Automaten und Formale Sprachen

SoSe 2013 in Trier

Henning Fernau
Universität Trier
fernau@uni-trier.de

24. Juli 2013

- Organisatorisches
- Einführung
- Endliche Automaten und reguläre Sprachen
- **Kontextfreie Grammatiken und kontrextfreie Sprachen**
- Chomsky-Hierarchie

Kontextfreie Grammatiken und kontrextfreie Sprachen

1. Automaten mit unendlichem Speicher
2. Kontextfreie Grammatiken
3. Normalformen
4. **Nichtkontextfreie Sprachen**
5. Algorithmen für kontextfreie Grammatiken

Ein Beispiel zur Schulung der Intuition

Betrachte

$$L = \{a^n b^n c^n \mid n \in \mathbb{N}\}$$

Ein Kellerautomat für L müsste irgendwie die Länge des a-Blocks mit der des folgenden b-Blocks abgleichen.
Dazu muss er den Keller verwenden.
Danach ist der Keller aber leer, d.h., die Länge des c-Blocks kann nicht mehr mit der eines vorhergehenden a- oder b-Blocks abgegeglichen werden.

Vermutung: L ist nicht kontextfrei.

Ein **binärer Wurzelbaum** ist gegeben durch ein Tripel $B = (V, \phi, r)$ mit ausgezeichneter *Wurzel* $r \in V$ und einer *Vater-Abbildung* $\phi : V \setminus \{r\} \to V$ mit der Eigenschaft

$$\forall v \in V : \#\underbrace{\{u \in V \mid \phi(u) = v\}}_{\kappa(v):=} \leq 2$$

$\kappa(v)$ liefert also die *Kinder* von v.
Knoten v mit $\kappa(v) = \emptyset$ heißen *Blätter*.

Lemma: Der Ableitungsbaum eines jeden von einer kontextfreien Grammatik in Chomsky-Normalform akzeptierten Wortes kann als binärer Wurzelbaum aufgefasst werden.

Die *Höhe* eines binären Wurzelbaumes $B = (V, \phi, r)$ ist gegeben durch

$$h(B) = \max_{v \in V \setminus \{r\}} \{k \in \mathbb{N} \mid \phi^k(v) = r\}.$$

Sonderfall $h(B) = 0$ bedeutet: $B = (\{r\}, \phi, r)$ mit trivialem ϕ.

Lemma: Hat ein binärer Wurzelbaum B mehr als 2^h Blätter, so gilt $h(B) > h$.

Beweis: Wir zeigen die *Kontraposition* per Induktion:

Gilt $h(B) \leq h$, so hat $B = (V, \phi, r)$ höchstens 2^h viele Blätter.

$h = 0$ ist trivial.

Es gelte $h > 0$. Daher gilt $\kappa(r) \neq \emptyset$.

Betrachte $r' \in \kappa(r)$. $V' := \{v \in V \mid \exists k \in \mathbb{N} : \phi^k(v) = r'\}$.

ϕ' sei die Einschränkung von ϕ auf $V' \setminus \{r'\}$.

$B' = (V', \phi', r')$ ist ein binärer Wurzelbaum der Höhe höchstens $h - 1$.

Auf B' ist die Induktionshypothese anwendbar: B' hat höchstens 2^{h-1} viele Blätter.

$\rightsquigarrow$ B hat höchstens $\#\kappa(r) \cdot 2^{h-1} \leq 2^h$ viele Blätter. $\square$

Lemma: Ist $G = (\Sigma, N, R, S)$ eine kfG in Chomsky-Normalform und ist $w \in L(G)$ mit $\ell(w) > 2^{\#N}$, so gilt für jeden Ableitungsbaum von w bzgl. G, dass es einen Weg von S zu einem Blatt gibt, auf dem mehr als $\#N$ viele Nichtterminalzeichen ersetzt werden.

<u>Beweis:</u> Nach dem vorigen Lemma hat der unterliegende binäre Wurzelbaum eine Höhe größer als $\#N$. Es gibt also einen Weg in besagtem Ableitungsbaum von der mit S beschrifteten Wurzel zu einem Blatt, dem mehr als $\#N$ Regelanwendungen entspricht $\rightsquigarrow$ Behauptung. $\square$

Folgerung: Auf besagtem Weg von der Wurzel zum Blatt im Ableitungsbaum von w finden wir also nach dem Schubfachprinzip zwei Regelanwendungen $A \to v$ und $A \to u$ mit gleicher linker Seite.

Erweiterte Chomsky-Normalform

Satz: Jede $L \in$ **KF** lässt sich beschreiben durch eine kfG $G = (\Sigma, N, R, S)$ mit Regeln der Form $N \times ((NN) \cup (\Sigma))$; zusätzlich darf eine Regel $S \to \lambda$ existieren, wobei dann gefordert ist, dass S in keiner rechten Regelseite vorkommt.

Beweis: Wie vorher erklärt, gibt es kfG $G' = (\Sigma, N', R', S')$ in Chomsky-Normalform für $L(G) \setminus \{\lambda\}$.
Definiere $R'' = \{S \to w \mid S' \to w \in R'\}$ für ein neues $S \notin N'$ und
$R = R' \cup R'' \cup \{S \to \lambda \mid \lambda \in L\}$, $N = N' \cup \{S\}$.
$\rightsquigarrow G = (\Sigma, N, R, S)$ beschreibt L und hat die geforderten Eigenschaften. $\square$

Hinweis: Die vorige Folie hat auch für die erweiterte Chomsky-Normalform Gültigkeit.

Ein Pumping-Lemma für KF

Satz: Zu jeder kfS L gibt es eine Konstante $n > 0$, sodass jedes Wort $w \in L$ mit $\ell(w) \geq n$ als Konkatenation $w = uvxyz$ dargestellt werden kann mit geeigneten u, v, x, y, z mit folgenden Eigenschaften:

1. $\ell(v) > 0$ oder $\ell(y) > 0$;
2. $\ell(vxy) \leq n$;
3. $\forall i \geq 0 : uv^ixy^iz \in L$.

Beweis: L werde durch kfG $G = (\Sigma, N, R, S)$ in erweiterter Chomsky-Normalform beschrieben.
Es sei $n = 2^{\#N}$.
Gibt es kein $w \in L$ mit $\ell(w) \geq n$, so stimmt die Aussage leer.
Sonst wähle ein $w \in L$ mit $\ell(w) \geq n$.
Nach obiger Folgerung gibt es zwei Regelanwendungen $A \to u'$ und $A \to v'$ auf einem Weg in einem Ableitungsbaum von w.
$\rightsquigarrow$ Es gibt Linksableitung

$$S \overset{*}{\Rightarrow} uA\zeta \overset{*}{\Rightarrow} uvA\eta\zeta \overset{*}{\Rightarrow} uvx\eta\zeta \overset{*}{\Rightarrow} uvxy\zeta \overset{*}{\Rightarrow} uvxyz.$$

Wegen Chomsky-NF (keine Kettenregeln oder λ-Regeln) gilt $\ell(v) > 0$ oder $\ell(y) > 0$.
$\ell(vxy) \leq n$ ergibt sich aus dem Beweis der Folgerung sowie wegen Chomsky-NF.
Daher gibt es auch Ableitungen

$$S \overset{*}{\Rightarrow} uA\zeta \overset{*}{\Rightarrow} ux\zeta \overset{*}{\Rightarrow} uxz$$

und allgemeiner

$$S \to uA\zeta \overset{*}{\Rightarrow} uvA\eta\zeta \overset{*}{\Rightarrow} uvvA\eta\eta\zeta \overset{*}{\Rightarrow} uv^iA\eta^i\zeta \overset{*}{\Rightarrow} uv^ixy^iz.$$

$\square$

Ein Beispiel zur Anwendung des Pumping-Lemmas:

$$L = \{a^k b^k c^k \mid k \in \mathbb{N}\} \notin \textbf{KF}$$

Wäre L kontext-frei, so gäbe es Pumping-Konstante n.
Wähle $w = a^n b^n c^n$ als genügend langes Wort.
Diskutiere Zerlegungen $w = uvxyz$.

Wegen $\ell(vxy) \leq n$ enthält vxy höchstens a's gefolgt von b's oder b's gefolgt von c's.

Beide Fälle sind analog; diskutiere also den ersten.
“Nullpumpen” liefert ein Wort, in dem alle a- und b-Vorkommen zusammen weniger als die doppelte Anzahl von c's ausmachen (Widerspruch zur Wortstruktur).

Das Pumping-Lemma kennzeichnet nicht Kontextfreiheit

Betrachte

$$L = \{a^k d^r a^k d^s a^k \mid k, r, s \in \mathbb{N}\}$$

Mit der gleichen Intuition wie vorher ist die Sprache nicht kontextfrei.
L erfüllt aber das Pumping-Lemma:
Ein Wort der Form $w = a^k d^r a^k d^s a^k$ mit $r > 0$ lässt sich zerlegen in: $w = uvxyz$ mit (z.B.) $u = a^k$, $v = d$ und $y = \lambda$.
Dann gilt $uv^ixy^iz \in L$ für alle $i \in \mathbb{N}$.
Der Fall $r = 0$ aber $s > 0$ geht analog.
Gilt $r = s = 0$, so gilt für $w = a^{3k}$, und die Wahl $v = a^3$, $y = \lambda$ führt wiederum auf $uv^ixy^iz \in L$ für alle $i \in \mathbb{N}$.

Wie kann man die **Intuition retten** ?

Betrachte

$$L' = \{a^k d a^k d a^k \mid k \in \mathbb{N}\}$$

Jetzt ermöglicht Nullpumpen einen Widerspruch zur angenommenen Kontextfreiheit. (Übungsaufgabe !)

Idee: Verwende Abschlusseigenschaften,
hier Durchschnitt mit regulären Sprachen:

Wäre $L = \{a^k d^r a^k d^s a^k \mid k, r, s \in \mathbb{N}\} \in$ **KF**, so auch $L' = L \cap \{a\}^*\{d\}\{a\}^*\{d\}\{a\}^*$.

Noch eine Übungsaufgabe...

Sei $\#_x w =$ Anzahl der Vorkommen von x in w.

Zeigen Sie, dass

$$L = \{w \in \{a, b, c\}^* \mid \#_a w = \#_b w = \#_c w\}$$

nicht kontextfrei ist.

Zur rechten Wahl des Pumpwortes

Betrachte $L = \{tt \mid t \in \{a, b\}^*\}$.

Wäre $L \in$ **KF**, so gäbe es Pump-Konstante n für L.

Diskutiere $w = a^n b^n a^n b^n \in L$.
Für vxy mit $\ell(vxy) \leq n$ gibt es drei Unterfälle:
(1) $vxy \in \{a\}^*\{b\}^*$, $z \in \{b\}^*\{a^n b^n\}$.
(2) $vxy \in \{b\}^*\{a\}^*$.
(3) $vxy \in \{a\}^*\{b\}^*$, $u \in \{a^n b^n\}\{a\}^*$.

Wir betrachten (3) eingehend (die anderen Fälle sind ähnlich):
$uxz = a^n b^n a^m b^k \in L$ mit $m < n$ oder $k < n$ (Nullpumpen).
$uxz = ss$ mit $\ell(s) < 2n$.
Das “erste s” muss mit a anfangen, aber das zweite muss mit b beginnen
(wegen $\ell(s) < 2n$ und da uxz nicht nur aus a’s besteht).

Nachtrag zu Abschlusseigenschaften I

Satz: **KF** ist nicht unter Durchschnitt abgeschlossen.

Beweis: $L_1 = \{a^n b^n c^k \mid k, n \in \mathbb{N}\}$

$L_2 = \{a^n b^k c^k \mid k, n \in \mathbb{N}\}$

L_1 und L_2 sind kontextfrei.

$L_1 \cap L_2$ ist jedoch nicht kontextfrei (s.o.). $\square$

Nachtrag zu Abschlusseigenschaften II

Satz: **KF** ist nicht unter Komplementbildung abgeschlossen.

Beweis: Betrachte das Komplement L' von $L = \{a^k b^k c^k \mid k \in \mathbb{N}\}$.

$$L' = \{w \in \{a, b, c\}^* \mid w \notin \{a\}^*\{b\}^*\{c\}^* \vee (\forall k \in \mathbb{N} : w \neq a^k b^k c^k)\}.$$

Mit $L'' = \{w \in \{a\}^*\{b\}^*\{c\}^* \mid (\forall k \in \mathbb{N} : w \neq a^k b^k c^k)\}$ wäre auch L' kontextfrei, denn
$L' = L'' \cup \overline{\{a\}^*\{b\}^*\{c\}^*}$ (Vereinigungsabschluss von **KF** und Komplementabschluss von **REG**).

$L'' = \{a^k b^\ell c^m \mid k \neq \ell \vee k \neq m \vee \ell \neq m\} \in$ **KF**, denn
$$\begin{aligned} L'' = {} & \{a^k b^\ell c^m \mid k < \ell\} \cup \{a^k b^\ell c^m \mid k > \ell\} \\ & \cup \{a^k b^\ell c^m \mid k < m\} \cup \{a^k b^\ell c^m \mid k > m\} \\ & \cup \{a^k b^\ell c^m \mid \ell < m\} \cup \{a^k b^\ell c^m \mid \ell > m\}. \end{aligned}$$
Jede dieser Teilsprachen ist “offensichtlich” kontextfrei, und wegen des Vereinigungsabschlusses von **KF** gilt dies auch für L'' und somit für L'.

Wäre **KF** komplementabgeschlossen, so wäre mit L' auch $L \in$ **KF**, ein Widerspruch ! $\square$

Entscheidbarkeitsfragen I

Frage: Gibt es einen Algorithmus, sodass ...

Satz: Das Leerheitsproblem ist für kfG entscheidbar.

1. Beweis: Ersetze alle Terminalzeichenvorkommen in der Grammatik G durch das leere Wort; das liefert neue Grammatik G'. Dann gilt: $L(G) \neq \emptyset$ gdw. $L(G') = \{\lambda\}$. $\lambda \in L(G')$ kann man entscheiden (siehe letzter Foliensatz). $\square$

2. Beweis: o.E. G in erweiterter Chomsky-Normalform.
Die betreffende Pump-Konstante n_G garantiert: Ist $L(G) \neq \emptyset$, so gibt es $w \in L(G)$ mit $\ell(w) < n_G$ (Nullpumpen).
Daher kann man $L(G) = \emptyset$ durch Testen aller Wörter bis zur Länge n_G entscheiden. $\square$

Entscheidbarkeitsfragen II

Endlichkeitsproblem: Gegeben Sprachbeschreibung G für L, ist L endlich ?

Satz: Das Endlichkeitsproblem ist für kfG entscheidbar.

Beweis: Dies folgt wie in obigem 2. Beweis zum Leerheitsproblem aus:

Lemma: Zu jeder Sprache $L \in$ **KF** gibt es eine Konstante $n > 0$, sodass gilt: L ist unendlich gdw. es gibt ein Wort $t \in L$ mit $n \leq \ell(t) < 2n$.

Lemma: Zu jeder Sprache $L \in$ **KF** gibt es eine Konstante $n > 0$, sodass gilt: L ist unendlich gdw. es gibt ein Wort $t \in L$ mit $n \leq \ell(t) < 2n$.

Beweis: Es sei n die Pump-Konstante aus dem Pumping-Lemma.

Ist L unendlich, so gibt es insbesondere Wörter mit Mindestlänge $2n$ in L.

Wähle unter diesen ein Wort t' minimaler Länge.

Nach dem Pumping-Lemma können wir $t' = uvxyz$ schreiben mit $\ell(vxy) \leq n$.

$\rightsquigarrow$ Nullpumpen liefert $t = uxz$ mit $n \leq \ell(t) < 2n$, da t' minimal.

Ein Wort $t \in L$ mit $n \leq \ell(t) < 2n$ können wir "aufpumpen",

d.h., mit der Existenz solch eines t ist L unendlich. $\square$