Linguagens Formais e Autômatos

Simplificação de Gramáticas Livre do Contexto (GLC)

Cristiano Lehrer, M.Sc.

Gramática Simplificada

- Gramática simplificada é uma gramática livre do contexto que não apresenta símbolos inúteis, produções vazias, nem produções do tipo $A \rightarrow B$.
  - Sequência de simplificação:
    - Exclusão das produções vazias.
    - Exclusão das produções da forma $A \rightarrow B$.
    - Exclusão dos símbolos inúteis.

Transformações de Gramática Livre do Contexto

- Uma vez que existem diferentes métodos de análise, cada qual exigindo gramáticas com características específicas, é importante que uma gramática possa ser transformada, porém, sem perder a qualidade de gerar a mesma linguagem.
- As gramáticas que, mesmo tendo conjuntos diferentes de produções, geram a mesma linguagem são ditas **gramáticas equivalentes**.

Exclusão de Produções Vazias (1/3)

- A exclusão de produções vazias (da forma $A \rightarrow \varepsilon$) pode determinar modificações diversas nas produções da gramática:
  - Variáveis que constituem produções vazias:
    - Considera, inicialmente, todas as variáveis que geram diretamente ε (exemplo $A \rightarrow \varepsilon$). A seguir, sucessivamente são determinadas as variáveis que indiretamente geram e (exemplo $B \rightarrow A$).
  - Exclusão de produções vazias:
    - Inicialmente, são consideradas todas as produções não-vazias. A seguir, cada produção cujo lado direito possui uma variável que gera a palavra vazia, determina uma produção adicional, sem esta variável.
  - Inclusão de geração da palavra vazia, se necessário:
    - Se a palavra vazia pertence à linguagem, então é incluída uma produção para gerar a palavra vazia.

Exclusão de Produções Vazias (2/3)

- Exemplo:
  - $G = (\{S, X, Y\}, \{a, b\}, P, S)$, sendo
  - $P = \{S \rightarrow aXa \mid bXb \mid \varepsilon, X \rightarrow a \mid b \mid Y, Y \rightarrow \varepsilon\}$
  - Conjunto de variáveis que geram ε:

iteração	variáveis
0	$\varnothing$
1	{S, Y}
2	{S, Y, X}
3	{S, Y, X}

Exclusão de Produções Vazias (3/3)

- Conjunto de produções sem produções vazias:

iteração	produções
0	$\{S \rightarrow aXa \mid bXb,\ X \rightarrow a \mid b \mid Y\}$
1	$\{S \rightarrow aXa \mid bXb \mid aa \mid bb,\ X \rightarrow a \mid b \mid Y\}$
2	$\{S \rightarrow aXa \mid bXb \mid aa \mid bb,\ X \rightarrow a \mid b \mid Y\}$

- Inclusão da geração da palavra vazia:
  - $G = (\{S, X, Y\}, \{a, b\}, P, S)$, sendo
  - $P = \{S \rightarrow aXa \mid bXb \mid aa \mid bb \mid \varepsilon,\ X \rightarrow a \mid b \mid Y\}$

Exclusão das Produções da Forma A → B (1/3)

- Uma produção da forma A → B não adiciona informação alguma em termos de geração de palavras, a não ser que a variável A pode ser substituída por B.
- Nesse caso, se B → α, então a produção A → B pode ser substituída por A → α:
  - Construção do fecho de cada variável:
    - Entende-se por fecho de uma variável o conjunto de variáveis que podem substituí-la transitivamente, ou seja, se A → B e B → C, então B e C pertencem ao fecho de A.
  - Exclusão das produções da forma A → B:
    - Substitui as produções da forma A → B por produções da forma A → α, onde α é atingível a partir de A através de seu fecho.

Exclusão das Produções da Forma A → B (2/3)

- Exemplo:
  - G = ({S, X}, {a, b}, P, S), sendo
  - P = {S → aXa | bXb, X → a | b | S | ε}
  - Construção do fecho de cada variável:
    - FECHO-S = ∅
    - FECHO-X = {S}

Exclusão das Produções da Forma A → B (3/3)

- Exclusão das produções da forma $A \rightarrow B$:

iteração	produções
0	$\{S \rightarrow aXa \mid bXb,\ X \rightarrow a \mid b \mid S \mid \varepsilon\}$
S	$\{S \rightarrow aXa \mid bXb,\ X \rightarrow a \mid b \mid S \mid \varepsilon\}$
X	$\{S \rightarrow aXa \mid bXb,\ X \rightarrow a \mid b \mid aXa \mid bXb \mid \varepsilon\}$

- $G = (\{S, X\}, \{a, b\}, P, S)$, sendo
- $P = \{S \rightarrow aXa \mid bXb,\ X \rightarrow a \mid b \mid aXa \mid bXb \mid \varepsilon\}$

Exclusão de Símbolos Inúteis (1/3)

- A exclusão de símbolos inúteis (não usados na geração de palavras de terminais) é realizada excluindo as produções que fazem referência a estes símbolos, bem como os próprios símbolos inúteis:
  - Qualquer variável gera palavra de terminais:
    - O algoritmo gera um novo conjunto de variáveis, analisando as produções da gramática a partir de terminais gerados. Inicialmente, considera todas as variáveis que geram terminais diretamente (exemplo $A \rightarrow a$). A seguir, sucessivamente são adicionadas as variáveis que geram palavras de terminais indiretamente (exemplo $B \rightarrow Ab$).
  - Qualquer símbolo é atingível a partir do símbolo inicial:
    - Após a execução da etapa acima, o algoritmo analisa as produções da gramática a partir do símbolo inicial. Inicialmente, considera exclusivamente o símbolo inicial. Após, sucessivamente as produções da gramática são aplicadas e os símbolos referenciados adicionados aos novos conjuntos.

Exclusão de Símbolos Inúteis (2/3)

- Exemplo:
  - $G = (\{S, A, B, C\}, \{a, b, c\}, P, S)$, onde
  - $P = \{S \rightarrow aAa \mid bBb, A \rightarrow a \mid S, C \rightarrow c\}$
  - Qualquer variável gera palavra de terminais:

iteração	**variáveis**
0	$\varnothing$
1	$\{A, C\}$
2	$\{A, C, S\}$
3	$\{A, C, S\}$

  - $G = (\{S, A, C\}, \{a, b, c\}, P, S)$, onde
  - $P = \{S \rightarrow aAa, A \rightarrow a \mid S, C \rightarrow c\}$

Exclusão de Símbolos Inúteis (3/3)

- G = ({S, A, C}, {a, b, c}, P, S), onde
- P = {S → aAa, A → a | S, C → c}
- Qualquer símbolo é atingido a partir do símbolo inicial.

iteração	variáveis	terminais
0	{S}	$\varnothing$
1	{S, A}	{a}
2	{S, A}	{a}

- G = ({S, A}, {a}, P, S), onde
- P = {S → aAa, A → a | S}

Formas Normais

- As formas normais estabelecem restrições rígidas na definição das produções, sem reduzir o poder de geração das Gramáticas Livres do Contexto.
- São usadas principalmente no desenvolvimento de algoritmos (com destaque para reconhecedores de linguagens) e na prova de teoremas;
- Forma Normal de Chomsky:
  - $< A > \rightarrow < B > < C >$
  - $< A > \rightarrow a$
- Forma Normal de Greibach
  - $< A > \rightarrow a\ \alpha$

Forma Normal de Chomsky (1/3)

- Simplificação da Gramática:
  - Simplifica a gramática, aplicando a seguinte sequencia de simplificação:
    - Exclusão das produções vazias (como a linguagem não possui a palavra vazia, todas as produções da forma < A > → ε pode ser excluídas).
    - Exclusão das produções da forma < A > → < B > (se o lado direito de alguma produção tiver somente um símbolo, este será terminal).
    - Exclusão dos símbolos inúteis.

Forma Normal de Chomsky (2/3)

- Transformação do lado direito das produções de comprimento maior ou igual a dois:
  - Garante que o lado direito das produções de comprimento maior ou igual a dois é composto exclusivamente por variáveis.
  - A exclusão de um terminal a pode ser realizada substituindo-o por uma variável intermediária $< Ca >$ e incluindo a produção $< Ca > \rightarrow a$

Forma Normal de Chomsky (3/3)

- Transformação do lado direito das produções de comprimento maior ou igual a três, em produções com exatamente duas variáveis:
  - Garante que o lado direito das produções de comprimento maior do que um é composto exatamente por duas variáveis.
  - Após a execução da etapa anterior, o lado direito das produções da forma < A > → < B1 > < B2 > … < Bn > (n > 2) é composto exclusivamente por variáveis.
  - Portanto, para concluir a transformação, é suficiente garantir que o lado direito é composto por exatamente duas variáveis.
  - Isto é possível gerando < B1 > < B2 > … < Bn > em diversas etapas, usando variáveis intermediárias.

Forma Normal de Greibach (1/6)

- Uma Gramática Livre do Contexto é dita na Foma Normal de Greibach (FNG) se todas as suas produções são da forma:
  - <A> → aα
    - A é uma variável.
    - a é um terminal.
    - α é uma palavra de variáveis.

Forma Normal de Greibach (2/6)

- Simplificação da Gramática:
  - Análoga à correspondente etapa do algoritmo referente a Forma Normal de Chomsky.
- Renomeação das variáveis em uma ordem crescente qualquer:
  - As variáveis da gramática são renomeadas em uma ordem crescente qualquer, como por exemplo, A_1, A_2, …., A_n, onde n é o cardinal do conjunto de variáveis.
  - Diferentes critérios de renomeação podem resultar em diferentes gramáticas na Forma Normal de Greibach. Entretanto, todas são equivalentes (geram a mesma linguagem).

Forma Normal de Greibach (3/6)

- Transformação de produções para a forma $A_r \to A_s\alpha$, onde $r \leq s$:
  - As produções são modificadas garantindo que a primeira variável do lado direito é maior ou igual que a do lado esquerdo, considerando a ordenação da etapa anterior.
  - As produções $A_r \to A_s\alpha$ tais que $r > s$ são modificadas substituindo a variável A_s pelas suas correspondentes produções ($A_s \to \beta_1 \mid \ldots \mid \beta_m$), resultando em $A_r \to \beta_1\alpha \mid \ldots \mid \beta_m\alpha$ e assim sucessivamente.
  - Entretando, como o conjunto de variáveis é finito, existe um limite para as produções crescentes, que pode ser a geração de um terminal ($A_r \to a\alpha$) ou de uma recursão ($A_r \to A_r\alpha$).

Forma Normal de Greibach (4/6)

- Exclusão das recursões da forma $A_r \rightarrow A_r\alpha$:
  - As recursões (à esquerda) podem existir originalmente na gramática, ou serem geradas pela etapa anterior.
  - A eliminação da recursão à esquerda pode ser realizada introduzindo variáveis auxiliares e incluindo recursão à direita ($B_r \rightarrow \alpha B_r$).

Forma Normal de Greibach (5/6)

- Um terminal no início do lado direito de cada produção:
  - Após a execução da etapa anterior, todas as produções da forma $A_r \rightarrow A_s\alpha$ são tais que $r < s$.
  - Consequentemente, as produções da maior variável A_n só podem iniciar por um terminal no lado direito.
  - Assim, se em $A_{n-1} \rightarrow A_n\alpha$ for substituído A_n pelas suas correspondentes produções ($A_n \rightarrow a\beta$), o lado direito das produções de A_{n-1} também iniciarão por um terminal ($A_{n-1} \rightarrow a\beta\alpha$).
  - A repetição do algoritmo para $A_{n-2}, \ldots, A_1$ resultará em produções exclusivamente da forma $A_r \rightarrow a\alpha$.

Forma Normal de Greibach (6/6)

- Produções da forma $A \rightarrow a\alpha$ onde α é composto por variáveis:
  - É análoga a correspondente etapa do algoritmo relativo à Forma Normal de Chomsky.

Transformação de uma GLC na FNG (1/6)

- A transformação da seguinte Gramática Livre do Contexto na correspondente Forma Normal de Greibach é como segue:

  $G = (\{S, A\}, \{a, b\}, P, S)$

  $P = \{ S \rightarrow AA \mid a$

  $\quad A \rightarrow SS \mid b\}$

- Simplificação da Gramática:
  - A gramática já está simplificada.

Transformação de uma GLC na FNG (2/6)

- Renomeação das variáveis em uma ordem crescente qualquer:
  - As variáveis S e A são renomeadas para A_1 e A_2, respectivamente.
  - As produções da gramática ficam como segue:

$$G = (\{A_1, A_2\}, \{a, b\}, P, S)$$

$$P = \{ A_1 \rightarrow A_2A_2 \mid a$$

$$A_2 \rightarrow A_1A_1 \mid b\}$$

Transformação de uma GLC na FNG (3/6)

- Transformação de produções para a forma $A_r \rightarrow A_s\alpha$, onde $r \leq s$:
  - A produção $A_2 \rightarrow A_1A_1$ necessita ser modificada.
  - As produções da gramática ficam como segue:

$G = (\{A_1, A_2\}, \{a, b\}, P, S)$

$P = \{ A_1 \rightarrow A_2A_2 \mid a$

$A_2 \rightarrow A_2A_2A_1 \mid aA_1 \mid b\}$

Transformação de uma GLC na FNG (4/6)

- Exclusão das recursões da forma $A_r \rightarrow A_r\alpha$:
  - A produção $A_2 \rightarrow A_2A_2A_1$ contém uma recursão.
  - Portanto, é necessário introduzir uma variável auxiliar B, como segue:

$$G = (\{A_1, A_2, B\}, \{a, b\}, P, S)$$

$$P = \{ A_1 \rightarrow A_2A_2 \mid a$$

$$A_2 \rightarrow aA_1 \mid b \mid aA_1B \mid bB$$

$$B \rightarrow A_2A_1 \mid A_2A_1B\}$$

Transformação de uma GLC na FNG (5/6)

- Um terminal no início do lado direito de cada produção:
  - O lado direito das produções da maior variável A_2 iniciam por um terminal.
  - Substituindo A_2 no lado direito de A_1 pelas suas correspondentes derivações, as produções de A_1 também iniciarão por um terminal.
  - As produções referentes à variável B também são modificadas:

$$G = (\{A_1, A_2, B\}, \{a, b\}, P, S)$$

$$\begin{aligned} P = \{ \; & A_1 \rightarrow aA_1A_2 \mid bA_2 \mid aA_1BA_2 \mid bBA_2 \mid a \\ & A_2 \rightarrow aA_1 \mid b \mid aA_1B \mid bB \\ & B \rightarrow aA_1A_1 \mid bA_1 \mid aA_1BA_1 \mid bBA_1 \\ & \quad \mid aA_1A_1B \mid bA_1B \mid aA_1BA_1B \mid bBA_1B\} \end{aligned}$$

Transformação de uma GLC na FNG (6/6)

- Produções da forma $A \rightarrow a\alpha$ onde α é composto por variáveis:
  - Nenhum procedimento é necessário, pois as produções já se encontram nesta forma.
- A gramática na Forma Normal de Greibach resultante é a seguinte:

$$G = (\{A_1, A_2, B\}, \{a, b\}, P, S)$$

$$\begin{aligned} P = \{ & A_1 \rightarrow aA_1A_2 \mid bA_2 \mid aA_1BA_2 \mid bBA_2 \mid a \\ & A_2 \rightarrow aA_1 \mid b \mid aA_1B \mid bB \\ & B \rightarrow aA_1A_1 \mid bA_1 \mid aA_1BA_1 \mid bBA_1 \\ & \quad \mid aA_1A_1B \mid bA_1B \mid aA_1BA_1B \mid bBA_1B\} \end{aligned}$$