1. $p \to a \in P$ akkor és csak akkor, ha $q_0 a \to p \in M_\delta$ valamely $q_0 \in Q_0$-ra,

2. $p \to qa \in P$ akkor és csak akkor, ha $qa \to p \in M_\delta$,

3. $S \to p \in P$ akkor és csak akkor, ha $p \in F$,

4. $S \to \varepsilon \in P$ akkor és csak akkor, ha $Q_0 \cap F \neq \emptyset$.

Bizonyításvázlat - folytatás

1. Nyilvánvaló, hogy $\varepsilon \in L(G)$ akkor és csak akkor, ha $\varepsilon \in L(A)$.

2. Tegyük fel, hogy $u \neq \varepsilon \in L(A)$. Akkor van olyan $q_0 \in Q_0$ kezdőállapot és olyan $p \in F$ elfogadó állapot, hogy $q_0 u \Longrightarrow_A^* p$ teljesül. Tekintve ennek a redukciónak a megfordítását, a 2. csoportba tartozó szabályok használatával meg tudjuk konstruálni a $p \Longrightarrow_G^* q_0 u$ levezetést.

3. A G-beli levezetés utolsó lépése $p_1 \to q_0 a$ alakú szabály alkalmazása lenne, ahol $q_0 \in Q_0$. Ezért az 1. szabálycsoportban van $p_1 \to a$ alakú szabály. Ebből az adódik, hogy $p \Longrightarrow_G^* u$.

4. Végezetül, tekintsük az $S \to p$ szabályt a 3. szabálycsoportból és így az $S \Longrightarrow_G p \Longrightarrow_G^* u$ levezetéshez jutunk.

Vagyis, $L(A) \subseteq L(G)$.

Bizonyításvázlat - folytatás

Megmutatjuk, hogy $L(G) \subseteq L(A)$.

1. Legyen $u \in L(G)$ és $u \neq \varepsilon$, azaz legyen $S \Longrightarrow_G^* u$, ahol $u \in T^+$.

2. A G konstrukciója alapján akkor létezik a

$$S \Longrightarrow_G p \Longrightarrow_G^* p_1 v \Longrightarrow_G av = u$$

levezetés, ahol $p \in F$ és $p_1 \to a \in P$.

3. Ebből az következik, hogy $q_0 u \Longrightarrow_A^* u$, így $u \in L(A)$.

A G grammatika szabályai bal-lineárisak, de ismeretes, hogy minden bal-lineáris grammatikához létezik vele ekvivalens jobb-lineáris grammatika, így a tétel állítása fennáll.

Tétel

Minden 3-típusú G grammatikához meg tudunk adni egy A véges automatát úgy, hogy $L(A) = L(G)$ teljesül.

Bizonyításvázlat

1. Az általánosság megszorítása nélkül feltehetjük, hogy $G = (N, T, P, S)$ normálformában van (minden szabály vagy $X \to aY$, vagy $X \to \varepsilon$ alakú, ahol $X, Y \in N$ és $a \in T$).

2. Legyen $A = (Q, T, \delta, Q_0, F)$ véges automata úgy, hogy $Q = N$, $Q_0 = \{S\}$ és $F = \{Z \in N \mid Z \to \varepsilon \in P\}$. Legyen M_δ úgy definiálva, hogy

$$Xa \to Y \in M_\delta \text{ akkor és csak akkor, ha } X \to aY \in P.$$

3. Könnyen belátható, hogy az $S \Longrightarrow_G^* u$ G-beli levezetésre van A-ban olyan redukció, amelyre $Su \Longrightarrow_A^* Z$ teljesül, ahol $Z \in F$.

4. Megfordítva, minden A-beli előbbi alakú redukcióhoz tudunk egy megfelelő levezetést találni G-ben. Azaz, az állítás fennáll.