

Grundlagen der Theoretischen Informatik

4. Kellerautomaten und kontextfreie Sprachen (II)

11.06.2015

Viorica Sofronie-Stokkermans

e-mail: sofronie@uni-koblenz.de

Übersicht

1. Motivation
2. Terminologie
3. Endliche Automaten und reguläre Sprachen
4. Kellerautomaten und kontextfreie Sprachen
5. Turingmaschinen und rekursiv aufzählbare Sprachen
6. Berechenbarkeit, (Un-)Entscheidbarkeit
7. Komplexitätsklassen P und NP

Umformung von Grammatiken

Startsymbol nur links

Einfache Annahme:

Im folgenden soll für alle cf-Grammatiken gelten:

Das Startsymbol S kommt nie auf einer rechten Regelseite vor.

Umformung

Ist das bei einer Grammatik nicht gegeben, kann man es wie folgt erreichen:

- Führe ein neues Startsymbol S_{neu} ein
- Füge die Regel

$$S_{neu} \rightarrow S$$

hinzu.

Nutzlose Symbole

Nutzlose Symbole und Regeln: Intuition

- Variablen und Symbole, die vom Startsymbol aus unerreichbar sind.
- Variablen, von denen aus kein Terminalwort abgeleitet werden kann.
- Regeln, die solche Variablen und Symbole enthalten

Nutzlose Symbole

Definition ((co-)erreichbare, nutzlose Symbole)

Sei $G = (V, T, R, S)$ eine Grammatik.

Ein Symbol $x \in (V \cup T)$ heißt

erreichbar: Es gibt $\alpha, \beta \in (V \cup T)^*$: $S \Longrightarrow_G^* \alpha x \beta$

co-erreichbar: Es gibt $w \in T^*$: $x \Longrightarrow_G^* w$

nutzlos: x ist nicht erreichbar oder nicht co-erreichbar.

Nutzlose Symbole

Theorem (cf-Grammatik ohne nutzlose Symbole)

Ist $G = (V, T, R, S)$ eine cf-Grammatik mit $L(G) \neq \emptyset$,
dann existiert eine cf-Grammatik $G' = (V', T', R', S')$ mit:

- G' ist äquivalent zu G .
- Jedes $x \in (V \cup T)$ ist erreichbar und co-erreichbar.

Nutzlose Symbole

Theorem (cf-Grammatik ohne nutzlose Symbole)

Ist $G = (V, T, R, S)$ eine cf-Grammatik mit $L(G) \neq \emptyset$,
dann existiert eine cf-Grammatik $G' = (V', T', R', S')$ mit:

- G' ist äquivalent zu G .
- Jedes $x \in (V \cup T)$ ist erreichbar und co-erreichbar.

Beweis

Man kann G' aus G effektiv konstruieren:

- Wie im folgenden beschrieben, die nutzlosen Symbole bestimmen.
- Diese Symbole und alle Regeln, die sie enthalten, entfernen.

Nutzlose Symbole

Algorithmus zur Berechnung der co-erreichbaren Variablen

Input: Grammatik $G = (V, T, R, S)$

Output: co-erreichbare Variablen

Alt := \emptyset

Neu := $\{A \in V \mid \exists w \in T^* (A \rightarrow w \in R)\}$

while Alt \neq Neu

{

 Alt := Neu

 Neu := Alt $\cup \{A \in V \mid \exists \alpha \in (T \cup \text{Alt})^* (A \rightarrow \alpha \in R)\}$

}

output Neu

Nutzlose Symbole

Bestimmung einer Grammatik $G'' = (V'', T'', R'', S)$ nur mit diesen co-erreichbaren Variablen.

if $S \in \text{Neu}$ $/* S \text{ ist co-erreichbar } */$

{

$V'' := \text{Neu}$

$T'' := T$

$R'' = R \cap (V'' \times (V'' \cup T'')^*)$

}

else $/* L(G) = \emptyset */$

Nutzlose Symbole

Bestimmung der Grammatik G' ohne nutzlose Symbole:

$G' = (V', T', R', S')$ mit:

$$V' := \text{Neu2} \cap V''$$

$$T' = \text{Neu2} \cap T$$

$$R' = R'' \cap (V' \times (V' \cup T')^*)$$

$$S' = S$$

Damit gilt dann: $L(G') = L(G)$ und G' enthält keine nutzlosen Symbole.

Normalform für Regeln

Normalform für Regeln

Theorem (Normalform)

Zu jeder Grammatik G (beliebigen Typs) existiert eine äquivalente Grammatik G' , bei der für alle Regeln $P \rightarrow Q \in R'$ gilt:

- $Q \in V^*$ und P beliebig
- $Q \in T$ und $P \in V$

Für alle Typen außer den linearen hat G' denselben Typ wie G .

Normalform für Regeln

Beweis

Für jedes Terminal $t \in T$ erzeuge man eine neue Variable V_t .

- $V' = V \cup \{V_t \mid t \in T\}$
- R' entsteht aus R , indem für jede Regel $P \rightarrow Q \in R$ in Q alle Vorkommen eines Terminals t durch die zugehörige Variable V_t ersetzt werden. Außerdem enthält R' für jedes $t \in T$ eine neue Regel $V_t \rightarrow t$.

Also $L(G') = L(G)$,

und für alle Sprachklassen außer \mathbf{L}_3 hat G' denselben Typ wie G .

Elimination von ε -Regeln

Idee: Variablen, aus denen ε ableitbar ist, sollten eliminiert werden

Elimination von ε -Regeln

Idee: Variablen, aus denen ε ableitbar ist, sollten eliminiert werden

Definition (ε -Regel, nullbare Variablen)

Eine Regel der Form

$$P \rightarrow \varepsilon \quad (P \text{ eine Variable})$$

heißt ε -Regel.

Eine Variable A heißt **nullbar**,
falls

$$A \Longrightarrow^* \varepsilon$$

Elimination von ε -Regeln

Theorem (ε -Regeln sind eliminierbar)

Zu jeder cf-Grammatik G existiert eine äquivalente cf-Grammatik G'

- ohne ε -Regeln und nullbare Variablen, falls $\varepsilon \notin L(G)$,
- mit der einzigen ε -Regel $S \rightarrow \varepsilon$ und der einzigen nullbaren Variablen S , falls $\varepsilon \in L(G)$ und S das Startsymbol ist.

Elimination von ε -Regeln

Algorithmus zur Berechnung der nullbaren Variablen

Input: Grammatik $G = (V, T, R, S)$ S o.B.d.A. in keiner Regel rechts

Output: nullbare Variablen

$Alt := \emptyset$

$Neu := \{A \in V \mid A \rightarrow \varepsilon \in R\}$

while $Alt \neq Neu$

{ $Alt := Neu$

für alle $(P \rightarrow Q) \in R$ **do**

 { **if** $Q = A_1 \dots A_n$ **and** $A_i \in Neu$ für $1 \leq i \leq n$ **and** $P \notin Neu$,
 then $Neu := Neu \cup \{P\}$

 }

}

output Neu

Elimination von ε -Regeln

Beweis (Forts.)

Ausgangsgrammatik G habe die Normalform, bei der für jede Regel $P \rightarrow Q$:
 $Q \in V^*$ oder $Q \in T$.

Für jede Regel $P \rightarrow A_1 \dots A_n$ generiere alle möglichen Kombinationen

$$P \rightarrow \alpha_1 \dots \alpha_n$$

mit

- $\alpha_i \in \{\varepsilon, A_i\}$ falls A_i nullbar
- $\alpha_i = A_i$ falls A_i nicht nullbar

Dann

- Füge alle diese neuen Regeln zur Grammatik hinzu
- Entferne alle Regeln der Form $A \rightarrow \varepsilon$ mit $A \neq S$

Elimination von ε -Regeln

Beweis ((Forts.))

Zu zeigen:

Für die neue Grammatik G' gilt: $L(G') = L(G)$

Vorgehen:

- G hat die Normalform:

Für jede Regel $P \rightarrow Q$ gilt $Q \in V^*$ oder $Q \in T$.

- Wir beweisen die etwas stärkere Behauptung

für alle $A \in V$ für alle $w \in (V \cup T)^* - \{\varepsilon\}$

$$\left((A \Longrightarrow_G^* w) \quad \underline{\text{gdw}} \quad (A \Longrightarrow_{G'}^* w) \right),$$

- Daraus folgt sofort $L(G') = L(G)$.

Elimination von ε -Regeln

Beweis (Forts.)

” \Rightarrow ” Wir zeigen: Aus $A \Longrightarrow_G^* w$ folgt $A \Longrightarrow_{G'}^* w$ (Induktion über Länge einer Ableitung von A nach w in G).

Induktionsanfang: Länge = 0.

Dann ist $w = A$, und $A \Longrightarrow_{G'}^* A$ gilt immer.

Induktionsschritt: Es sei schon gezeigt: Wenn in G in n Schritten eine Ableitung $B \Longrightarrow_G^* u$ durchgeführt werden kann, dann folgt, dass in G' die Ableitung $B \Longrightarrow_{G'}^* u$ möglich ist.

Elimination von ε -Regeln

Beweis (Forts.)

Außerdem gelte in der Ausgangsgrammatik G : $A \Longrightarrow_G^* w \neq \varepsilon$ in $n + 1$ Schritten.

Dann gilt:

- $A \Longrightarrow_G w' \Longrightarrow_G^* w$,
- $w' = A_1 \dots A_\ell \Longrightarrow_G^* w_1 \dots w_\ell = w$,
- und es wird jeweils A_i zu w_i in höchstens n Schritten für geeignete $w', A_1, \dots, A_\ell, w_1, \dots, w_\ell$.
- Für $1 \leq i \leq \ell$ gilt:
 - Entweder $w_i \neq \varepsilon$ und $A_i \Longrightarrow_G^* w_i$
also (per Induktionsvoraussetzung) $A_i \Longrightarrow_{G'}^* w_i$
 - oder $w_i = \varepsilon$ und $A_i \Longrightarrow_G^* w_i$.

Elimination von ε -Regeln

Beweis (Forts.)

Fall 1: $w_i = \varepsilon$, A_i ist nullbar.

Dann gibt es in G' eine Regel $A \rightarrow A_1 \dots A_{i-1} A_{i+1} \dots A_\ell$ nach der obigen Konstruktionsvorschrift für G' , falls $A_1 \dots A_{i-1} A_{i+1} \dots A_\ell \neq \varepsilon$. Das ist der Fall, denn sonst hätten wir: $A \Longrightarrow w' = \varepsilon \Longrightarrow^* w = \varepsilon$ (aus nichts wird nichts), aber $w = \varepsilon$ ist ausgeschlossen.

Fall 2: $w_i \neq \varepsilon$. Dann gilt nach Induktionsvoraussetzung

$$A_i \Longrightarrow_{G'}^* w_i.$$

Elimination von ε -Regeln

Beweis (Forts.)

Wir haben also folgendes gezeigt:

Sei $I = \{i \in \{1 \dots \ell\} \mid w_i \neq \varepsilon\} \neq \emptyset$.

Dann gibt es in R' eine Regel $A \rightarrow A_{i_1} \dots A_{i_m}$ mit $I = \{i_1, \dots, i_m\}$, und die A_i sind so angeordnet wie in der ursprünglichen Regel $A \rightarrow A_1 \dots A_\ell$.

Mit dieser neuen Regel können wir w so ableiten:

$$A \Longrightarrow_{G'} A_{i_1} \dots A_{i_m} \Longrightarrow_{G'}^* w_{i_1} \dots w_{i_m} = w$$

Elimination von ε -Regeln

Beweis (Forts.)

" \Leftarrow " Wir zeigen: Aus $A \Longrightarrow_{G'}^* w$ folgt $A \Longrightarrow_G^* w$ (Induktion über Länge einer Ableitung von A nach w in G'):

Induktionsanfang: Länge = 0. Dann ist $w = A$, und $A \Longrightarrow_G^* A$ gilt immer.

Induktionsschritt: Es gelte für alle Ableitungen $A \Longrightarrow_{G'}^* w$ einer Länge von höchstens n , dass $A \Longrightarrow_G^* w$.

Ist $A \Longrightarrow_{G'}^* w$ eine Ableitung der Länge $n + 1$, so gibt es ein ℓ , Wörter w_1, \dots, w_ℓ und Variablen A_1, \dots, A_ℓ mit $A \Longrightarrow_{G'} A_1 \dots A_\ell \Longrightarrow_{G'}^* w = w_1 \dots w_\ell$. Es gilt jeweils $A_i \Longrightarrow_{G'}^* w_i$ in höchstens n Schritten, und $w_i \neq \varepsilon$.

Elimination von ε -Regeln

Beweis (Forts.)

Nach der Induktionsvoraussetzung folgt daraus:

- für die Originalgrammatik G gibt es Ableitungen $A_i \Longrightarrow_G^* w_i$
- damit gibt es auch eine Ableitung $A_1 \dots A_\ell \Longrightarrow_G^* w$.

Elimination von ε -Regeln

Beweis (Forts.)

Nach der Induktionsvoraussetzung folgt daraus:

- für die Originalgrammatik G gibt es Ableitungen $A_i \Longrightarrow_G^* w_i$
- damit gibt es auch eine Ableitung $A_1 \dots A_\ell \Longrightarrow_G^* w$.

Da es in G' eine Ableitung $A \Longrightarrow_{G'} A_1 \dots A_\ell$ gibt, gibt es in R' eine Regel $A \rightarrow A_1 \dots A_\ell$. Wie ist diese Regel aus R entstanden?

Elimination von ε -Regeln

Beweis (Forts.)

Nach der Induktionsvoraussetzung folgt daraus:

- für die Originalgrammatik G gibt es Ableitungen $A_i \Longrightarrow_G^* w_i$
- damit gibt es auch eine Ableitung $A_1 \dots A_\ell \Longrightarrow_G^* w$.

Da es in G' eine Ableitung $A \Longrightarrow_{G'} A_1 \dots A_\ell$ gibt, gibt es in R' eine Regel $A \rightarrow A_1 \dots A_\ell$. Wie ist diese Regel aus R entstanden?

Eine Regel in R' entsteht aus einer Regel in R , indem einige nullbare Variablen gestrichen werden. Es gab also in G nullbare Variablen B_1 bis B_m , so dass R die Regel

$$A \rightarrow A_1 \dots A_{\ell_1} B_1 A_{\ell_1+1} \dots A_{\ell_2} B_2 \dots A_m B_m A_{m+1} \dots A_\ell$$

enthält. (m kann auch 0 sein, dann war die Regel selbst schon in R .)

Elimination von ε -Regeln

Beweis (Forts.)

Also gilt in G :

$$\begin{aligned} A &\Longrightarrow_G A_1 \dots A_{l_1} B_1 A_{l_1+1} \dots A_{l_2} B_2 \dots A_m B_m A_{m+1} \dots A_\ell \\ &\Longrightarrow_G^* A_1 \dots A_{l_1} A_{l_1+1} \dots A_{l_2} \dots A_m A_{m+1} \dots A_\ell \Longrightarrow_G^* w \end{aligned}$$

da ja $B_i \Longrightarrow_G^* \varepsilon$ möglich ist. \square

Elimination von ε -Regeln: Beispiel

$R :$

$S \rightarrow ABD$

$A \rightarrow ED \mid BB$

$B \rightarrow AC \mid \varepsilon$

$C \rightarrow \varepsilon$

$D \rightarrow d$

$E \rightarrow e$

$R' :$

$S \rightarrow ABD \mid AD \mid BD \mid D$

$A \rightarrow ED \mid BB \mid B$

$B \rightarrow AC \mid A \mid C$

$D \rightarrow d$

$E \rightarrow e$

Elimination von ε -Regeln: Beispiel

$R :$	$R' :$
$S \rightarrow ABD$	$S \rightarrow ABD \mid AD \mid BD \mid D$
$A \rightarrow ED \mid BB$	$A \rightarrow ED \mid BB \mid B$
$B \rightarrow AC \mid \varepsilon$	$B \rightarrow AC \mid A \mid C$
$C \rightarrow \varepsilon$	
$D \rightarrow d$	$D \rightarrow d$
$E \rightarrow e$	$E \rightarrow e$

Für die Regelmengemenge R in der linken Spalte sind die Variablen A, B, C nullbar.

Elimination von ε -Regeln: Beispiel

$R :$	$R' :$
$S \rightarrow ABD$	$S \rightarrow ABD \mid AD \mid BD \mid D$
$A \rightarrow ED \mid BB$	$A \rightarrow ED \mid BB \mid B$
$B \rightarrow AC \mid \varepsilon$	$B \rightarrow AC \mid A \mid C$
$C \rightarrow \varepsilon$	
$D \rightarrow d$	$D \rightarrow d$
$E \rightarrow e$	$E \rightarrow e$

Für die Regelmenge R in der linken Spalte sind die Variablen A, B, C nullbar.

Der obige Algorithmus erzeugt aus R die rechts aufgeführte Regelmenge R' .

Elimination von ε -Regeln

Beobachtung

- Der Algorithmus lässt nutzlose Variablen zurück, die nicht in Prämissen auftauchen (und deshalb nicht co-erreichbar sind).
Hier: C .
- Der Algorithmus lässt nutzlose Regeln zurück.
Hier: $B \rightarrow AC \mid C$.