Finite automata theory and formal languages (DIT321, TMV027)

Nils Anders Danielsson

2019-02-21

Today

Context-free grammars: syntax and semantics.

Syntax

Context-free grammars

A context-free grammar has the form (N, Σ, P, S):

- N is a finite set of *nonterminals*.
- Σ is a finite set of *terminals* satisfying $\Sigma \cap N = \emptyset$.
- $P \subseteq N \times (N \cup \Sigma)^*$ is a finite set of *productions*.
- The start symbol $S \in N$.

Notation

- A production (A, α) can be written $A \to \alpha$.
- Multiple productions $A \to \alpha_1$, ..., $A \to \alpha_n$ can be written $A \to \alpha_1 \mid \cdots \mid \alpha_n$ (at least if $n \geq 2$).

Which of the following expressions are well-formed context-free grammars?

1. $(\mathbb{N}, \{ a, b \}, P, 0)$, where P contains the following productions: $0 \to a1$, $1 \to b$.
2. $(\{ 0, 1 \}, \{ a, b \}, P, 0)$, where P contains the following productions: $0 \to a1$, $1 \to b$.
3. $(\{ 0, 1 \}, \{ 0, 1 \}, P, 0)$, where P contains the following productions: $0 \to 01$, $1 \to 1$.
4. $(\{ 0, 1 \}, \{ 0', 1' \}, P, 0)$, where P contains the following productions: $0 \to 01$, $1 \to 1 \mid 0$.
5. $(\{ 0, 1 \}, \{ 0', 1' \}, P, 2)$, where P contains the following productions: $0 \to 01$, $1 \to 1 \mid 0$.

Examples

An example

A context-free grammar for the non-regular language $\{ 0^n 1^n \mid n \in \mathbb{N} \}$ over $\{ 0, 1 \}$:

$$(\{ S \}, \{ 0, 1 \}, S \to 0S1 \mid \varepsilon, S)$$

An example

A context-free grammar for the non-regular language $\{ 0^n 1^n \mid n \in \mathbb{N} \}$ over $\{ 0, 1 \}$:

$$(\{ S \}, \{ 0, 1 \}, S \to 0S1 \mid \varepsilon, S)$$

Generated strings:

- ε.
- $0\varepsilon 1 = 01$.
- 0011.
- $\vdots$

An example

A context-free grammar for the non-regular language $\{ 0^n 1^n \mid n \in \mathbb{N} \}$ over $\{ 0, 1 \}$:

$$(\{ S \}, \{ 0, 1 \}, S \to 0S1 \mid \varepsilon, S)$$

An inductive definition of the language $L \subseteq \{ 0, 1 \}^*$ generated by the grammar:

$$\frac{w \in L}{0w1 \in L} \qquad \frac{}{\varepsilon \in L}$$

Another example

Consider the grammar $(\{ S, A \}, \{ 0, 1 \}, P, S)$, where P is defined in the following way:

$$S \to 0A1 \mid \varepsilon \qquad\qquad A \to 1A0 \mid S$$

Another example

Consider the grammar $(\{ S, A \}, \{ 0, 1 \}, P, S)$, where P is defined in the following way:

$$S \to 0A1 \mid \varepsilon \qquad\qquad A \to 1A0 \mid S$$

Sentential forms:

- S.
- ε.
- $0A1$.
- $01A01$.
- $01S01$.
- 0101.
- $\vdots$

Another example

Consider the grammar $(\{ S, A \}, \{ 0, 1 \}, P, S)$, where P is defined in the following way:

$$S \to 0A1 \mid \varepsilon \qquad\qquad A \to 1A0 \mid S$$

An inductive definition of the languages $L_S, L_A \subseteq \{ 0, 1 \}^*$ generated by S and A:

$$\frac{w \in L_A}{0w1 \in L_S} \qquad\qquad \frac{}{\varepsilon \in L_S}$$

$$\frac{w \in L_A}{1w0 \in L_A} \qquad\qquad \frac{w \in L_S}{w \in L_A}$$

Construct a context-free grammar for the language $\{ 0^{3n}1^{2n} \mid n \in \mathbb{N} \}$ over $\{ 0, 1 \}^*$ by filling in the missing part of the following definition.

$(\{ S \}, \{ 0, 1 \}, S \to ???, S)$

Semantics

Derivations

For the grammar $G = (N, \Sigma, P, S)$ one can define the following two binary relations on $(N \cup \Sigma)^*$ inductively:

$$\frac{\alpha, \beta \in (N \cup \Sigma)^* \qquad A \in N \qquad (A, \gamma) \in P}{\alpha A \beta \Rightarrow \alpha \gamma \beta}$$

$$\frac{}{\alpha \Rightarrow^* \alpha} \qquad \frac{\alpha \Rightarrow \beta \qquad \beta \Rightarrow^* \gamma}{\alpha \Rightarrow^* \gamma}$$

The language $L(G) = \{ w \in \Sigma^* \mid S \Rightarrow^* w \}$.

Leftmost derivations

A variant:

$$\frac{w \in \Sigma^* \qquad A \in N \qquad \alpha \in (N \cup \Sigma)^* \qquad (A, \beta) \in P}{wA\alpha \Rightarrow_{\mathsf{lm}} w\beta\alpha}$$

$$\frac{}{\alpha \Rightarrow^*_{\mathsf{lm}} \alpha} \qquad \frac{\alpha \Rightarrow_{\mathsf{lm}} \beta \qquad \beta \Rightarrow^*_{\mathsf{lm}} \gamma}{\alpha \Rightarrow^*_{\mathsf{lm}} \gamma}$$

Which of the following propositions are valid?

1. $\alpha \Rightarrow^* \beta \quad \Leftrightarrow \quad \alpha \Rightarrow^*_{\mathsf{lm}} \beta$
2. $\alpha \Rightarrow^* \beta \quad \Leftrightarrow \quad \beta \Rightarrow^* \alpha$
3. $\exists w \in \Sigma^*.\ \varepsilon \Rightarrow w$
4. $\exists w \in \Sigma^*.\ \varepsilon \Rightarrow^*_{\mathsf{lm}} w$

Recursive inference

If the grammar $G = (N, \Sigma, P, S)$, then one can define certain languages over Σ inductively:

- The language generated by the nonterminal $A \in N$, $L(G, A)$.
- The language generated by a list $\alpha \in (N \cup \Sigma)^*$, $L^*(G, \alpha)$.

Recursive inference

$$\frac{(A, \alpha) \in P \qquad w \in L^*(G, \alpha)}{w \in L(G, A)}$$

$$\frac{}{\varepsilon \in L^*(G, \varepsilon)} \qquad \frac{a \in \Sigma \qquad w \in L^*(G, \alpha)}{aw \in L^*(G, a\alpha)}$$

$$\frac{A \in N \qquad v \in L(G, A) \qquad w \in L^*(G, \alpha)}{vw \in L^*(G, A\alpha)}$$

Consider the grammar $(\{A, B\}, \{0, 1\}, P, A)$, where P is defined in the following way:

$$A \to 0B0$$
$$B \to 1A1 \mid \varepsilon$$

Which of the following propositions are true?

1. $1001 \in L(G, A)$
2. $1001 \in L(G, B)$
3. $00 \in L^*(G, AB)$
4. $0000 \in L^*(G, AB)$

Recursive inference

A derivation:

$$\dfrac{A \to 0B0 \in P \qquad \dfrac{\dfrac{\dfrac{B \to \varepsilon \in P \quad \dfrac{}{\varepsilon \in L^*(G,\varepsilon)}}{\varepsilon \in L(G,B)} \qquad \dfrac{\dfrac{}{\varepsilon \in L^*(G,\varepsilon)}}{0 \in L^*(G,0)}}{0 \in L^*(G,B0)}}{00 \in L^*(G,0B0)}}{00 \in L(G,A)}$$

Due to lack of space I have omitted "$a \in \Sigma$" and "$A \in N$".

Parse trees

Consider the following definitions again:

$$\frac{(A, \alpha) \in P \qquad w \in L^*(G, \alpha)}{w \in L(G, A)}$$

$$\frac{}{\varepsilon \in L^*(G, \varepsilon)} \qquad \frac{a \in \Sigma \qquad w \in L^*(G, \alpha)}{aw \in L^*(G, a\alpha)}$$

$$\frac{A \in N \qquad v \in L(G, A) \qquad w \in L^*(G, \alpha)}{vw \in L^*(G, A\alpha)}$$

Parse trees

Parse trees:

$$\frac{(A, \alpha) \in P \qquad ts \in P^*(G, \alpha)}{\mathsf{node}(A, ts) \in P(G, A)}$$

$$\frac{}{\mathsf{nil} \in P^*(G, \varepsilon)} \qquad \frac{a \in \Sigma \qquad ts \in P^*(G, \alpha)}{\mathsf{term}(a, ts) \in P^*(G, a\alpha)}$$

$$\frac{A \in N \qquad t \in P(G, A) \qquad ts \in P^*(G, \alpha)}{\mathsf{nonterm}(t, ts) \in P^*(G, A\alpha)}$$

Parse trees

The yield of a parse tree:

$$
\begin{aligned}
&\mathit{yield} \in P(G, A) \to \Sigma^* \\
&\mathit{yield}(\mathsf{node}(A, ts)) = \mathit{yield}^*(ts)
\end{aligned}
$$

$$
\begin{aligned}
&\mathit{yield}^* \in P^*(G, \ ) \to \Sigma^* \\
&\mathit{yield}^*(\mathsf{nil}) &&= \varepsilon \\
&\mathit{yield}^*(\mathsf{term}(a, ts)) &&= a\ \mathit{yield}^*(ts) \\
&\mathit{yield}^*(\mathsf{nonterm}(t, ts)) &&= \mathit{yield}(t)\ \mathit{yield}^*(ts)
\end{aligned}
$$

Yields containing nonterminals

The inductive definitions of recursive inference and parse trees can be extended to support strings containing both terminals and nonterminals:

$$\frac{}{A \in L_N(G, A)} \qquad \frac{}{\mathsf{leaf}(A) \in P_N(G, A)}$$

$$\begin{aligned}
&\mathit{yield} \in P_N(G, A) \to (N \cup \Sigma)^* \\
&\mathit{yield}(\mathsf{leaf}(A)) \qquad = A \\
&\mathit{yield}(\mathsf{node}(A, ts)) = \mathit{yield}^*(ts) \\
&\mathit{yield}^* \in P_N^*(G, \ ) \to (N \cup \Sigma)^* \\
&\vdots
\end{aligned}$$

Which of the following propositions are valid?

1. $\forall t \in P(G, A).\ yield(t) \in L(G, A)$
2. $\forall ts \in P^*(G, \alpha).\ yield^*(ts) \in L^*(G, \alpha)$
3. $\forall \alpha \in (N \cup \Sigma)^*.\ A \Rightarrow^* \alpha \iff \alpha \in L(G, A)$
4. $\forall \alpha \in (N \cup \Sigma)^*.\ A \Rightarrow^* \alpha \iff \alpha \in L_N(G, A)$
5. $w \in L^*(G, \alpha\beta) \Leftrightarrow$
   $\exists u \in L^*(G, \alpha), v \in L^*(G, \beta).\ w = uv$
6. $uv \in L^*(G, \alpha\beta) \Leftrightarrow$
   $u \in L^*(G, \alpha) \wedge v \in L^*(G, \beta)$

Proofs about grammars

A proof

Recall:

$$G = (\{ S \}, \{ 0, 1 \}, S \to 0S1 \mid \varepsilon, S)$$

$$\frac{w \in L}{0w1 \in L} \qquad \frac{}{\varepsilon \in L}$$

Let us prove that $L(G, S) \subseteq L$.

A proof

Let us prove $\forall w \in L(G, S).\ w \in L$ by complete induction on the length of the string:

- $w \in L(G, S)$ implies that $w \in L^*(G, \varepsilon)$ or $w \in L^*(G, 0S1)$.
- If $w \in L^*(G, \varepsilon)$, then $w = \varepsilon \in L$.
- If $w \in L^*(G, 0S1)$, then...

A proof

- If $w \in L^*(G, 0S1)$, then:
  - $w = 0w'$ for some $w' \in L^*(G, S1)$.
  - $w' = uv$ for some $u \in L(G, S)$, $v \in L^*(G, 1)$.
  - $v = 1$.
  - $|u| < |w|$, so by the inductive hypothesis $u \in L$.
  - Thus $w = 0u1 \in L$.

A proof

- If $w \in L^*(G, 0S1)$, then:
  - $w = 0w'$ for some $w' \in L^*(G, S1)$.
  - $w' = uv$ for some $u \in L(G, S)$, $v \in L^*(G, 1)$.
  - $v = 1$.
  - $|u| < |w|$, so by the inductive hypothesis $u \in L$.
  - Thus $w = 0u1 \in L$.

Another kind of induction can also be used.

Induction on the structure of the recursive inference

- Let $G = (N, \Sigma, P, -)$.
- Let Q be a relation on Σ^* and N.
- Let R be a relation on Σ^* and $(N \cup \Sigma)^*$.

Induction on the structure of the recursive inference

One form of induction for G:

$$
\begin{aligned}
&(\forall (A,\alpha) \in P, w \in L^*(G,\alpha).\ R(w,\alpha) \Rightarrow Q(w,A)) \land \\
&R(\varepsilon,\varepsilon) \land \\
&(\forall a \in \Sigma, \alpha \in (N \cup \Sigma)^*, w \in L^*(G,\alpha). \\
&\quad R(w,\alpha) \Rightarrow R(aw, a\alpha)) \land \\
&(\forall A \in N, \alpha \in (N \cup \Sigma)^*, v \in L(G,A), w \in L^*(G,\alpha). \\
&\quad Q(v,A) \land R(w,\alpha) \Rightarrow R(vw, A\alpha)) \Rightarrow \\
&(\forall A \in N, w \in L(G,A).\ Q(w,A)) \land \\
&(\forall \alpha \in (N \cup \Sigma)^*, w \in L^*(G,\alpha).\ R(w,\alpha))
\end{aligned}
$$

Induction on the structure of the recursive inference

One form of induction for G:

$$
\begin{aligned}
&(\forall (A,\alpha) \in P, w \in L^*(G,\alpha).\ R(w,\alpha) \Rightarrow Q(w,A)) \wedge \\
&R(\varepsilon,\varepsilon) \wedge \\
&(\forall a \in \Sigma, \alpha \in (N \cup \Sigma)^*, w \in L^*(G,\alpha). \\
&\quad R(w,\alpha) \Rightarrow R(aw, a\alpha)) \wedge \\
&(\forall A \in N, \alpha \in (N \cup \Sigma)^*, v \in L(G,A), w \in L^*(G,\alpha). \\
&\quad Q(v,A) \wedge R(w,\alpha) \Rightarrow R(vw, A\alpha)) \Rightarrow \\
&(\forall A \in N, w \in L(G,A).\ Q(w,A)) \wedge \\
&(\forall \alpha \in (N \cup \Sigma)^*, w \in L^*(G,\alpha).\ R(w,\alpha))
\end{aligned}
$$

A proof, take two

Recall:

$$G = (\{ S \}, \{ 0, 1 \}, S \to 0S1 \mid \varepsilon, S)$$

$$\frac{w \in L}{0w1 \in L} \qquad \frac{}{\varepsilon \in L}$$

Let us prove $L(G, S) \subseteq L$ by induction on the structure of the recursive inference.

- Let us prove that $L(G, S) \subseteq L$.
- Let $Q(w, \_)$ be $w \in L$.
- Let $R(w, \alpha)$ be

$$(\alpha \in \{ 0S1, \varepsilon \} \Rightarrow w \in L) \wedge$$
$$(\alpha = S1 \Rightarrow ??? \in L).$$

Suggest some replacement for ??? that will make the proof go through.

Today

- Context-free grammars.
- Derivations.
- Left-most derivations.
- Recursive inference.
- Parse trees.
- Proofs about grammars.

Next lecture

- More about context-free grammars.

- Deadline for the next quiz: 2019-02-25, 10:00.
- Deadline for the fourth assignment: 2019-02-24, 23:59.