Autômatos e computabilidade

Gramáticas Livre de Contexto

Pedro A D Rezende
UnB – IE – CIC

A & C: Gramáticas livre de contexto

Motivação

Linguagens como

- $\{a^n b^n \mid n \in \mathbb{N}\}$
- $\{\alpha \in \{a,b\}^* \mid |\alpha|_a = |\alpha|_b \land \alpha = uv \Rightarrow |u|_a \geq |u|_b\}$

  (linguagem de parêntesis)

não são regulares, não se expressam por ERs

A & C: Gramáticas livre de contexto

Motivação

Linguagens como

- $\{a^n b^n \mid n \in \mathbb{N}\}$
- $\{\alpha \in \{a,b\}^* \mid |\alpha|_a = |\alpha|_b \wedge \alpha = uv \Rightarrow |u|_a \geq |u|_b\}$

  (linguagem de parêntesis)

não são regulares, não se expressam por ERs

Que modelos computacionais existem para reconhecer ou gerar essas linguagens?

Quais os mais simples?

A & C: Gramáticas livre de contexto

Definições formais: Gramática gerativa (Chomsky)

Um Gramática Gerativa é uma quádrupla

$G = (V , T , P , S)$

Variáveis gramaticais	Alfabeto terminal	Regras de produção	Símbolo inicial

A & C: Gramáticas livre de contexto

Definições formais: Gramática gerativa (Chomsky)

Um Gramática Gerativa é uma quádrupla

$G = (V , T , P , S)$

V	T	P	S
Variáveis gramaticais	Alfabeto terminal	Regras de produção	Símbolo inicial

satisfazendo:

- V, T são conjuntos *finitos disjuntos*;

A & C: Gramáticas livre de contexto

Definições formais: Gramática gerativa (Chomsky)

Um Gramática Gerativa é uma quádrupla

$G = (V , T , P , S)$

V	T	P	S
Variáveis gramaticais	Alfabeto terminal	Regras de produção	Símbolo inicial

satisfazendo:

- V, T são conjuntos *finitos disjuntos*;
- $S \in V$ (em linguística, `<sentença>`);

A & C: Gramáticas livre de contexto

Definições formais: Gramática gerativa (Chomsky)

Um Gramática Gerativa é uma quádrupla

$$G = (\underbrace{V}_{\text{Variáveis gramaticais}} , \underbrace{T}_{\text{Alfabeto terminal}} , \underbrace{P}_{\text{Regras de produção}} , \underbrace{S}_{\text{Símbolo inicial}})$$

satisfazendo:

- V, T são conjuntos *finitos disjuntos*;
- $S \in V$ (em linguística, `<sentença>`);
- $P \subseteq V^* \times (V \cup T)^*$ é uma relação *finita*

A & C: Gramáticas livre de contexto

Definições formais: Gramática gerativa (Chomsky)

Um Gramática Gerativa é uma quádrupla

$G = (\; V \; , \; T \; , \; P \; , \; S \;)$

Variáveis gramaticais	Alfabeto terminal	Regras de produção	Símbolo inicial

satisfazendo:

- V, T são conjuntos *finitos disjuntos*;
- $S \in V$ (em linguística, `<sentença>`);
- $P \subseteq V^* \times (V \cup T)^*$ é uma relação *finita*

Def: Se $P \subseteq V \times (V \cup T)^*$, G é dita livre de contexto

A & C: Gramáticas livre de contexto

CFG (Gramática Livre de Contexto)

Dado $G = (V, T, P, S)$ gramática gerativa satisfazendo $P \subseteq V \times (V \cup T)^*$ (livre de contexto)

Denotamos, para $(A,\alpha) \in P$, a produção

$$A \to \alpha$$

Uma tal produção permite a opção de se substituir a variável gramatical $A \in V$ pela expressão $\alpha \in (V \cup T)^*$ em *derivações* de S

A & C: Gramáticas livre de contexto

Derivação em CFGs

Dado G=(V, T, P, S) CFG, uma derivação direta em G é a substituição de A por α em cadeias $\gamma \in (V \cup T)^*$ denotada por

$$\gamma \Rightarrow \gamma'$$

onde $\gamma = \beta A \beta'$, $\gamma' = \beta\alpha\beta'$, $\beta, \beta' \in (V \cup T)^*$, $A \rightarrow \alpha \in P$

A & C: Gramáticas livre de contexto

Derivação em CFGs

Dado G=(V, T, P, S) CFG, uma derivação direta em G é a substituição de A por α em cadeias $\gamma \in (V \cup T)^*$. Denota-se uma derivação direta por

$$\gamma \Rightarrow \gamma'$$

onde $\gamma = \beta A \beta'$, $\gamma' = \beta \alpha \beta'$, $\beta, \beta' \in (V \cup T)^*$, $A \rightarrow \alpha \in P$

Notação: (seguindo HU)

A, B, C...S $\in$ V (variáveis); W, X, Y, Z $\in$ V$\cup$T (símbolos)

a, b, c... $\in$ T (terminais); u, v, w... $\in$ T^* (cadeias de terminais);

$\alpha, \beta, \gamma, \varepsilon \in (V \cup T)^*$ (cadeias de símbolos)

A & C: Gramáticas livre de contexto

Derivação em CFGs

Dado G=(V, T, P, S) CFG, uma derivação em G é uma cadeia de derivações diretas em G

$\gamma_0 \Rightarrow \gamma_1 \Rightarrow \ldots \Rightarrow \gamma_n$ denotada por

$$\gamma \Rightarrow^* \gamma'$$

A & C: Gramáticas livre de contexto

Derivação em CFGs

Dado G=(V, T, P, S) CFG, uma derivação em G é uma cadeia de derivações diretas em G

$\gamma_0 \Rightarrow \gamma_1 \Rightarrow \ldots \Rightarrow \gamma_n$ denotada por

$$\gamma \Rightarrow^* \gamma'$$

onde $\gamma = \gamma_0$, $\gamma' = \gamma_n$,

$\gamma_i = \beta_i A_i \beta_i'$, $\gamma_{i+1} = \beta_i \alpha_i \beta_i'$,

$A_i \to \alpha_i \in P$ β_i, β_i', $\alpha_i = X_{i1} \ldots X_{is} \in (V \cup T)^*$,

A & C: Gramáticas livre de contexto

Derivação em CFGs

Dado G=(V, T, P, S) CFG, uma derivação em G é uma cadeia de derivações diretas em G

$\gamma_0 \Rightarrow \gamma_1 \Rightarrow \ldots \Rightarrow \gamma_n$ denotada por

$$\gamma \Rightarrow^* \gamma'$$

onde $\gamma = \gamma_0$, $\gamma' = \gamma_n$,

$\gamma_i = \beta_i A_i \beta_i'$, $\gamma_{i+1} = \beta_i \alpha_i \beta_i'$,

$A_i \rightarrow \alpha_i \in P$ $\beta_i, \beta_i', \alpha_i = X_{i1} \ldots X_{is} \in (V \cup T)^*$,

γ_i são chamadas formas sentenciais de G

A & C: Gramáticas livre de contexto

Linguagem gerada por uma CFG

Dado G CFG, a linguagem gerada por G é o conjunto L(G) das cadeias $x \in T^*$ tal que

$$S \Rightarrow^* x$$

A & C: Gramáticas livre de contexto

Linguagem gerada por uma CFG

Dado G CFG, a linguagem gerada por G é o conjunto L(G) das cadeias $x \in T^*$ tal que

$$S \Rightarrow^* x$$

Notação: $S \Rightarrow^* \gamma$ é dita derivação inicial em G

A & C: Gramáticas livre de contexto

Linguagem gerada por uma CFG

Dado G CFG, a linguagem gerada por G é o conjunto L(G) das cadeias $x \in T^*$ tal que

$$S \Rightarrow^* x$$

Notação: $S \Rightarrow^* \gamma$ é dita derivação inicial em G

$\gamma \Rightarrow^* x$, $x \in T^*$ é dita derivação final em G

A & C: Gramáticas livre de contexto

Linguagem gerada por uma CFG

Dado G CFG, a linguagem gerada por G é o conjunto L(G) das cadeias $x \in T^*$ tal que

$$S \Rightarrow^* x$$

Notação: $S \Rightarrow^* \gamma$ é dita derivação inicial em G

$\gamma \Rightarrow^* x$, $x \in T^*$ é dita derivação final em G

$\gamma_0 \Rightarrow \gamma_1 \Rightarrow \ldots \Rightarrow \gamma_n$ é dita derivação à esquerda se

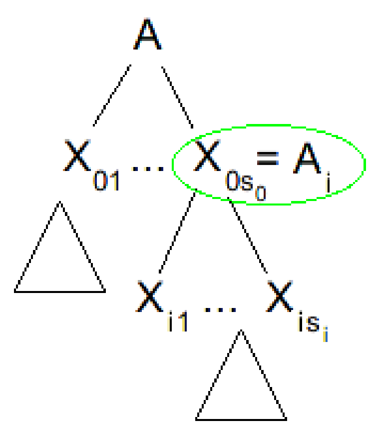
$\forall i \, [\gamma_i \Rightarrow \gamma_{i+1}$ substitui a 1a. ocorrência de A_i em $\gamma_i]$

(derivação à direita, se substitui a última)

A & C: Gramáticas livre de contexto

Representações de derivações em CFGs

Uma derivação $A \Rightarrow^* \gamma$ pode ser representada pela árvore formada por suas derivações diretas, chamada árvore de derivação de γ



onde $A_i \rightarrow X_{i1} \ldots X_{is_i}$ são as substituições das derivações diretas,

γ é formada pela concatenação das folhas (da esq. à direita)

A & C: Gramáticas livre de contexto

Árvores de derivações e CFGs (HU T 4.3)

Numa árvore *A* de nós rotulados por símbolos de G ($V \cup T \cup \{\lambda\}$), se a cada descendência corresponde uma produção de G então *A* representa derivações de G que derivam a forma setencial γ correspondente à concatenação das folhas (da esq. à dir.)

A & C: Gramáticas livre de contexto

Árvores de derivações e CFGs (HU T 4.3)

Numa árvore *A* de nós rotulados por símbolos de G ($V \cup T \cup \{\lambda\}$), se a cada descendência corresponde uma produção de G então *A* representa derivações de G que derivam a forma setencial γ correspondente à concatenação das folhas (da esq. à dir.)

OBS: *A* é dita árvore sintática (*parse tree*) de γ

A representa uma única derivação à esquerda, e uma única derivação à direita de γ.

A & C: Gramáticas livre de contexto

Linguagens de Livre Contexto (CFL)

$L \subseteq \Sigma^*$ é dita Linguagem de Livre Contexto (CFL) se existir CFG G tal que L = L(G)

.

A & C: Gramáticas livre de contexto

Linguagens de Livre Contexto (CFL)

$L \subseteq \Sigma^*$ é dita Linguagem de Livre Contexto (CFL) se existir CFG G tal que L = L(G)

Ambiguidade de CFGs (HU T 4.3)

Dado uma CFG G, se para algum $x \in L(G)$ existir mais de uma derivação à esquerda de x em G (mais de uma árvore sintática de x), dizemos que G é uma gramática ambígua.

A & C: Gramáticas livre de contexto

Linguagens de Livre Contexto (CFL)

$L \subseteq \Sigma^*$ é dita Linguagem de Livre Contexto (CFL) se existir CFG G tal que L = L(G)

Ambiguidade de CFGs (HU T 4.3)

Dado uma CFG G, se para algum $x \in L(G)$ existir mais de uma derivação à esquerda de x em G (mais de uma árvore sintática de x), dizemos que G é uma gramática ambígua.

OBS: Em aplicações tais como a construção de compiladores, evita-se a ambiguidade de G.