Propriedades das GLC e Lema do Bombeamento

Prof^a Jerusa Marchi

Departamento de Ciência da Computação

Universidade Federal de Santa Catarina

e-mail: jerusa.marchi@ufsc.br

Forma Normal de Chomsky

- Uma GLC G está na Forma Normal de Chomsky se suas produções estiverem em uma de duas formas:
  - $A \to BC$ onde A, B, C são não terminais; ou
  - $A \to a$ onde A é um não terminal e a é um terminal.
- Além disso, G não tem símbolos inúteis

Forma Normal de Chomsky

- Para transformar uma G em uma GLC na FNC (forma normal de Chomsky)
  1. Eliminam-se as produções vazias ($A \to \varepsilon$)
  2. Eliminam-se as produções unitárias ($A \to B$)
  3. Eliminam-se os símbolos inúteis (improdutivos e inalcançáveis)

Forma Normal de Chomsky

- Fazendo-se isto, G terá dois tipos de produções:
  - $A \rightarrow a$ ou
  - produções cujo corpo possui tamanho maior ou igual a 2
- O processo de transformação é concluído
  1. Fazendo com que todo corpo de produção de tamanho 2 ou mais consista apenas de não terminais
  2. Quebrando as produções de tamanho 3 ou mais em uma cascata de produções, com corpos de até 2 não terminais

Forma Normal de Chomsky

- Fazendo com que todo corpo de produção de tamanho 2 ou mais consista apenas de não terminais
  - Para cada terminal a que apareça no corpo de uma procução com tamanho maior ou igual a 2, crie um novo não terminal A cuja única produção seja $A \to a$
  - Use A no lugar de a em toda produção (≥ 2) em que a aparecer

Forma Normal de Chomsky

- Quebrando as produções de tamanho 3 ou mais em uma cascata de produções, com corpos de até 2 não terminais
  - Para cada produção na forma $A \rightarrow B_1 B_2 \cdots B_k$ para $k \geq 3$, introduza $k-2$ novas variáveis $C_1, C_2, \cdots, C_k - 2$. A produção original é substituida por $k-1$ novas produções da forma:

$$A \rightarrow B_1 C_1, \quad C1 \rightarrow B_2 C_2, \quad \cdots, \quad C_{k-3} \rightarrow B_{k-2} C_{k-2}, \quad C_{k-2} \rightarrow B_{k-1} B_k$$

Forma Normal de Chomsky

- Exemplo

$$E \to E + T \mid T * F \mid (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$

$$T \to T * F \mid (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$

$$F \to (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$

$$I \to a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$

Propriedades das LLC

- Linguagens Livres de Contexto são fechadas sob as seguintes propriedades:
  - União
  - Concatenação
  - Fechamento
  - Reverso

Propriedades das LLC

- Linguagens Livres de Contexto não são fechadas sob as propriedades de:
  - Interseção

$$L_1 = \{a^n b^n c^k \mid n \geq 1, k \geq 0\}$$

$$L_2 = \{a^k b^n c^n \mid n \geq 1, k \geq 0\}$$

$$L_1 \cap L_2$$

  - Complemento

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$$

Lema do Bombeamento para GLC

- Ferramenta usada para demonstrar que algumas linguagens não são livres de contexto
  - Para uma palavra suficientemente grande em um LLC, é possível encontrar duas substrings que podem ser “bombeadas” (repetidas por até i vezes)
  - A string resultante também pertencerá a LLC
- a classe das LR é um subconjunto próprio da classe das LLC, o lema precisa valer para elas, inclusive.
- para estabelecer o lema, precisamos olhar como são as árvores de derivação das LLC

Tamanho da Árvore de Derivação

- Em uma árvore de derivação, as folhas da árvore formam a sentença w
- Seja n a altura da árvore (da raíz às folhas), onde n é o número de derivações.
- Suponha uma árvore de derivação gerada por uma Gramática $G = (N, T, P, S)$ na Forma Normal de Chomsky. Suponha que a sentença nas folhas da árvore é w. Se a altura da árvore é n, então $\mid w \mid \leq 2^{n-1}$

Tamanho da Árvore de Derivação

Teorema: Suponha uma árvore de derivação gerada por uma Gramática $G = (N, T, P, S)$ em uma Forma Normal de Chomsky. Suponha que a sentença nas folhas da árvore é w. Se a altura da árvore é n, então $\mid w \mid \leq 2^{n-1}$

Prova (por indução):

- Para $n = 1$. Apenas uma produção de G foi usada. Como G está em FNC e w consiste de terminais, a produção usada é da forma $S \rightarrow a$. Logo $\mid a \mid = 1$ que é $2^{n-1} = 2^0 = 1$.
- Para $n > 1$. Se $n > 1$ a produção usada foi da forma $A \rightarrow BC$. As subárvores a partir de B e C não podem ter altura maior do que $n - 1$, pois apenas uma aresta foi excluída da raiz para os filhos B e C. Então pela hipótese indutiva, estas duas subárvores geram sentenças de, no máximo, comprimento igual a 2^{n-2} e a sentença da árvore inteira é dada pela soma das subsentenças geradas a partir de B e C, logo $2^{n-2} + 2^{n-2} = 2^{n-1}$.

Lema do Bombeamento para GLC

- Assim como o Lema do Bombeamento para Linguagens Regulares, o Lema do Bombeamento para Linguagens Livres de Contexto é uma ferramenta para demonstrar a não pertinência de alguma linguagem ao conjunto das Linguagens Livre de Contexto.
- A sentença z é particionada em 5 partes ($uvwxy$), onde a 2a. e a 4a. são bombeadas.

Lema do Bombeamento para GLC

Teorema: Seja L uma LLC. Então há uma constante n tal que, se z é qualquer cadeia em L tal que $\mid z \mid$ é pelo menos n, então pode-se dividir $z = uvwxy$ sujeito as seguintes condições:

1. $\mid vwx \mid \leq n$, ou seja a porção do meio da sentença não é tão longa
2. $vx \neq \varepsilon$, uma vez que v e x são as partes bombeadas, então pelo menos uma delas não pode ser vazia.
3. Para todo $i \geq 0$, uv^iwx^iy está em L, isto é, as duas cadeias v e x podem ser "bombeadas" qualquer número de vezes, incluindo 0 e a cadeia resultante deve pertencer a linguagem L.

Lema do Bombeamento para GLC

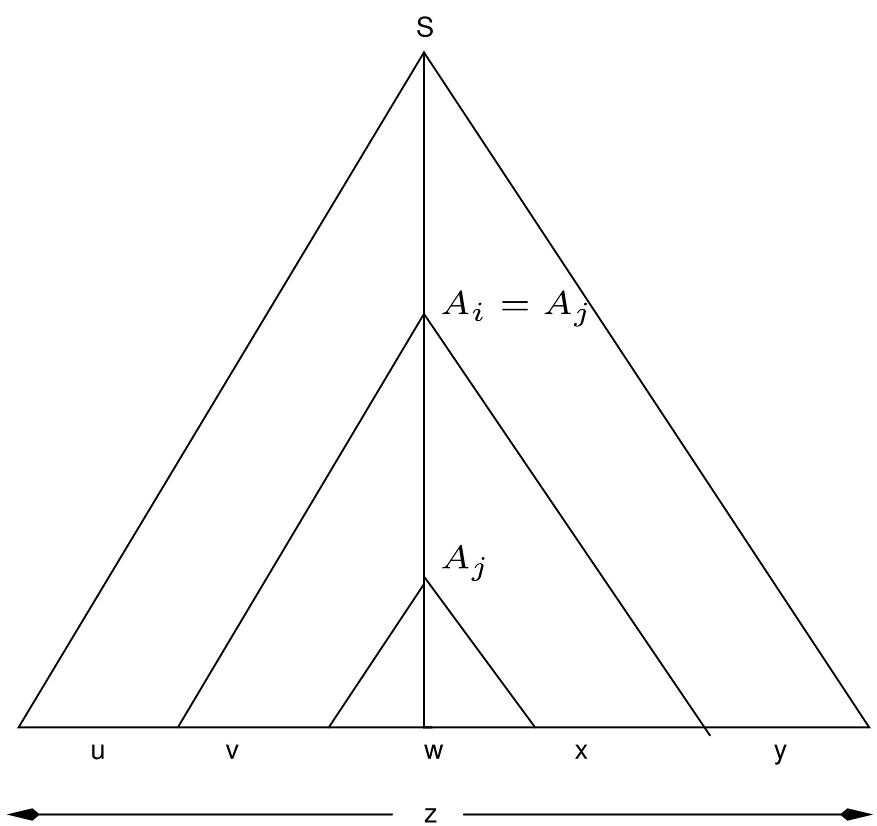
Prova: O primeiro passo é obter uma gramática em uma FNC para a linguagem L. Tecnicamente, a linguagem não pode ser $\emptyset$ ou $\{\varepsilon\}$.
Contudo, se L = $\emptyset$ então não há nenhum $z \in L$ e portanto, o teorema foi violado.
Da mesma forma, a gramática G na FNC gerará $L - \{\varepsilon\}$, mas isso não importa, pois devemos tomar um $n > 0$, e portanto z não pode ser ε.

Lema do Bombeamento para GLC

Prova (cont): Agora, iniciando com uma gramática $G = (N, T, P, S)$ na FNC tal que L(G) = L - $\{\varepsilon\}$, deixe G ter m variáveis. Escolha $n = 2^m$. Em seguida, suponha que $z \in L$ tem tamanho pelo menos igual a n. Pelo Teorema que estabelece a altura da árvore de derivação, qualquer árvore de derivação cujo caminho mais longo tem tamanho m ou menor tem abertura de tamanho $2^{m-1} = n/2$ ou menor. Esta árvore não pode ter gerado z, pois z é muito longa. Então, qualquer árvore que gere z tem um caminho de comprimento pelo menos igual a $m + 1$.

Lema do Bombeamento para GLC

Prova (cont): Portanto, há pelo menos, $m+1$ ocorrências das variáveis $A_0, A_1, A_2, ..., A_k$ no caminho. Como há somente m variáveis diferentes em V, pelo menos duas das últimas $m+1$ variáveis no caminho, são a mesma variável. Suponha que $A_i = A_j$, onde $k - m \leq i < j \leq k$. Então haverá pelo menos 2 subárvores com raiz A_j.



Lema do Bombeamento para GLC

Prova (cont):Então a string w é gerada na subárvore enraizada em A_j. As strings v e x são strings à esquerda e à direita de w na geração de uma string maior com raíz em A_i. Como não há produções unitárias v e x não podem ser ambas ε, ainda que uma possa (condição 2 do lema). Finalmente, u e y são partes de z posicionadas mais à esquerda e mais à direita com raiz em A_i. Se $A_i = A_j = A$ então pode-se construir novas árvores sintáticas a partir da original.

Lema do Bombeamento para GLC

Prova (cont): Primeiro, pode-se substituir a árvore original enraizada em A_i, que gera vwx pela subárvore com raiz em A_j que gera w, o que resulta na string uwy e corresponde ao caso quando $i = 0$ no padrão uv^iwx^iy. Outra possibilidade é substituir a subárvore enraizada em A_j pela árvore enraizada em A_i, resultando em uv^2wx^2y e assim por diante para cada expoente i. (condição 3 do lema)

Lema do Bombeamento para GLC

Prova (cont): Ainda falta a condição 1 que diz que $| vwx | \leq n$. Nós pegamos A_i para ser próxima a base da árvore, isto é, $k - i \leq m$. Então o caminho mais longo na subárvore enraizada em A_i não é maior do que $m + 1$. Pelo teorema da altura da árvore, a subárvore enraizada em A_i tem uma abertura de tamanho no máximo igual a $2^m = n$.

Aplicação do Lema do Bombeamento

- Assim como o lema para linguagens regulares, as provas são desenvolvidas por contradição. Seja

$$L = \{0^n 1^n 2^n \mid n \geq 1\}$$

  - Suponha para fins de obter uma contradição que L é uma LLC. Então há um inteiro n (comprimento do bombeamento) para o qual a linguagem pode ser bombeada. Escolha $z = 0^n 1^n 2^n$. Se $z = uvwxy$, onde $\mid vwx \mid \leq n$ e v e x não são ambos vazios. Então é sabido que vwx não pode envolver 0's e 2's pois há entre eles pelo menos $n + 1$ posições.

Aplicação do Lema do Bombeamento

- Para demonstrar que ao bombear v e x a cadeia resultante não pertence a L, os casos possíveis são:
  1. vwx não possui 2's. Então vx consiste de somente 0's e 1's e tem pelo menos um destes símbolos (v e x não podem ser simultaneamente iguais a ε). Então uwy, que deveria pertencer a L pela aplicação do lema do bombeamento, tem n 2's, mas um número inferior a n de 0's ou 1's ou ambos. Portanto L não é uma LLC.
  2. vwx não tem 0's. Então, similarmente, vx tem n 0's, mas menos 1's ou 2's ou ambos.Portanto L não é uma LLC.

Aplicação do Lema do Bombeamento

- Para $L = 0^i 1^j 2^i 3^j \mid i \geq 1$ e $j \geq 1\}$
  - Suponha para fins de obter uma contradição que L é uma LLC. Então seja n o comprimento do bombeamento. Escolha $z = 0^n 1^n 2^n 3^n$. Se $z = uvwxy$, onde $\mid vwx \mid \leq n$ e $vx \neq \varepsilon$. Então vwx está contido em uma substring de um símbolo ou em uma substring de 2 símbolos adjacentes.
  - Se vwx consiste de somente um símbolo, então uwy tem n de três símbolos e menos que n do 4^o símbolo. Portanto $z \notin L$
  - Se vwx consiste em 2 símbolos adjacentes, digamos 1's e 2's, então uwy terá menos 1's ou 2's ou ambos. Suponha a ausência de 1's. Como há n 3's, esta cadeia não pode estar em L. Similarmente na ausência de 2's, haverão n 0's e uwy não pode estar em L. Portanto L não é uma LLC.