Grundzüge der Theoretischen Informatik

28.1.2022

Markus Bläser

Universität des Saarlandes


TI-Schokolade für vergessene Aufnahmen.
Ritter SPORT
KNUSPERFLAKES
BUNTE VIELFALT
Ritter SPORT
NUGAT
BUNTE VIELFALT

Kapitel 29: Kontextfreie Grammatiken

Ableitungsbäume und Mehrdeutigkeit

$$E \to E * E \mid E + E \mid (E) \mid x$$

Beispiel: $x + x * x$

Linksableitung:

$$E \Rightarrow E + E \Rightarrow x + E \Rightarrow x + E * E \Rightarrow x + x * E \Rightarrow x + x * x$$

Rechtsableitung:

$$E \Rightarrow E + E \Rightarrow E + E * E \Rightarrow E + E * x \Rightarrow E + x * x \Rightarrow x + x * x$$

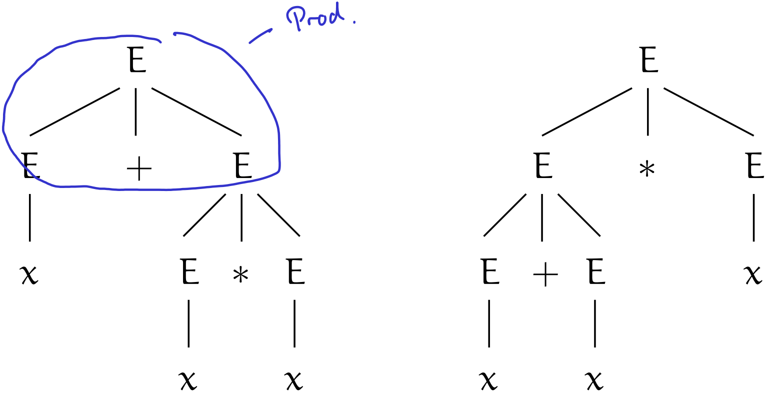
Ableitungsbäume

Definition (29.1)

Sei $G = (V, \Sigma, P, S)$ eine kontextfreie Grammatik.

1. Ein *Ableitungsbaum* ist ein geordneter Baum mit einer Knotenbeschriftung:
   1.1 Die Wurzel ist mit S beschriftet.
   1.2 Alle Blätter sind mit $V \cup \Sigma$ beschriftet oder mit ε. Im letzten Fall gibt es nur ein Blatt.
   1.3 Alle inneren Knoten sind mit Elementen aus V beschriftet. Falls A eine Knotenbeschriftung ist und $x_1, x_2, \ldots, x_t$ (in dieser Reihenfolge) die Beschriftungen der Kinder, dann ist $A \to x_1 x_2 \ldots x_t \in P$.
2. Das *Blattwort* ist die Konkatenation der Beschriftungen der Blätter.

Beispiel



Eindeutigkeit

Definition (29.2)

1. Eine kfG heißt *mehrdeutig*, falls ein Wort zwei oder mehr Ableitungsbäum hat. Sonst heißt sie *eindeutig*.
2. Eine kfS L heißt *eindeutig*, falls es eine eindeutige Grammatik für L gibt. Sonst ist Sie *inhärent mehrdeutig*.

$$
\begin{aligned}
E &\longrightarrow T \mid T + E \\
T &\longrightarrow F \mid F * T \\
F &\longrightarrow x \mid (E)
\end{aligned}
$$

Theorem (29.3)

$\{0^n 1^n 2^m 3^m \mid n, m \geq 1\} \cup \{0^n 1^m 2^m 3^n \mid n, m \geq 1\}$ *ist kontextfrei und inhärent mehrdeutig.*

Entfernen nutzloser Symbole

Kann man feststellen, ob Variablen wirklich gebraucht werden?

* nicht erreichbaren.

A heißt generierend, falls es ein $v \in \Sigma^*$ gibt mit $A \Rightarrow^* v$

A heißt erreichbar, falls es $x, y \in (V \cup \Sigma)^*$ mit $S \Rightarrow^* xAy$

A heißt nutzlos, falls es nicht generierend oder nicht erreichbar ist

Man kann nutzlose Symbole eliminieren, indem man erst alle nicht generierenden Var entfernt und dann alle *

Kapitel 30: Die Chomsky-Normalform

Chomsky-Normalform

$E \to T \mid T + E$

$A \to B$

Definition (30.5)

Eine kfG $G = (V, \Sigma, P, S)$ ist in *Chomsky-Normalform*, falls jede Produktion die Form

$$A \to BC$$

oder

$$A \to \sigma$$

hat mit $A, B, C \in V$ und $\sigma \in \Sigma$.

ε-Produktionen und kfGs

$A \to B \qquad B \to \varepsilon$

- Bei Typ-1-Grammatiken sind ε-Produktionen im Wesentlichen verboten und damit auch bei Typ-2-Grammatiken (kfGs).
- Während Typ-1-Grammatiken mit beliebigen ε-Produktionen so mächtig sind wie Typ-0-Grammatiken, erhöhen ε-Produktionen die Mächtigtkeit von kfGs *nicht*.
- Deswegen kann man beliebige ε-Produktionen bei kfGs zulassen.

Eliminierung von ε-Produktionen

$A \to BC$ $\quad B \to C$ $\quad \cancel{C \to \varepsilon}$

Definition (30.1)

Sei $G = (V, \Sigma, P, S)$ eine kfG. $A \in V$ heißt nullierbar, falls $A \Rightarrow^*_G \varepsilon$.

Theorem (30.2)

Sei $G = (V, \Sigma, Q, S)$ *eine kfG. Sei* $H = (V, \Sigma, Q, S)$ *definiert wie folgt:*

$V \cup \Sigma$

1. *Ersetze jedes* $A \to a_1 a_2 \ldots a_k$ *durch* 2^ℓ *Produktionen, eine für jede Möglichkeit eines der* a_{i_λ} *auszulassen, wobei* $a_{i_1}, \ldots, a_{i_\ell}$ *die nullierbaren Variablen aus* $a_1, a_2, \ldots, a_k$ *sind*
2. *Entferne alle* ε*-Produktionen.*

Es gilt $L(G) \setminus \{\varepsilon\} = L(H)$.

$A \to BC$ $\quad A \to C$

$A \to B$ $\quad \cancel{A \to \varepsilon}$

Beweis 30.1

Wir zeigen:

$\forall A \in V$ und $u \in (V \cup \Sigma)^*$:

$$A \Rightarrow_H^* u \iff A \Rightarrow_G^* u \text{ und } u \neq \varepsilon$$

Beweis über die Länge der Ableitung

Ind. anf.: $A \to u \in Q$. $u \neq \varepsilon$ nach Def. von H

Nach Konstruktion gibt eine Produktion

$A \to a_1 \dots a_k \in P$ und Indizes $1 \le j_1, \dots, j_t \le k$,

so dass: u ist die Konkatenation aller

a_i mit $i \notin \{j_1, \dots, j_t\}$ und jedes

a_{j_i}, $1 \le i \le t$, ist nullierbar.

Damit gilt:

$$A \Rightarrow_G a_1 \dots a_k \Rightarrow_G^* u$$

Induktionsschritt.

Falls $A \Rightarrow^* u$ und die Länge der Ableitung mindestens zwei ist, dann gibt es ein $w \in (V \cup \Sigma)^*$ so dass $A \Rightarrow_H^* w \Rightarrow_H u$

Das bedeutet, dass $w = xBz$ mit $x, z \in (V \cup \Sigma)^*$ so dass $u = xyz$ und $B \to y \in H$

Wie beim Induktionsanfang gilt $B \Rightarrow_G^* y$ und $y \neq \varepsilon$.

Nach Induktionsvor. $A \Rightarrow_G^* w$ und $w \neq \varepsilon$

$$\Rightarrow \quad A \Rightarrow_G^* w \Rightarrow_G^* u \,, \quad u \neq \varepsilon$$

$\Leftarrow$: Übung $\square$

Wenn es eine Produktion $A \to \varepsilon$ gibt, dann ist A nullierbar.

Wenn es eine Produktion $A \to a_1 \ldots a_k$ und a_1 bis a_k sind nullierbar, dann ist A nullierbar

Eliminierung von Kettenproduktionen

Definition (30.3)

Sei $G = (V, \Sigma, P, S)$ eine kfg: Eine Produktion $A \to B$ mit $A, B \in V$ heißt *Kettenproduktion*.

$\subseteq (V \cup \Sigma)^{+} \times (V \cup \Sigma)^{*}$

Theorem (30.4)

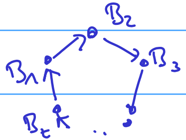
Sei $G = (V, \Sigma, P, S)$ *eine kfG, so dass* $H = (V, P \cap V \times V)$ *azyklisch ist. Dann gibt eine kfG* $G' = (V, \Sigma, P', S)$ *ohne Kettenproduktionen mit* $L(G) = L(G')$.

Baue einen Hilfsgraphen $H = (V, E)$

Variablen

$(A, B) \in E \iff A \to B \in P$ gerichtet!

H kann Kreise haben



Die Variablen innerhalb eines Kreises sind austauschbar!

Entferne daher $B_2, \dots, B_t$ und ersetze jedes Vorkommen durch B_1 (streiche $B_1 \to B_1$)

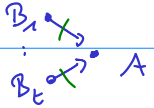
$B_2 \to C \quad \to \quad B_1 \to C$

Beweis 29.4.

Per Induktion über der Größe der Menge der Kanten von H

Ind.anf: $|E| = 0$ ✓

Ind.schritt, Da H azyklisch ist, gilt es mindestens einen Knoten, aus dem keine Kante herausgeht



Seien $A \to u_1, \ldots, A \to u_\ell$ alle Produktionen, bei denen A auf der linken Seite steht.

Terminale oder Länge ≥ 2

Ersetze $B_j \to A$ durch $B_j \to u_x$, $1 \leq x \leq \ell$ für alle $1 \leq j \leq t$.

Sei G" die resultierende Grammatik
Auf G" kann die Induktionsvor. angewendet
werden, da der entsprechende Hilfsgraph
weniger Kanten hat. $\rightarrow$ G'

„ $\square$ “

Die Chomsky-Normalform

Theorem (30.6)

Für jedes $G = (V, \Sigma, P, S)$ *mit* $L(G) \neq \emptyset$ *gibt es eine kfG* $G' = (V', \Sigma, P', S)$ *in Chomsky-Normalform mit* $L(G') = L(G) \setminus \{\varepsilon\}$.

G hat o.B.d.A keine ε-Prod. und keine Kettenproduktion.

$A \to BC$ $\qquad$ $A \to \sigma$

1. Schritt

Neue Variablen T_σ für jedes $\sigma \in \Sigma$.

Neue Prod $\boxed{T_\sigma \to \sigma}$

In allen anderen Produktionen ersetzen wir σ durch T_σ.

$\Rightarrow$ jede Produktion hat die Form

$A \to A_1 \dots A_t$ mit $A, A_1, \dots, A_t \in V$, $t \geq 2$

Wenn $t = 2$, dann ist konform mit der CNF.

2. Schritt:

Was passiert, wenn $t \geq 3$. ?

$$A \to A_1 \dots A_t$$

$$A \to A_1 C_2$$

$$C_2 \to A_2 C_3$$

$$C_3 \to A_3 C_4$$

$$\vdots$$

$$C_{t-2} \to A_{t-2} C_{t-1}$$

$$C_{t-1} \to A_{t-1} A_t$$

$$\cancel{C_2 \to A_2 \dots A_t}$$

$$\cancel{C_3 \to A_3 \dots A_t}$$

Für jede Prod. brauchen wir neue Hilfsvar.