Automata Theory :: Chomsky Normal Form

Jörg Endrullis

Vrije Universiteit Amsterdam

Lambda Rules and Erasable Variables

Lambda Rules and Erasable Variables

A production rule $A \to \lambda$ is called **λ-production rule.**

A variable A is called **erasable** if $A \Rightarrow^+ \lambda$.

The set of erasable variables can be computed as follows:

- If $A \to \lambda$, then A is erasable.
- If $A \to B_1 \cdots B_n$ and $B_1, \ldots, B_n$ are erasable, then so is A.

$$S \to AcB \qquad A \to CBC \qquad B \to abB \qquad C \to cCd$$
$$B \to \lambda \qquad C \to BB$$

We determine the set of erasable variables:

- B is erasable because of the rule $B \to \lambda$
- C is erasable because of $C \to BB$ and B is erasable
- A is erasable because of $A \to CBC$ and B, C are erasable

So the variables A, B, C are erasable.

Removal of Lambda Rules

Theorem

For every context-free language L there exists a context-free grammar G without λ-rules such that $L(G) = L \setminus \{\lambda\}$.

Construction

Let G be a context-free grammar with $L(G) = L$.

- Determine all erasable variables (that is, variables $A \Rightarrow^* \lambda$).
- For every rule $A \rightarrow xBy$ with $B \Rightarrow^* \lambda$, add a rule $A \rightarrow xy$.
- Remove all λ-production rules.

The resulting grammar G has the property $L(G) = L \setminus \{\lambda\}$.

Exercise

Consider the following grammar

$$S \rightarrow ABaC \qquad A \rightarrow BC \qquad B \rightarrow b \mid \lambda \qquad D \rightarrow d$$
$$C \rightarrow D \mid \lambda$$

What variables are erasable?

- A, B and C

Construct the resulting grammar after removing all λ-rules:

$$S \rightarrow ABaC \mid BaC \mid AaC \mid ABa \mid aC \mid Ba \mid Aa \mid a$$
$$A \rightarrow BC \mid C \mid B \mid \lambda$$
$$B \rightarrow b \mid \lambda \qquad C \rightarrow D \mid \lambda \qquad D \rightarrow d$$

Exercise

Consider the following grammar

$$S \rightarrow ABaC \qquad A \rightarrow BC \qquad B \rightarrow b \mid \lambda \qquad D \rightarrow d$$
$$C \rightarrow D \mid \lambda$$

What variables are erasable?

- A, B and C

Construct the resulting grammar after removing all λ-rules:

$$S \rightarrow ABaC \mid BaC \mid AaC \mid ABa \mid aC \mid Ba \mid Aa \mid a$$
$$A \rightarrow BC \mid C \mid B$$
$$B \rightarrow b \qquad C \rightarrow D \qquad D \rightarrow d$$

Unit Production Rules

Removal of Unit Production Rules

A rule $A \to B$ is called **unit production rule** (here $B \in V$).

Theorem

For every context-free language L there is a context-free grammar G without λ- and unit-productions with $L(G) = L \setminus \{\lambda\}$.

Construction

Let G be context-free, without λ-rules, and $L(G) = L \setminus \{\lambda\}$.

- Determine all pairs $A \neq B$ with $A \Rightarrow^+ B$.
- Whenever $A \Rightarrow^+ B$ and $B \to y$ is a rule, add a rule $A \to y$.
- Remove all unit production rules.

The resulting grammar G has no λ- and unit-productions and it has the property $L(G) = L \setminus \{\lambda\}$.

Exercise

Remove all unit production rules from

$$S \to Aa \mid B \qquad A \to a \mid bc \mid B \qquad B \to A \mid bb$$

Note that there are no λ-productions.
(So no need to first remove λ-productions.)

We determine all pairs $A \neq B$ with $A \Rightarrow^+ B$:

$$S \Rightarrow^+ B \qquad A \Rightarrow^+ B \qquad B \Rightarrow^+ A \qquad S \Rightarrow^+ A$$

Thus we add the following rules:

$$S \to Aa \mid B \mid a \mid bc \mid A \mid bb$$
$$A \to a \mid bc \mid B \mid A \mid bb$$
$$B \to A \mid bb \mid a \mid bc \mid B$$

Removing all unit production rules yields the final result:

$$S \to a \mid bb \mid bc \mid Aa \qquad A \to a \mid bb \mid bc \qquad B \to a \mid bb \mid bc$$

Chomsky Normal Form

Chomsky Normal Form

In a grammar in **Chomsky normal form** all rules have the form

$$A \rightarrow BC \quad \text{or} \quad A \rightarrow a$$

Note that a grammar in Chomsky normal form contains

- no λ-production rules,
- no unit production rules.

Theorem

For every context-free language L there is a grammar G in Chomsky normal form with $L(G) = L \setminus \{\lambda\}$.

Chomsky Normal Form

Construction

Let G be a context-free grammar without λ- and unit-productions and $L(G) = L \setminus \{\lambda\}$.

- Introduce variables C_a and rules $C_a \to a$ for every $a \in T$.
- Replace every rule $A \to x_1 \cdots x_n$ ($x_i \in V \cup T$) with $n \geq 2$ by

$$A \to \sigma(x_1) \cdots \sigma(x_n) \quad \text{where} \quad \sigma(x) = \begin{cases} x, & \text{if } x \in V \\ C_x, & \text{if } x \in T \end{cases}$$

- Replace every $A \to B_1 \cdots B_n$ with $n \geq 3$ by

$$A \to B_1 \cdots B_{n-2} C \qquad\qquad C \to B_{n-1} B_n$$

  where C is a fresh variable.

Repeat the last step until all rules are in Chomsky normal form.

Exercise

Transform the following context-free grammar

$$S \to aSbX \mid a$$
$$X \to Xa \mid aba$$

into Chomsky normal form.

The λ-rules and unit-rules have already been removed.

We continue the transformation:

$$S \to aSbX \mid a$$
$$X \to Xa \mid aba$$
$$C_a \to a$$
$$C_b \to b$$

Exercise

Transform the following context-free grammar

$$S \rightarrow aSbX \mid a$$
$$X \rightarrow Xa \mid aba$$

into Chomsky normal form.

The λ-rules and unit-rules have already been removed.

We continue the transformation:

$$S \rightarrow C_a S C_b X \mid a$$
$$X \rightarrow Xa \mid aba$$
$$C_a \rightarrow a$$
$$C_b \rightarrow b$$

Exercise

Transform the following context-free grammar

$$S \to aSbX \mid a$$
$$X \to Xa \mid aba$$

into Chomsky normal form.

The λ-rules and unit-rules have already been removed.

We continue the transformation:

$$S \to C_a S C_b X \mid a$$
$$X \to X C_a \mid aba$$
$$C_a \to a$$
$$C_b \to b$$

Exercise

Transform the following context-free grammar

$$S \rightarrow aSbX \mid a$$
$$X \rightarrow Xa \mid aba$$

into Chomsky normal form.

The λ-rules and unit-rules have already been removed.

We continue the transformation:

$$S \rightarrow C_a S C_b X \mid a$$
$$X \rightarrow X C_a \mid C_a C_b C_a$$
$$C_a \rightarrow a$$
$$C_b \rightarrow b$$

Exercise

Transform the following context-free grammar

$$S \rightarrow aSbX \mid a$$
$$X \rightarrow Xa \mid aba$$

into Chomsky normal form.

The λ-rules and unit-rules have already been removed.

We continue the transformation:

$$S \rightarrow C_a S_2 \mid a \qquad S_2 \rightarrow S S_3 \qquad S_3 \rightarrow C_b X$$
$$X \rightarrow X C_a \mid C_a C_b C_a$$
$$C_a \rightarrow a$$
$$C_b \rightarrow b$$

Exercise

Transform the following context-free grammar

$$S \to aSbX \mid a$$
$$X \to Xa \mid aba$$

into Chomsky normal form.

The λ-rules and unit-rules have already been removed.

We continue the transformation:

$$S \to C_a S_2 \mid a \qquad S_2 \to S S_3 \qquad S_3 \to C_b X$$
$$X \to X C_a \mid C_a X_2 \qquad X_2 \to C_b C_a$$
$$C_a \to a$$
$$C_b \to b$$