Ацикличные грамматики и нормальная форма Хомского

Михайлова Инна Анатольевна

Институт математики и естественных наук.
Кафедра алгебры и фундаментальной информатики.

Ацикличные грамматики

- Циклом в КС грамматике $G = (\Gamma, \Sigma, P, S)$ называется вывод $A \Rightarrow_G^+ A$.

Ацикличные грамматики

- Циклом в КС грамматике $G = (\Gamma, \Sigma, P, S)$ называется вывод $A \Rightarrow_G^+ A$.
- Грамматика называется ациклической, если в ней нет циклов.

Ацикличные грамматики

- Циклом в КС грамматике $G = (\Gamma, \Sigma, P, S)$ называется вывод $A \Rightarrow_G^+ A$.
- Грамматика называется ациклической, если в ней нет циклов.

Лемма

Пусть в грамматике $G = (\Gamma, \Sigma, P, S)$ есть правила вывода $A_1 \to A_2, \ldots, A_{k-1} \to A_k, A_k \to A_1$, тогда грамматика G', полученная заменой каждого A_i на нетерминал A и удалением правила $A \to A$, эквивалентна исходной.

Ацикличные грамматики

- Циклом в КС грамматике $G = (\Gamma, \Sigma, P, S)$ называется вывод $A \Rightarrow_G^+ A$.
- Грамматика называется ациклической, если в ней нет циклов.

Лемма

Пусть в грамматике $G = (\Gamma, \Sigma, P, S)$ есть правила вывода $A_1 \to A_2, \ldots, A_{k-1} \to A_k, A_k \to A_1$, тогда грамматика G', полученная заменой каждого A_i на нетерминал A и удалением правила $A \to A$, эквивалентна исходной.

- Пусть $w \in L(G)$ и $S \Rightarrow_G^* w$, тогда стирая индексы у нетерминалов A_i, получим вывод w в G'.

Ацикличные грамматики

- Циклом в КС грамматике $G = (\Gamma, \Sigma, P, S)$ называется вывод $A \Rightarrow_G^+ A$.
- Грамматика называется ациклической, если в ней нет циклов.

Лемма

Пусть в грамматике $G = (\Gamma, \Sigma, P, S)$ есть правила вывода $A_1 \to A_2, \ldots, A_{k-1} \to A_k, A_k \to A_1$, тогда грамматика G', полученная заменой каждого A_i на нетерминал A и удалением правила $A \to A$, эквивалентна исходной.

- Пусть $w \in L(G)$ и $S \Rightarrow_G^* w$, тогда стирая индексы у нетерминалов A_i, получим вывод w в G'.
- Пусть $w \in L(G')$ и вывод $S \Rightarrow_{G'} \alpha_1 \Rightarrow_{G'} \ldots \Rightarrow_{G'} w$, восстановим индексы у нетерминалов A_i.

Ацикличные грамматики

- Циклом в КС грамматике $G = (\Gamma, \Sigma, P, S)$ называется вывод $A \Rightarrow_G^+ A$.
- Грамматика называется ациклической, если в ней нет циклов.

Лемма

Пусть в грамматике $G = (\Gamma, \Sigma, P, S)$ есть правила вывода $A_1 \to A_2, \ldots, A_{k-1} \to A_k, A_k \to A_1$, тогда грамматика G', полученная заменой каждого A_i на нетерминал A и удалением правила $A \to A$, эквивалентна исходной.

- Пусть $w \in L(G)$ и $S \Rightarrow_G^* w$, тогда стирая индексы у нетерминалов A_i, получим вывод w в G'.
- Пусть $w \in L(G')$ и вывод $S \Rightarrow_{G'} \alpha_1 \Rightarrow_{G'} \ldots \Rightarrow_{G'} \; w$, восстановим индексы у нетерминалов A_i.
- Рассмотрим две соседние цепочки:

$$\alpha_i = \gamma_1 A_i \gamma_2, \alpha_{i+1} = \gamma_1 \beta \gamma_2$$

По построению G' существует $A_j \in \Gamma$ и правило $A_j \to \beta \in P$, тогда в грамматике G есть вывод

$$\alpha_i = \gamma_1 A_i \gamma_2 \Rightarrow_G^* \gamma_1 A_j \gamma_2 \Rightarrow_G \gamma_1 \beta \gamma_2$$

Ацикличные грамматики

- Циклом в КС грамматике $G = (\Gamma, \Sigma, P, S)$ называется вывод $A \Rightarrow_G^+ A$.
- Грамматика называется ациклической, если в ней нет циклов.

Лемма

Пусть в грамматике $G = (\Gamma, \Sigma, P, S)$ есть правила вывода $A_1 \to A_2, \ldots, A_{k-1} \to A_k, A_k \to A_1$, тогда грамматика G', полученная заменой каждого A_i на нетерминал A и удалением правила $A \to A$, эквивалентна исходной.

- Пусть $w \in L(G)$ и $S \Rightarrow_G^* w$, тогда стирая индексы у нетерминалов A_i, получим вывод w в G'.
- Пусть $w \in L(G')$ и вывод $S \Rightarrow_{G'} \alpha_1 \Rightarrow_{G'} \ldots \Rightarrow_{G'} w$, восстановим индексы у нетерминалов A_i.
- Рассмотрим две соседние цепочки:

$$\alpha_i = \gamma_1 A_i \gamma_2, \alpha_{i+1} = \gamma_1 \beta \gamma_2$$

  По построению G' существует $A_j \in \Gamma$ и правило $A_j \to \beta \in P$, тогда в грамматике G есть вывод

$$\alpha_i = \gamma_1 A_i \gamma_2 \Rightarrow_G^* \gamma_1 A_j \gamma_2 \Rightarrow_G \gamma_1 \beta \gamma_2$$

- Следовательно, можно восстановить вывод цепочки w в грамматике G.

- Говорят, что КС грамматика $G = (\Gamma, \Sigma, P, S)$ находится в нормальной форме Хомского (НФХ), если G ε-свободная и каждое ее правило вывода имеет вид $A \to BC$, либо $A \to a$.

Нормальная форма Хомского

- Говорят, что КС грамматика $G = (\Gamma, \Sigma, P, S)$ находится в нормальной форме Хомского (НФХ), если G ε-свободная и каждое ее правило вывода имеет вид $A \to BC$, либо $A \to a$.

Теорема

Для любой КС грамматики существует эквивалентная ей грамматика в НФХ.

Нормальная форма Хомского

- Говорят, что КС грамматика $G = (\Gamma, \Sigma, P, S)$ находится в нормальной форме Хомского (НФХ), если G ε-свободная и каждое ее правило вывода имеет вид $A \to BC$, либо $A \to a$.

Теорема

Для любой КС грамматики существует эквивалентная ей грамматика в НФХ.

- Пусть $G = (\Gamma, \Sigma, P, S)$, будем считать, что G — ε-свободная грамматика.

Нормальная форма Хомского

- Говорят, что КС грамматика $G = (\Gamma, \Sigma, P, S)$ находится в нормальной форме Хомского (НФХ), если G ε-свободная и каждое ее правило вывода имеет вид $A \to BC$, либо $A \to a$.

Теорема

Для любой КС грамматики существует эквивалентная ей грамматика в НФХ.

- Пусть $G = (\Gamma, \Sigma, P, S)$, будем считать, что G — ε-свободная грамматика.
- Для каждого нетерминала $a \in \Sigma$ введем новый нетерминал C_a. Заменим каждое вхождение терминала в правых частях правил вывода грамматики P на соответствующий ему нетерминал. Добавим правила $C_a \to a$ и рассмотрим грамматику $G_1 = (\Gamma \cup \{C_a \mid a \in \Sigma\}, \Sigma, P_1, S)$, которая эквивалентна исходной.

Нормальная форма Хомского

- Говорят, что КС грамматика $G = (\Gamma, \Sigma, P, S)$ находится в нормальной форме Хомского (НФХ), если G ε-свободная и каждое ее правило вывода имеет вид $A \to BC$, либо $A \to a$.

Теорема

Для любой КС грамматики существует эквивалентная ей грамматика в НФХ.

- Пусть $G = (\Gamma, \Sigma, P, S)$, будем считать, что G — ε-свободная грамматика.
- Для каждого нетерминала $a \in \Sigma$ введем новый нетерминал C_a. Заменим каждое вхождение терминала в правых частях правил вывода грамматики P на соответствующий ему нетерминал. Добавим правила $C_a \to a$ и рассмотрим грамматику $G_1 = (\Gamma \cup \{C_a \mid a \in \Sigma\}, \Sigma, P_1, S)$, которая эквивалентна исходной.
- Пусть $A \to B_1 B_2 \dots B_m \in P_1$, здесь B_i — нетерминалы. Рассмотрим случаи:

- Говорят, что КС грамматика $G = (\Gamma, \Sigma, P, S)$ находится в нормальной форме Хомского (НФХ), если G ε-свободная и каждое ее правило вывода имеет вид $A \to BC$, либо $A \to a$.

Теорема

Для любой КС грамматики существует эквивалентная ей грамматика в НФХ.

- Пусть $G = (\Gamma, \Sigma, P, S)$, будем считать, что G — ε-свободная грамматика.
- Для каждого нетерминала $a \in \Sigma$ введем новый нетерминал C_a. Заменим каждое вхождение терминала в правых частях правил вывода грамматики P на соответствующий ему нетерминал. Добавим правила $C_a \to a$ и рассмотрим грамматику $G_1 = (\Gamma \cup \{C_a \mid a \in \Sigma\}, \Sigma, P_1, S)$, которая эквивалентна исходной.
- Пусть $A \to B_1 B_2 \dots B_m \in P_1$, здесь B_i — нетерминалы. Рассмотрим случаи:
- Случай $m = 1$ Если в грамматике G_1 есть цикл, то используя лемму можно от него избавиться. Будем считать, что грамматика ациклична.

Нормальная форма Хомского

- Говорят, что КС грамматика $G = (\Gamma, \Sigma, P, S)$ находится в нормальной форме Хомского (НФХ), если G ε-свободная и каждое ее правило вывода имеет вид $A \to BC$, либо $A \to a$.

Теорема

Для любой КС грамматики существует эквивалентная ей грамматика в НФХ.

- Пусть $G = (\Gamma, \Sigma, P, S)$, будем считать, что G — ε-свободная грамматика.
- Для каждого нетерминала $a \in \Sigma$ введем новый нетерминал C_a. Заменим каждое вхождение терминала в правых частях правил вывода грамматики P на соответствующий ему нетерминал. Добавим правила $C_a \to a$ и рассмотрим грамматику $G_1 = (\Gamma \cup \{C_a \mid a \in \Sigma\}, \Sigma, P_1, S)$, которая эквивалентна исходной.
- Пусть $A \to B_1 B_2 \ldots B_m \in P_1$, здесь B_i — нетерминалы. Рассмотрим случаи:
- Случай $m = 1$ Если в грамматике G_1 есть цикл, то используя лемму можно от него избавиться. Будем считать, что грамматика ациклична.
- Пусть $A_1 \to A_2$ — правило вывода такое, что нет правила вида $A_2 \to A_3$. Удалим его и для каждого правила $A_2 \to \beta$ добавим правило $A_1 \to \beta$. Тогда полученная грамматика будет эквивалентна исходной и в ней будет меньше правил вида нетерминал в нетерминал.

Нормальная форма Хомского

- Говорят, что КС грамматика $G = (\Gamma, \Sigma, P, S)$ находится в нормальной форме Хомского (НФХ), если G ε-свободная и каждое ее правило вывода имеет вид $A \to BC$, либо $A \to a$.

Теорема

Для любой КС грамматики существует эквивалентная ей грамматика в НФХ.

- Пусть $G = (\Gamma, \Sigma, P, S)$, будем считать, что G — ε-свободная грамматика.
- Для каждого нетерминала $a \in \Sigma$ введем новый нетерминал C_a. Заменим каждое вхождение терминала в правых частях правил вывода грамматики P на соответствующий ему нетерминал. Добавим правила $C_a \to a$ и рассмотрим грамматику $G_1 = (\Gamma \cup \{C_a \mid a \in \Sigma\}, \Sigma, P_1, S)$, которая эквивалентна исходной.
- Пусть $A \to B_1 B_2 \dots B_m \in P_1$, здесь B_i — нетерминалы. Рассмотрим случаи:
- Случай $m = 1$ Если в грамматике G_1 есть цикл, то используя лемму можно от него избавиться. Будем считать, что грамматика ациклична.
- Пусть $A_1 \to A_2$ — правило вывода такое, что нет правила вида $A_2 \to A_3$. Удалим его и для каждого правила $A_2 \to \beta$ добавим правило $A_1 \to \beta$. Тогда полученная грамматика будет эквивалентна исходной и в ней будет меньше правил вида нетерминал в нетерминал.
- Случай $m \geq 3$ Введем новые нетерминалы $D_1, \dots, D_{m-2}$ и заменим это правило на набор новых: $A \to B_1 D_1, D_1 \to B_2 D_2, \dots, D_{m-3} \to B_{m-2} D_{m-2}, D_{m-2} \to B_{m-1} B_m$.

Пример

Привести грамматику к нормальной форме Хомского

$$
\begin{aligned}
S &\to aBa \mid B \\
A &\to b \mid C \\
B &\to C \mid a \\
C &\to A \mid bb
\end{aligned}
$$

Пример

Привести грамматику к нормальной форме Хомского

$$
\begin{aligned}
S &\to aBa \mid B \\
A &\to b \mid C \\
B &\to C \mid a \\
C &\to A \mid bb
\end{aligned}
$$

- Сначала заметим, что есть цикл $A \Rightarrow C \Rightarrow A$, отождествим нетерминалы A и C:

$$
\begin{aligned}
S &\to aBa \mid B \\
A &\to b \mid bb \\
B &\to A \mid a
\end{aligned}
$$

Пример

Привести грамматику к нормальной форме Хомского

$$
\begin{aligned}
S &\to aBa \mid B \\
A &\to b \mid C \\
B &\to C \mid a \\
C &\to A \mid bb
\end{aligned}
$$

- Сначала заметим, что есть цикл $A \Rightarrow C \Rightarrow A$, отождествим нетерминалы A и C:

$$
\begin{aligned}
S &\to aBa \mid B \\
A &\to b \mid bb \\
B &\to A \mid a
\end{aligned}
$$

- Введем нетерминалы A_1 и B_1 такие, что $A_1 \to a, B_1 \to b$, заменим в каждом правиле терминалы на нетерминалы:

$$
\begin{aligned}
S &\to A_1BA_1 \mid B \\
A &\to B_1 \mid B_1B_1 \\
B &\to A \mid A_1 \\
A_1 &\to a \\
B_1 &\to b
\end{aligned}
$$

Пример(продолжение)

- Текущая грамматика

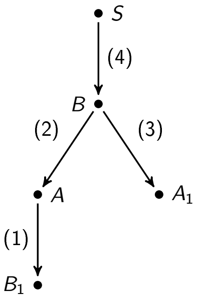
$$
\begin{aligned}
&S \to A_1BA_1 \mid B\\
&A \to B_1 \mid B_1B_1\\
&B \to A \mid A_1\\
&A_1 \to a, B_1 \to b
\end{aligned}
$$

Пример(продолжение)

- Текущая грамматика

$$S \to A_1BA_1 \mid B$$
$$A \to B_1 \mid B_1B_1$$
$$B \to A \mid A_1$$
$$A_1 \to a, B_1 \to b$$

Порядок удаления правил



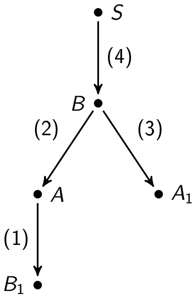
- Текущая грамматика

$$S \to A_1BA_1 \mid B$$
$$A \to B_1 \mid B_1B_1$$
$$B \to A \mid A_1$$
$$A_1 \to a, B_1 \to b$$

(1)

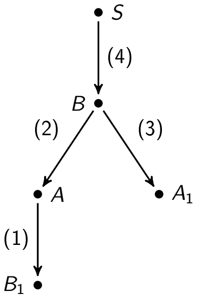
$$S \to A_1BA_1 \mid B$$
$$A \to b \mid B_1B_1$$
$$B \to A \mid A_1$$
$$A_1 \to a, B_1 \to b$$

Порядок удаления правил



Пример(продолжение)

Порядок удаления правил



- Текущая грамматика

$$\begin{aligned} S &\to A_1BA_1 \mid B \\ A &\to B_1 \mid B_1B_1 \\ B &\to A \mid A_1 \\ A_1 &\to a, B_1 \to b \end{aligned}$$

(1)

$$\begin{aligned} S &\to A_1BA_1 \mid B \\ A &\to b \mid B_1B_1 \\ B &\to A \mid A_1 \\ A_1 &\to a, B_1 \to b \end{aligned}$$

(2)

$$\begin{aligned} S &\to A_1BA_1 \mid B \\ A &\to b \mid B_1B_1 \\ B &\to b \mid B_1B_1 \mid A_1 \\ A_1 &\to a, B_1 \to b \end{aligned}$$

Пример(продолжение)

- Текущая грамматика

$$
\begin{aligned}
&S \to A_1BA_1 \mid B\\
&A \to B_1 \mid B_1B_1\\
&B \to A \mid A_1\\
&A_1 \to a, B_1 \to b
\end{aligned}
$$

(1)

$$
\begin{aligned}
&S \to A_1BA_1 \mid B\\
&A \to b \mid B_1B_1\\
&B \to A \mid A_1\\
&A_1 \to a, B_1 \to b
\end{aligned}
$$

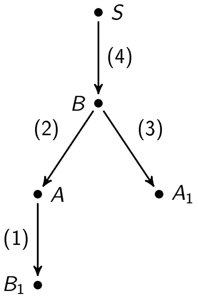
(2)

$$
\begin{aligned}
&S \to A_1BA_1 \mid B\\
&A \to b \mid B_1B_1\\
&B \to b \mid B_1B_1 \mid A_1\\
&A_1 \to a, B_1 \to b
\end{aligned}
$$

(3)

$$
\begin{aligned}
&S \to A_1BA_1 \mid B\\
&A \to b \mid B_1B_1\\
&B \to b \mid B_1B_1 \mid a\\
&A_1 \to a, B_1 \to b
\end{aligned}
$$

Порядок удаления правил



Пример(продолжение)

- Текущая грамматика

$$\begin{aligned} &S \to A_1BA_1 \mid B \\ &A \to B_1 \mid B_1B_1 \\ &B \to A \mid A_1 \\ &A_1 \to a, B_1 \to b \end{aligned}$$

(1)

$$\begin{aligned} &S \to A_1BA_1 \mid B \\ &A \to b \mid B_1B_1 \\ &B \to A \mid A_1 \\ &A_1 \to a, B_1 \to b \end{aligned}$$

(2)

$$\begin{aligned} &S \to A_1BA_1 \mid B \\ &A \to b \mid B_1B_1 \\ &B \to b \mid B_1B_1 \mid A_1 \\ &A_1 \to a, B_1 \to b \end{aligned}$$

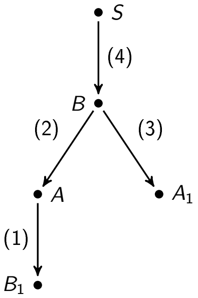
(3)

$$\begin{aligned} &S \to A_1BA_1 \mid B \\ &A \to b \mid B_1B_1 \\ &B \to b \mid B_1B_1 \mid a \\ &A_1 \to a, B_1 \to b \end{aligned}$$

(4)

$$\begin{aligned} &S \to A_1BA_1 \mid b \mid B_1B_1 \mid a \\ &A \to b \mid B_1B_1 \\ &B \to b \mid B_1B_1 \mid a \\ &A_1 \to a, B_1 \to b \end{aligned}$$

Порядок удаления правил



Пример(продолжение)

- Текущая грамматика:

$$S \to A_1BA_1 \mid b \mid B_1B_1 \mid a$$
$$A \to b \mid B_1B_1$$
$$B \to b \mid B_1B_1 \mid a$$
$$A_1 \to a, B_1 \to b$$

- Осталось разобраться с правыми частями, у которых длины больше двух, и мы получим грамматику в нормальной форме Хомского:

$$S \to A_1D \mid b \mid B_1B_1 \mid a$$
$$A \to b \mid B_1B_1$$
$$B \to b \mid B_1B_1 \mid a$$
$$D \to BA_1$$
$$A_1 \to a, B_1 \to b$$