DIE GRÖSSE DES ZUSTANDSMINIMALEN LR(O)-ANALYSATORS

R.Kemp,Universität des Saarlandes,Saarbrücken

Die Analyse eines Wortes w,welches mittels einer LR(k)-Grammatik G (/6/) erzeugbar ist,läßt sich bekanntlich in linearer Zeit mittels eines deterministischen "push-down-transducer" DPDT(a) (<u>LR(k)-Analysator</u>) durchführen,dessen Schaltwerk ein endlicher Automat a (<u>LR(k)-Automat</u>) ist,welcher die Menge aller Anfangswörter bis zur ersten Reduktion (Handle) und die folgenden k Terminalzeichen erkennt. Dabei treten die Endzustände (Reduce-Zustände) im LR(k)-Automaten getrennt auf,d.h.es fallen keine Endzustände in Zwischenzustände (Shift-Zustände) und keine Endzustände zusammen.Im allgemeinen geht allerdings die rasche Analysezeit auf Kosten aufwendiger Analysatoren,d.h. die Zustandsmenge des LR(k)-Automaten wird sehr groß.(vergl./3/).Aus diesem Grund ist man an Verfahren interessiert,welche die Zustandsmenge vermindern,ohne auf eine korrekte Fehlererkennung bei der Analyse zu verzichten.

In der vorliegenden Arbeit werden zunächst zwei Typen von LR(k)-Automaten definiert: Der <u>"A-minimale LR(k)-Automat"</u> und der <u>"$\mathfrak{z}$-minimale LR(k)-Automat"</u>.Beide definieren einen LR(k)-Analysator,welcher genau die von der gegebenen LR(k)-Grammatik G erzeugte Sprache $\mathfrak{L}(G)$ analysiert.Der A-minimale LR(k)-Analysator,welcher als Schaltwerk den A-minimalen LR(k)-Automaten besitzt,hat die Eigenschaft,daß die Entscheidung,ob das Wort in der erzeugten Sprache liegt oder nicht,in kürzester Zeit getroffen wird.Der $\mathfrak{z}$-minimale LR(k)-Analysator mit dem $\mathfrak{z}$-minimalen LR(k)-Automaten als Schaltwerk trifft die gleiche Entscheidung,allerdings mit stärkerer Verzögerung , besitzt dagegen aber die geringste Anzahl von Zuständen unter allen Analysatoren,in deren Schaltwerk die Endzustände getrennt auftreten und die zu vorgegebenem Wort aus der Sprache $\mathfrak{L}(G)$ den zugehörigen Ableitungsbaum bzgl.G liefern.Dabei arbeiten beide Analysatoren auf Wörtern der Sprache gleich schnell.Im wesentlichen stellt sich die Konstruktion des $\mathfrak{z}$-minimalen LR(k)-Automaten aus dem A-minimalen LR(k)-Automaten als Minimierung spezieller partieller endlicher Automaten dar.Als bemerkenswertes Ergebnis erhält man,daß i.a.auch die Mächtigkeit der Zustandsmenge des $\mathfrak{z}$-minimalen LR(O)-Automaten exponentiell mit der Mächtigkeit des Hilfsalphabetes der gegebenen LR(O)-Grammatik wächst.

1. NOTATIONEN - DEFINITIONEN

Definition 1

Ein 4-tupel $G=(I,T,P,\sigma)$ heißt kontextfreie Grammatik (CFG) über dem Hilfsalphabet I, dem Endalphabet T, mit dem Produktionssystem P und dem Axiom σ ($\notin I$). Jede Regel $f_i \in P, 1 \leq i \leq \#P$, ist von der Gestalt:

$$f_i : Q(f_i) \longrightarrow Z(f_i)$$

mit $Q(f_1)=\sigma, Q(f_j) \in I, Z(f_i) \in (I \cup T)^*, 1 \leq i \leq \#P, j \geq 2$. $Q(f)$ heißt Quelle der Regel f und $Z(f)$ heißt Ziel der Regel f.

Bemerkung 1

Wie in /1/ beschrieben, läßt sich jeder CFG eine freie X-Kategorie $\mathcal{F}(P,I\cup T)$ mit der Morphismenmenge $MOR(\mathcal{F}(P,I\cup T))$ und dem freien Erzeugendensystem P zuordnen. Jedem aus dem Axiom σ mittels der $f \in P$ ableitbaren Wort w entspricht damit ein Morphismus $g \in MOR(\mathcal{F}(P,I\cup T))$ mit $Q(g)=\sigma$ und $Z(g)=w$, wobei $Q(g)$ bzw. $Z(g)$ Quelle und Ziel von g darstellen.

Definition 2

Ist $G=(I,T,P,\sigma)$ eine CFG, dann ist

$$\mathcal{T}(G) := \left\{ w \in (I\cup T)^* \;\middle|\; (\exists g \in MOR(\mathcal{F}(P,I\cup T)))(Q(g)=\sigma \wedge Z(g)=w \right\}$$

die Menge aller aus dem Axiom σ ableitbaren Satzformen. Die Menge $\mathcal{L}(G) := \mathcal{T}(G) \cap T^*$ ist die von der CFG erzeugte kontextfreie Sprache.

Definition 3

Eine CFG $G=(I,T,P,\sigma)$ heißt chomskyreduziert (CCFG), falls gilt:

(i) $(\forall A \in I)(\exists f \in P)(Q(f)=A \wedge Z(f) \in T^*)$

(ii) $(\forall A \in I)(\exists g \in MOR(\mathcal{F}(P,I\cup T)) \wedge \exists u,v \in (I\cup T)^*)(Q(g)=\sigma \wedge Z(g)=uAv)$

Definition 4

Ist $G=(I,T,P,\sigma)$ eine CCFG und $g \in MOR(\mathcal{F}(P,I\cup T))$, dann heißt

$$g=(1_{u_m} \times f_m \times 1_{v_m}) \circ \ldots \circ (1_{u_1} \times f_1 \times 1_{v_1})$$

mit $u_i, v_i \in (I \cup T)^*, f_i \in P, 1 \leq i \leq m$, sequentielle Darstellung von g. Besitzt g genau eine sequentielle Darstellung, dann ist g total unzerlegbar. Eine sequentielle Darstellung heißt antikanonisch, wenn $l(v_i) < l(v_{i+1})+1$, $1 \leq i \leq m-1$, gilt. Dabei ist $l(w)$ die Länge des Wortes w. Wenn es zu g eine antikanonische Darstellung obiger Form mit $Q(g)=\sigma$ gibt, dann heißt $(1_{u_m} \times f_m \times 1_{v_m})$ H-Faktor von g.

Definition 5

Ist $G=(I,T,P,\sigma)$ eine CCFG, dann sei mit $f \in P$ die Menge $R(f)$ wie folgt definiert:

$$R(f) := \left\{ wZ(f) \,\middle|\, (\exists v \in (I \cup T)^* \wedge \exists g \in MOR(\mathfrak{F}(P, I \cup T))(Q(g) = \mathfrak{S} \wedge g \text{ total unzerlegbar} \wedge (1_w \times f \times 1_v) \text{ H-Faktor von } g) \right\}$$

Die Menge R(p) mit

$$R(P) := \bigcup_{f \in P} R(f)$$

heißt Menge der <u>regulären Anfangswörter</u>.

<u>Definition 6</u>

Ein <u>endlicher Automat</u> (EA) ist ein 7-tupel $a=(\mathcal{E},\mathcal{J},\mathcal{F},\alpha,\mathcal{O},\delta,\lambda)$ mit

(i) $\mathcal{E},\mathcal{J},\mathcal{F},\mathcal{O}$ sind endliche Mengen (<u>Eingabezeichen</u>,<u>Zwischenzustände</u>, <u>Endzustände</u>,<u>Ausgabezeichen</u>)

(ii) $\delta:\mathfrak{D} \longrightarrow (\mathcal{J} \cup \mathcal{F}), \mathfrak{D} \subseteq (\mathcal{J} \cup \mathcal{F}) \times \mathcal{E}$ (<u>Überführungsfunktion</u>)

(iii) $\lambda:\mathcal{H} \longrightarrow \mathcal{O}\quad, \mathcal{H} \subseteq (\mathcal{J} \cup \mathcal{F}) \times \mathcal{E}$ (<u>Ausgabefunktion</u>)

(iv) $\alpha \in \mathcal{J}$ (<u>Startzustand</u>)

Ein EA a heißt

<u>deterministisch</u> (DEA) $\Longleftrightarrow (\forall\, (s,x) \in \mathfrak{D})(\#\delta(s,x) \leqslant 1)$

<u>vollständig</u> (VEA) $\Longleftrightarrow \mathfrak{D} = \mathcal{H} = (\mathcal{J} \cup \mathcal{F}) \times \mathcal{E}$

<u>streng partiell</u> (SPEA) $\Longleftrightarrow \mathfrak{D} = \mathcal{H}$

<u>Bemerkung 2</u>

Ist a ein deterministischer VEA,dann bezeichnen wir mit $\delta(\alpha,w)$ den Zustand,in den man durch Lesen des Wortes w vom Startzustand α aus gelangt.

2. DER $\mathfrak{G}$-MINIMALE LR(0)-ANALYSATOR

Bei der Analyse eines Wortes w,welches mittels einer LR(0)-Grammatik G erzeugbar ist,liest man das Wort von <u>l</u>inks nach <u>r</u>echts und die Entscheidung,ob an der gerade gelesenen Stelle eine Reduktion f,d.h.eine Ersetzung von Z(f) durch Q(f),vorzunehmen ist oder nicht,kann eindeutig aus dem bereits gelesenen Teilwort getroffen werden.Da in der Menge der regulären Anfangswörter (1.Def.5) alle Anfangswörter bis zur ersten Reduktion aufgelistet sind,genügt bei der Analyse eines Wortes w die Kenntnis der Menge R(P).In /1/ wurde gezeigt,daß die Menge R(P) für jede CCFG als homomorphes Bild eines Standardereignisses regulär ist und damit von einem endlichen Automaten A erkannt wird.In /2/ wurde die Berechnung der Überführungsfunktion δ dieses deterministischen VEA (DVEA) $A=(I \cup T,\mathcal{J},\{ⓕ \mid f \in P\},\alpha,P \cup \{1,s\},\delta,\lambda)$ effizient gestaltet und darüberhinaus gezeigt,daß dieser VEA im automatentheoretischen Sinne reduziert und damit minimal ist.Ist $w=uxv \in (I \cup T)^*, uv \in (I \cup T)^*, x \in (I \cup T)$, dann gilt für die Überführungsfunktion δ und die Ausgabefunktion λ:

$$\delta : (\mathcal{J} \cup \mathcal{F}) \times (I \cup T) \longrightarrow \mathcal{J} \cup \mathcal{F}$$

$$\delta(Z,x) = \begin{cases} (f) & \text{falls } ux \in R(f), f \in P \\ Z' & \text{falls } (\exists y \in (I \cup T)^*)(uxy \in R(P)) \\ q & \text{sonst} \end{cases}$$

und

$$\lambda : (\mathcal{J} \cup \mathcal{F}) \times (I \cup T) \longrightarrow P \cup \{1, s\}$$

$$\lambda(Z,x) = \begin{cases} f & \text{falls } \delta(Z,x) = (f) \\ 1 & \text{falls } \delta(Z,x) \in \mathcal{J} \setminus \{q\} \\ s & \text{falls } \delta(Z,x) = q \end{cases}$$

Dabei ist q ein ausgezeichneter Zustand, in den man im Falle $w \notin \mathfrak{T}(G)$ gelangt.

Der Automat A besitzt bei vorliegender LR(0)-Grammatik die Eigenschaft, daß keine Endzustände zusammen- und keine Endzustände in Zwischenzustände fallen. Die Aequivalenz dieser Aussage mit dem Vorliegen einer LR(0)-Grammatik ist in /1/,/2/ gezeigt und gibt Anlaß zu folgender

Definition 7

Eine CCFG ist genau dann vom Typ LR(0), wenn für alle $f, f' \in P, f \neq f'$ gilt:

(i) $R(f) \cdot T^* \cap R(f') = \emptyset$ und (ii) $R(f) \cdot T^+ \cap R(f) = \emptyset$

Es gilt der in /5a/ bewiesene

SATZ 1

Ist G eine CCFG vom Typ LR(0), dann meldet der Analysator DPDT(A) mit dem endlichen Automaten A als Schaltwerk bei dem Eingabewort w genau dann einen Analyseabbruch, wenn w ein Anfangswort w_1 besitzt, zu dem keine Verlängerung v zu einem aus $\mathfrak{S}$ ableitbaren Wort existiert.

Damit besitzt der DPDT(A) die Eigenschaft, daß er zum frühest möglichen Zeitpunkt die Analyse abbricht. Wir bezeichnen ihn als A-minimalen LR(0)-Analysator der Sprache $\mathcal{L}(G)$. (Analysezeitminimal). Der Automat A heißt A-minimaler LR(0)-Automat.

3. REGULÄRE LR(0)-ÜBERDECKUNGEN

Definition 8

Ist $G=(I,T,P,\mathfrak{S})$ eine CCFG vom Typ LR(0) und A der zugehörige A-minimale LR(0)-Automat, welcher die reguläre Menge R(P) erkennt, dann heißt jedes Mengensystem $\hat{R}(P) := \{ \hat{R}(f) \mid f \in P \}$ vermöge

(i) $(\forall f \in P)(\hat{R}(f) \text{ regulär})$ (iii) $(\forall f \in P)(\hat{R}(f) \subseteq \hat{R}(f) \subseteq (I \cup T)^*)$

(ii) $(\forall f \in P)(\hat{R}(f) \cdot T^+ \cap \hat{R}(f) = \emptyset)$ (iv) $(\forall f, f' \in P)(f \neq f' \Longrightarrow \hat{R}(f') \cdot T^* \cap \hat{R}(f) = \emptyset)$

reguläre LR(0)-Überdeckung von R(P).

Es gilt der in /5a/ bewiesene

SATZ 2

Ist $G=(I,T,P,\sigma)$ eine CCFG vom Typ LR(0) und $\hat{R}(P)$ eine reguläre LR(0)-Überdeckung von $R(P)$,welche von dem DEA $\hat{A}=(I\cup T,\hat{\mathcal{Z}},\mathcal{F},\hat{\alpha},P\cup\{1,s\},\hat{\delta},\hat{\lambda})$ erkannt wird,dann stellt jeder DPDT($\hat{A}$) ein Analysator der Sprache $\mathcal{L}(G)$ dar.

Bemerkung 3

Liefert der A-minimale LR(0)-Automat A eine Reduktion f,dann ist diese auch anwendbar,d.h.der Automat gelangte durch Lesen von Z(f) in den Endzustand f.Liefert dagegen der Automat $\hat{A}$ eine Reduktion f,dann muß diese nicht notwendigerweise anwendbar sein,da $\hat{A}$ u.U.nur durch Lesen eines Endwortes von Z(f) in den Endzustand f gelangte.Aus diesem Grund müssen die zu löschenden Kellerzeichen bei einer Ausführung einer Reduktion f mittels des DPDT($\hat{A}$) mit den entsprechenden Zeichen in Z(f) auf Übereinstimmung verglichen werden.

4. DER $\mathcal{Z}$-MINIMALE LR(0)-ANALYSATOR

Ist A der A-minimale LR(0)-Analysator bzgl.der LR(0)-Grammatik G und $\hat{A}$ ein EA,welcher eine reguläre LR(0)-Überdeckung $\hat{R}(P)$ erkennt,dann bezeichnen wir mit $\underset{p}{A}$ bzw.$\underset{p}{\hat{A}}$ diejenigen SPEA,welche durch Elimination des ausgezeichneten Zustandes q aus A bzw.$\hat{A}$ entstehen,d.h. z.B.

$$\underset{p}{A} = (I\cup T,\mathcal{Z}\setminus\{q\},\mathcal{F},\alpha,P\cup\{1\},\delta/(\mathcal{Z}\setminus\{q\})x(I\cup T),\lambda/(\mathcal{Z}\setminus\{q\})x(I\cup T))$$

Falls keine Verwechslungen auftreten,setzen wir künftig wieder:

$$\underset{p}{A} = (I\cup T,\mathcal{Z},\mathcal{F},\alpha,P\cup\{1\},\delta,\lambda) \quad \text{und} \quad \underset{p}{\hat{A}} = (I\cup T,\hat{\mathcal{Z}},\mathcal{F},\hat{\alpha},P\cup\{1\},\hat{\delta},\hat{\lambda})$$

Es gilt der in /5a/ gezeigte

SATZ 3

Es existiert i.a. kein Automatenhomomorphismus Φ mit

$$\text{(i)}\ \Phi(A) = \hat{A} \qquad \text{bzw.} \qquad \text{(ii)}\ \Phi(\underset{p}{A}) = \underset{p}{\hat{A}}$$

Die Menge aller regulärer LR(0)-Überdeckungen von R(P) bezeichnen wir mit $\hat{\mathcal{R}}(P)$. $\hat{\mathfrak{A}}(P)$ ist die Menge aller EA $\underset{p}{\hat{A}}$,welche $\hat{R}(P)\in\hat{\mathcal{R}}(P)$ erkennen.

Wir geben nun die

Definition 9

Ist $G=(I,T,P,\sigma)$ eine CCFG vom Typ LR(0),dann heißt ein Analysator DPDT($\underset{p}{\bar{A}}$) vermöge

$$\#\bar{\mathcal{Z}} = \min_{\underset{p}{\hat{A}}\in\hat{\mathfrak{A}}(P)} \#\hat{\mathcal{Z}}$$

$\mathcal{Z}$-minimaler LR(0)-Analysator.(Zustandsminimal).Der Automat $\underset{p}{\bar{A}}$ heißt $\mathcal{Z}$-minimaler LR(0)-Automat.

Wie man leicht einsieht, erhält man jeden EA $\hat{A}_P \in \hat{\mathfrak{A}}(P)$ aus A_P durch geeignete Identifikation von Zuständen. Es dürfen solche Zustände $z, z' \in \mathfrak{I}$ nicht identifiziert werden, für die gilt:

$$(\exists v \in (I \cup T)^*)(\delta(\eta, v) = (f) \wedge \delta(\xi, v) \in (\mathfrak{I} \cup \mathfrak{F}) \setminus \{(f)\}), \quad \eta, \xi \in \{z, z'\}, \quad \eta \neq \xi$$

da ansonsten der Endzustand (f) in einen Zwischenzustand oder einen von (f) verschiedenen Endzustand fallen würde. Zustände, welche nicht identifizierbar sind, d.h. der angegebenen Bedingung genügen, bezeichnen wir künftig als <u>inkompatibel</u> (ansonsten: <u>kompatibel</u>). Ein Minimierungsalgorithmus zur Konstruktion des $\mathfrak{z}$-minimalen Analysators $\bar{A}_P$ ist in /5b/ angegeben.

5. DIE GRÖSSE DES $\mathfrak{z}$-MINIMALEN LR(O)-ANALYSATORS $\bar{A}_P$

Wir wollen nun einige Aussagen über die Mächtigkeit der Zustandsmenge $\bar{\mathfrak{I}} \cup \mathfrak{F}$ machen. Hierzu geben wir die

<u>Definition 10</u>

Ist $G=(I,T,P,\sigma)$ eine CCFG, dann ist

$$F(f,X) := \{f \in P \mid Z(f) \in (I \cup T)^* X, \; X \in (I \cup T)\}$$

die Menge der <u>X-finalen Regeln</u> und

$$M(f,X) := \{f \in P \mid Z(f) \in (I \cup T)^* \cdot X \cdot (I \cup T)^+ \wedge (\exists g \in P)(g \in F(f,X))\}$$

die Menge des <u>X-medialen Regeln.</u>

Es gilt der

$\boxed{\text{SATZ 4}}$

Ist $G=(I,T,P,\sigma)$ eine CCFG vom Typ LR(O), dann gilt für die Mächtigkeit der Zustandsmenge $\bar{\mathfrak{I}} \cup \mathfrak{F}$ eines $\mathfrak{z}$-minimalen LR(O)-Automaten $\bar{A}_P$:

$$\#(\mathfrak{I} \cup \mathfrak{F}) \geqslant \#P + \underset{x \in (I \cup T)}{\mathrm{MAX}} (\# F(f,X) - \delta_{0, \# M(f,X)} + 1)$$

Dabei ist $\delta_{i,k}$ das "Kronecker Symbol".

<u>Beweisskizze</u>:

Betrachte im A-minimalen LR(O)-Automaten A_P alle Zustände, in die man durch Lesen eines $X \in (I \cup T)$ gelangt. Seien diese o.B.d.A. $(f_1), \ldots (f_k), s_1, \ldots, s_m$ mit $(f_i) \in \mathfrak{F}$ und $s_j \in \mathfrak{I}, 1 \leq i \leq k, 1 \leqslant j \leqslant m$. Es gilt:

a) Jedes $(f_i) \in \mathfrak{F}$ rührt von einer X-finalen Regel

b) Jedes $s_j \in \mathfrak{I}$ rührt von einer X-medialen Regel

Für jedes $(f_i) \in \mathfrak{F}$ muß im Automaten $\bar{A}_P$ mindestens ein Zwischenzustand $\bar{z}_i \in \bar{\mathfrak{I}}$ mit $\bar{\delta}(\bar{z}_i, X) = (f_i)$ existieren, d.h. mindestens $\# F(f,X)$ Zustände. Ungünstigenfalls sind die Zustände $s_j \in \mathfrak{I}$ paarweise kompatibel und werden in $\bar{A}_P$ zu einem Zustand $\bar{s} \in \bar{\mathfrak{I}}$ identifiziert, was einen weiteren Zustand s mit $\bar{\delta}(s,X) = \bar{s}$ erfordert. Damit folgt unmittelbar die Behauptung.

Eine unmittelbare Folgerung von SATZ 4 ist das

KOROLLAR

Ist $G=(I,T,P,\sigma)$ eine CCFG vom Typ LR(0) und endet jedes $Z(f), f \in P$, mit einem anderen Buchstaben $X \in (I \cup T)$, welcher nicht als Teilwort eines $Z(f'), f' \in P$, auftritt, dann gilt:

$$\# \mathcal{J} \cup \mathcal{F} \geq \# P + 1$$

●

Definition 11

Sei $G=(I,T,P,\sigma)$ eine CCFG und $\varphi \in \mathfrak{P}(I)$ eine Hilfsalphabetkombination (HK). Wir definieren folgende Mengen:

$\{f\}_A := \{f \in P \mid Q(f)=A\}$ — Bezeichnung: Regelpaket (RP)

$\{f\}_\varphi := \bigcup_{A \in \varphi} \{f\}_A$ — Bezeichnung: Regelpaketkombination (RPK)

$V_A := \{w \in (I \cup T)^+ \mid (\exists f \in \{f\}_A)(Z(f)=wu \wedge u \in (I \cup T)^*)\}$ — Bezeichnung: Zielmenge (ZM)

$V_\varphi := \bigcap_{A \in \varphi} V_A$ — Bezeichnung: Zielmenge der Regelpaketkombination $\{f\}_\varphi$ (ZMRPK)

Wie in /2/,/3/ gezeigt, werden bei der Konstruktion des A-minimalen LR(0)-Automaten sukzessive HK erzeugt, deren Gesamtheit wir mit $\mathbb{T}$ bezeichnen. Jedes $\varphi \in \mathbb{T}$ erzeugt gerade $\# V_\varphi$ Zustände. Mit $i \in \mathbb{N}$ sei

$$\mathbb{T}_i := \{\varphi \in \mathbb{T} \mid \# \varphi = i\}$$

Es gilt nun der /3/,/5b/ bewiesene

SATZ 5

Ist $G=(I,T,P,\sigma)$ eine CCFG vom Typ LR(0), dann gilt für die Zustandsmenge $\mathcal{J} \cup \mathcal{F}$ des A-minimalen LR(0)-Automaten $\underset{p}{A}$:

$$\# \mathcal{J} \cup \mathcal{F} = 1 + \sum_{\varphi \in \mathbb{T}} \# V_\varphi + \sum_{A \in I \setminus \mathbb{T}_1} \Big[\# V_A - \# \bigcup_{\varphi \in \mathcal{H}_A} V_\varphi \Big] \leq$$

$$\leq 1 + \sum_{A \in I} \# V_A + \sum_{i=1}^{m} \# \varphi(x_i) \leq 1 + \sum_{A \in I} \# V_A + m \cdot (2^{\# I - 1} - \# I)$$

mit $\mathcal{H}_A := (\mathbb{T} \setminus \mathbb{T}_1) \cap (\mathfrak{P}(I) \setminus \mathfrak{P}(I \setminus \{A\}))$, $\varphi(x_i) := \{\varphi \in \mathbb{T} \setminus \mathbb{T}_1 \mid x_i \in V_\varphi\}$, $1 \leq i \leq m$

$$\bigcup_{\varphi \in \mathbb{T} \setminus \mathbb{T}_1} V_\varphi := \{x_i \mid 1 \leq i \leq m\}$$

●

Definition 12

Ist M eine endliche Menge und $N \subseteq \mathfrak{P}(M)$ mit

(i) $(\forall \varphi, \varphi' \in N)(\varphi \nsubseteq \varphi' \wedge \varphi' \nsubseteq \varphi)$

(ii) $(\forall \varphi, \varphi' \in N)(\exists \{A,B\})(\forall \varphi'' \in N)(A \in \varphi \setminus \varphi' \wedge B \in \varphi' \setminus \varphi \wedge \{A,B\} \nsubseteq \varphi'')$

dann heißt N reduziertes Spernersystem (RSS). Ist nur (i) erfüllt, dann bezeichnen wir N in Anlehnung an /7/ als Spernersystem (SS).

Mit

$$\Theta_{RSS} := \left\{ \Phi \subseteq \mathcal{P}(M) \mid \Phi \text{ ist RSS} \right\}$$

sei

$$F_{RSS}: \mathbb{N} \longrightarrow \mathbb{N}$$

$$F_{RSS}: \# M \longmapsto \underset{\Phi \in \Theta_{RSS}}{MAX} (\# \Phi)$$

Es gilt der

SATZ 6

Ist $G=(I,T,P,\sigma)$ eine CCFG vom Typ LR(0), dann gilt für die Zustandsmenge $\mathfrak{Z} \cup \mathfrak{F}$ eines $\mathfrak{Z}$-minimalen LR(0)-Automaten $\bar{A}_p$:

$$\# \bar{\mathfrak{Z}} \cup \mathfrak{F} \leq 1 + \sum_{A \in I} \# V_A + m \cdot F_{RSS}(\# I - 1) \qquad \text{mit } m = \# \bigcup_{\varphi \in \Pi \setminus \Pi_1} V_\varphi$$

Beweisskizze:

Nach SATZ 5 gilt: $\# \mathfrak{Z} \cup \mathfrak{F} \leq 1 + \sum_{A \in I} \# V_A + \sum_{i=1}^{m} \# \varphi(x_i)$

Betrachte für ein $i \in [1:m]$ die HK-en $\varphi(x_i) := \{\varphi_1, \varphi_2, \ldots, \varphi_k\}$ d.h. $x_i \in V_{\varphi_j}$, $1 \leq j \leq k$

a) o.B.d.A. sei $\varphi_1 \subseteq \varphi_2 \subseteq \ldots \subseteq \varphi_r, r < k$

$$\Longrightarrow (\forall A \in \bigcup_{i=1}^{r} V_{\varphi_i})(\exists f \in P)(Q(f)=A \wedge Z(f)=x_i w)$$

Diese Regeln erzeugen im Automaten A_p r verschiedene Zustände. Da $\varphi_i \subseteq \varphi_r$, $1 \leq i \leq r-1$ und $\varphi_r \in \Pi \Longrightarrow$ Die Zustände sind kompatibel und können zu einem Zustand identifiziert werden. Dies gilt für alle Inklusionsketten in $\varphi(x_i)$, so daß $\varphi(x_i)$ höchstens soviele Zustände in $\bar{A}_p$ erzeugt, wie SS über I möglich sind.

b) Nach den Identifizierungen in Teil a) ist $\varphi(x_i)$ zu $\psi(x_i)$ mit

$$(\forall \varphi, \varphi' \in \psi(x_i))(\varphi \nsubseteq \varphi' \wedge \varphi' \nsubseteq \varphi)$$

vermindert worden. Existiert in $\psi(x_i)$ eine Teilmenge η mit

$$(\exists \varphi, \varphi' \in \eta)(\forall \{A,B\})(\exists \varphi'' \in \eta)(A \in \varphi \setminus \varphi' \wedge B \in \varphi' \setminus \varphi \wedge \{A,B\} \subseteq \varphi'')$$

dann ist leicht einzusehen, daß die Zustände, welche durch φ, φ' erzeugt werden, kompatibel sind, da jedes $\{A,B\}$ mit $A \in \varphi \setminus \varphi' \wedge B \in \varphi' \setminus \varphi$ in einem $\varphi'' \in \eta$ auftritt. Sukzessive Anwendung dieses Verfahrens ergibt, daß $\varphi(x_i)$ bzgl. des $\mathfrak{Z}$-minimalen Automaten $\bar{A}_p$ höchstens soviele Zustände erzeugt, wie RSS über I möglich sind, d.h. $\psi(x_i)$ wird zu $\varrho(x_i)$ mit $\# \varrho(x_i) \leq F_{RSS}(\# I - 1)$ vermindert. Damit folgt die Beh. ●

Eine Untersuchung der Funktion F_{RSS} liefert den

SATZ 7

Ist M eine endliche Menge mit $\# M \geq 2$, dann gilt: $F_{RSS}(\# M) > \sqrt{2^{\# M}}$ ●

Bemerkung 4

Ist $G=(I,T,P,\sigma)$ eine CCFG und $\varphi = \{A_1, \ldots, A_s\} \in \mathcal{P}(I)$, dann sei

$$f_i : Q(f_i) \longrightarrow w\varphi = Z(f_i) \quad , 1 \leq i \leq s$$

die Menge der Produktionen

$$f_i : Q(f_i) \longrightarrow wA_i \qquad 1 \leq i \leq s$$

Es gilt der

$\boxed{\text{SATZ 8}}$

Die Mächtigkeit der Zustandsmenge des $\mathfrak{z}$-minimalen Automaten wächst i.a. exponentiell mit der Mächtigkeit des Hilfsalphabetes der zugrunde liegenden CCFG vom Typ LR(0).

Beweisskizze:

(a) Betrachte folgende Grammatikfolge $G_{2n} := (I_{2n}, T_{2n}, P_{2n}, \sigma)$ mit

$$I_{2n} := \{A_i \mid 1 \leq i \leq 2n\} \qquad T_{2n} := \{a\} \cup X \cup Y \quad \text{mit}$$

$$X := \{x_i \mid 1 \leq i \leq 2^n\}$$
$$Y := \{y_i \mid 1 \leq i \leq n\}$$
$$X \cap Y = \emptyset, \; a \notin X \cup Y$$

$$P_{2n} := \{f_i \mid 1 \leq i \leq n(2^n+5)\} \quad \text{mit}$$

$$\begin{array}{lll}
f_{\&n+i} & : \sigma \longrightarrow x_{\&+1}\,\varphi_{\&+1} & 0 \leq \& \leq 2^n-1,\; 1 \leq i \leq n \\
f_{n2^n+j} & : \sigma \longrightarrow y_j & 1 \leq j \leq n \\
f_{n(2^n+1)+2\beta-1} & : A_\beta \longrightarrow aA_\beta & 1 \leq \beta \leq 2n \\
f_{n(2^n+1)+2l} & : A_\beta \longrightarrow y_{\left[\frac{l+1}{2}\right]} & 1 \leq l \leq 2n
\end{array}$$

wobei gilt:

$$\varphi_{i+1} := M_{2^n-1+i} \quad \text{mit} \quad \begin{array}{l}
M_0 := \emptyset \\
M_{2^k+2\beta-1} := M_{2^{k-1}+\beta-1} \cup \{A_{2k-1}\} \\
M_{2^k+2\beta} := M_{2^{k-1}+\beta-1} \cup \{A_{2k}\} \\
0 \leq k \leq n,\; 0 \leq \beta \leq 2^{k-1}
\end{array}$$

Es gilt:

$$\mathcal{L}(G_{2n}) = Xa^*Y \cup Y$$

Man zeigt nun folgende Aussagen:

(b) Mit $\Phi := \varphi_1 \cup \varphi_2 \cup \ldots \cup \varphi_{2^n}$ und $y \in Y$ gilt:

$$(\forall \varphi_1, \varphi_2 \in \Phi)(\exists A_1 \in \varphi_1)(\exists A_2 \in \varphi_2)(A_1 \neq A_2 \Longrightarrow \{A_1 \to y,\; A_2 \to y\} \subseteq P_{2n})$$

(c) $\Pi = \{\sigma\} \cup \{\varphi_i \mid 1 \leq i \leq 2^n\}$

(d) Jede CCFG G_{2n} ist vom Typ LR(0).

(e) Alle Zustände im A-minimalen LR(0)-Automaten sind inkompatibel.

(f) Der A-minimale LR(0)-Automat und der $\mathfrak{z}$-minimale LR(0)-Automat sind isomorph.

(g) Mit SATZ 5 folgt: $\# \mathfrak{J} \cup \mathfrak{F} = (n+2)2^n + 5n+1 =$

$$= (\#I_{2n}/2 + 2) \cdot \sqrt{2^{\#I_{2n}}} + 5 \cdot \#I_{2n}/2 + 1$$

6. SCHLUSSBEMERKUNGEN

(a) Der $\mathfrak{Z}$-minimale LR(O)-Automat besitzt die gleiche Eigenschaft wie der A-minimale LR(O)-Automat,nämlich daß zur Reduktionsentscheidung keine Berechnung von "look-aheads" notwendig ist.Fordert man diese Eigenschaft der LR(k)-Invarianz nicht mehr,dann kann die Zustandsmenge noch weiter unter die Zustandsmenge des Automaten $\bar{A}_P$ vermindert werden.(vergl./4/,/5b/)

(b) Nach Bemerkung 3 wird die Anwendbarkeit einer Reduktion auf dem Kellerband des DPDT($\bar{A}_P$) getroffen.Der definierte Kompatibilitätsbegriff von Zuständen kann leicht dahingehend verallgemeinert werden, daß gewisse Endzustände im Automaten $\bar{A}_P$ zusammenfallen,ohne die Analyse über linear anwachsen zu lassen.Es dürfen solche Endzustände nicht zusammenfallen,denen Regeln $f,f'\in P, f\neq f'$,mit $Z(f)=wZ(f')$ entsprechen,da in diesem Fall die Entscheidung zwischen f und f'auf dem Kellerband i.a. nicht mehr getroffen werden kann.

(c) Eine Minimierung des A-minimalen LR(k)-Automaten A_P zum $\mathfrak{Z}$-minimalen LR(k)-Automaten mit $k > 1$,kann auf die gleiche Art und Weise wie für den Fall k=O durchgeführt werden.(vergl./5b/)

7. LITERATUR

/1/ G.Hotz–V.Claus "Automatentheorie und formale Sprachen" III.Formale Sprachen,BI.Hochschulskripten (1971), 823a

/2/ R.Kemp "LR(k)-Analysatoren", Dissertation Saarbrücken (1973)

/3/ R.Kemp "An Estimation of the Set of States of the minimal LR(O)-Acceptor" Automata,Languages and Programming(M.Nivat.ed) North Holland Publishing Co,Amsterdam (1973) 563-574

/4/ R.Kemp "Die Größe des minimalen Analysators einer kontextfreien Grammatik" Lecture Notes in Ec.and Math.Syst.,78,(1972) 99-1o6

/5a/,/5b/ R.Kemp "Minimierung von LR(k)-Analysatoren"(1,2) Berichte des Math.Inst.,Saarbrücken (A75/1,2)

/6/ D.Knuth "On the Translation of Languages from Left to Right" ,Inf.and Control 8,(1965),607-639

/7/ E.Sperner "Ein Satz über Untermengen einer endlichen Menge" ,Math.Zeitsch. 27 (1928) ,544-548