Context-Free Languages

Outline

- Decision Properties of CFLs
- Closure Properties of CFLs

Decision Properties of CFLs

Summary of CFL Decision Properties

- Can answer:
  - Membership: $\omega \in \mathrm{C}$
  - Emptiness: $\mathrm{C} = \emptyset$
  - Regular Subset: $\mathrm{C} \subseteq \mathrm{R}$
- Cannot Answer
  - Equivalence: $\mathrm{C}_1 = \mathrm{C}_2$
  - Disjoint: $\mathrm{C}_1 \cap \mathrm{C}_2 = \emptyset$
  - Subset: $\mathrm{C}_1 \subseteq \mathrm{C}_2$

Membership

Given: CFL L, string w

Find: Is ω in L?

Solution (simple): Simulate PDA for L on ω

Solution (fast): Many specialized parsing algorithms. For string of length n:

- Earley: Any CFL, $\mathrm{O}(\mathrm{n}^3)$
- CYK: Any CFL, $\mathrm{O}(\mathrm{n}^3)$
- LALR: Some CFL, O(n)
- LL(1): Some CFL, O(n)
- LL(*): Some CFLs and some non-CFLs, $\mathrm{O}(\mathrm{n}^2)$

Emptiness

Given: CFL L

Find: Is L empty

Idea: In a CFG for L, can the start symbol derive a string of terminals

Useful Symbols

Useful: Nonterminal <X> is useful if:

> There exists a derivation <S> $\rightsquigarrow \alpha$<X>$\beta \rightsquigarrow \omega$, where <S> is the start symbol and $\omega \in$ T*

Requirements: Two requirements for usefulness:

1. A terminal string is derivable from <X>
2. <X> is in string derivable from <S>

Useless: no such derivation using <X>

Algorithm 1: Find terminal-deriving symbols

$V_{\text{old}} \leftarrow \emptyset;$

$$V_{\text{new}} \leftarrow \left\{ A \;\middle|\; \underbrace{A \to \sigma \text{ for some } \sigma \in T^*}_{A \text{ derives string of terminals}} \right\};$$

while $V_{\text{old}} \neq V_{\text{new}}$ **do**

- $V_{\text{old}} \leftarrow V_{\text{new}};$
- $$V_{\text{new}} \leftarrow V_{\text{old}} \cup \left\{ A \;\middle|\; \underbrace{A \to \alpha \text{ for } \alpha \in (T \cup V_{\text{old}})^*}_{A \text{ derives string of terminals and terminal-deriving nonterminals}} \right\};$$

return $V_{\text{new}};$

Algorithm 2: Find reachable symbols

$V_{\text{old}} \leftarrow \emptyset;$

$V_{\text{new}} \leftarrow \{S\};$

while $V_{\text{old}} \neq V_{\text{new}}$ **do**

- $V_{\text{old}} \leftarrow V_{\text{new}};$
- $$V_{\text{new}} \leftarrow V_{\text{old}} \cup \left\{ B \;\middle|\; \underbrace{\exists A \in V_{\text{old}},\ A \to \alpha B \beta}_{A \text{ derives } B} \right\};$$

return $V_{\text{new}};$

Useless Symbols

$V_{\text{useful}} = V_{\text{terminal-deriving}} \cap V_{\text{reachable}}$

P_{useful} = are productions containing only V_{useful}

Emptiness (Redefined)

Given: Context-Free Grammar G

Find: Is the language of G empty?

Solution: Is start symbol S a useless symbol

Closure Properties of CFLs

CFL Closure under Concatenation

Given: CFGs

$A = (V_a, T_a, P_a, S_a)$

$B = (V_b, T_b, P_b, S_b)$

Find: $C = (V_c, T_c, P_c, S_c)$ where

$L(C) = L(A)L(B)$

Solution: Concatenation of start symbols:

$V_C = V_A \cup V_B \cup \{S'\}$

$T_C = T_A \cup T_B$

$S_C = S'$

$P_C = P_A \cup P_B \cup \{S' \rightarrow S_A S_B\}$

CFL Closure under Union

Given: CFGs

$A = (V_a, T_a, P_a, S_a)$

$B = (V_b, T_b, P_b, S_b)$

Find: $C = (V_c, T_c, P_c, S_c)$ where

$L(C) = L(A) \cup L(B)$

Solution: Union on start symbols:

$V_C = V_A \cup V_B \cup \{S'\}$

$T_C = T_A \cup T_B$

$S_C = S'$

$P_C = P_A \cup P_B \cup \{S' \rightarrow S_A, S' \rightarrow S_B\}$

CFL Closure under Repetition

Given: CFGs

$A = (V_a, T_a, P_a, S_a)$

Find: $C = (V_c, T_c, P_c, S_c)$ where

$L(C) = L(A)^*$

Solution: Epsilon transition back to start symbols:

$V_C = V_A \cup \{S'\}$

$T_C = T_A$

$S_C = S'$

$P_C = P_A \cup \{S' \rightarrow S_A S', S' \rightarrow \epsilon\}$

Intersection and Difference

Regular languages are closed under intersection and difference

Context-free languages are **NOT** closed under intersection and difference

CFL Non-closure under Intersection

Theorem: The context-free languages are not closed under intersection[

Proof by counterexample:

1. $C_1 = \{a^n b^n c^i \mid n \geq 1, i \geq 1\}$ is context-free:

   <S> → <A><B>   <A> → a<A>b|ab   <B> → c<B>|c
2. $C_1 = \{a^i b^n c^n \mid n \geq 1, i \geq 1\}$ is context-free:

   <S> → <A><B>   <A> → a<A>|a   <B> → b<B>c|bc
3. $N = C_1 \cap C_2$
4. $N = \{a^n b^n c^n \mid n \geq 1\}$ is not context free (can prove with pumping lemma)

CFL Non-closure under Difference

Corollary: The context-free languages are not closed under difference

Proof by contradiction:

1. Assume the CFLs were closed under difference
2. Generally: $L \cap M = L \quad (L \quad M)$
3. If CFLs were closed under difference, they would also be closed under intersection
4. But CFLs are not closed under intersection, contradiction

CFL-Regular Intersection

Given: A context-free language and regular language as:

$\mathrm{PDA}\ \mathrm{P} = (\mathrm{Q_P}, \Sigma, \Gamma_\mathrm{P}, \delta_\mathrm{P}, \mathrm{q_{0,P}}, \mathrm{F_P})$

$\mathrm{DFA}\ \mathrm{D} = (\mathrm{Q_D}, \Sigma, \delta_\mathrm{D}, \mathrm{q_{0,D}}, \mathrm{F_D})$

Find: PDA C such that $L(\mathrm{C}) = L(\mathrm{P}) \cap L(\mathrm{D})$

Solution: Simulate P and D in parallel. Result is a PDA

$\mathrm{Q_C} = \mathrm{Q_P} \times {}_\mathrm{D}$ and $\mathrm{q_{0,C}} = (\mathrm{q_{0,P}}, \mathrm{q_{0,D}})$ and $\mathrm{F_C} = \mathrm{F_P} \times \mathrm{F_D}$

$\Gamma_\mathrm{C} = \Gamma_\mathrm{P}$

$$\delta_C\left(\overbrace{(q_P, q_D)}^{\text{predecessor}}, \overbrace{a}^{\text{input}}, \overbrace{\gamma}^{\text{popped}}\right) \triangleq$$

$$\left\{ \left(\underbrace{(q'_P, q'_D)}_{\text{successor}}, \underbrace{\gamma'}_{\text{pushed}} \right) \middle| \underbrace{((q'_P, \gamma') \in \delta_P(q_P, a, \gamma))}_{\text{PDA Transition}} \land \underbrace{(q'_D = \delta_D(q_D))}_{\text{DFA Transition}} \right\}$$

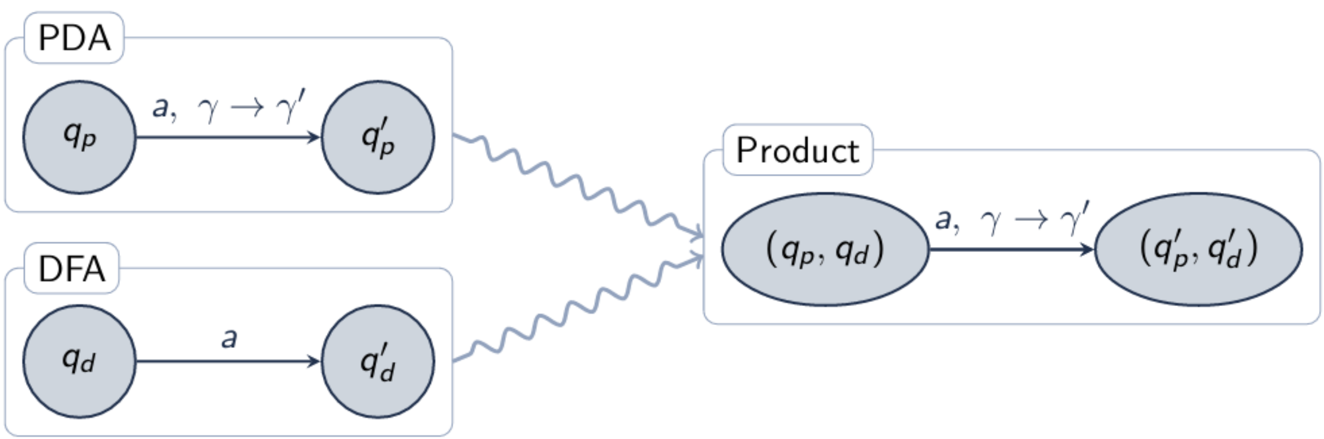

Figure 1: Example of CFL-Regular Intersection