The Pumping Lemma for CFLs

Mark Greenstreet, CpSc 421, Term 1, 2008/09

- A language that is not context free: $a^n b^n c^n$.
- A pumping lemma for context free languages
- Examples

$\mathbf{a}^n\mathbf{b}^n\mathbf{c}^n$ is not context free

Proof strategy:

- Let $A = \{w \mid \exists n \in \mathbb{Z}^{\geq 0}.\ w = \mathsf{a}^n\mathsf{b}^n\mathsf{c}^n\}$.
- A is not context free.
- Use proof by contradiction.
  - Assume $G = (V, \Sigma, R, S_0)$ be a CNF CFG that supposedly generates A (note: $\Sigma = \{\mathsf{a}, \mathsf{b}, \mathsf{c}\}$).
- Pick a really big n (how big depends on G) and consider the parse tree, $\mathcal{T}$ for a derivation for $w = \mathsf{a}^n\mathsf{b}^n\mathsf{c}^n$.
- We'll take a closer look at this parse tree and derivation on the next slide.

Deriving w

- CNF parse-trees are binary-trees, with nodes of degree-1 at the level just above the leaves.
- Thus, for $k \geq 2$, a CNF parse-tree of height k has at most 2^{k-2} leaves.
- If we choose n large enough, we can find a vertex, v of of $\mathcal{T}$ (the parse tree for $w = \mathtt{a}^n\mathtt{b}^n\mathtt{c}^n$) where:
  - the height of v is $|G| + 1$, and
  - each leaf of the subtree rooted at v is either an a or a b.
- Because v is at height $|G| + 1$, we can find a path from v to a leaf that has two nodes, u and u' that are labeled with the same variable from V.
- We can replace the subtree rooted at u' with the subtree rooted at u to produce a new parse tree, $\mathcal{T}'$.
  - Note that parse tree $\mathcal{T}'$ has more leaves for a's and/or b's than tree $\mathcal{T}$;
  - however, both parse-trees have the same number of leaves for c's.
  - Therefore, the strings produced by $\mathcal{T}$ and $\mathcal{T}'$ cannot both be in A.
- $\therefore A$ is not regular.

The Pumping Lemma for CFLs

Let A be a CFL. There is a constant p (that depends on the choice of A) such that for any string $s \in A$ with $|s| \geq p$, there are strings u, w, x, y and z such that:

- $s = uwxyz$;
- $|wxy| \leq p$;
- $|wy| \geq 1$;
- For all $i \in \mathbb{Z}^{\geq 0}$, $uw^ixy^iz \in A$.

As with the pumping lemma for regular languages, we will typically use the contrapositive form to show that a language is not context free.

- If we can show that for any choice of p, there is some string $s \in A$ with $|s| \geq p$ such that at least one of the four conditions listed above is violated,
- Then A is not context free.

Proof strategy

- We'll assume a CNF grammar for A.
  Thus, it's parse-trees are binary trees.
- If we make p big enough, then parse tree must be fairly high (roughly $\log_2(p)$ to generate a string of length p).
- Consider a path from S_0 to a leaf that is longer than $|G|$.
  This path must encounter two nodes that correspond to the same variable, $u \in V$.
- These two subtrees that are derivations from u are "interchangable".

  This gives us what we need for pumping.

Height of a derivation

- Let $G = (V, \Sigma, R, S_0)$ be a CNF CFG.
- Let $\Psi = V \cup \Sigma$ be the union of the variables and terminals.
- Let $\alpha \in \Psi^+$ and $s \in \Sigma^*$ such that $\alpha \overset{*}{\Rightarrow} s$.
- We define $height(v, s)$ as shown below:

$$
\begin{array}{lcll}
height(\mathsf{v}, \mathsf{c}) & = & 1, & (v \Rightarrow \mathsf{c}) \in R \\
 & & & ((v \rightarrow v_1\, v_2) \in R) \\
height(v, s) & = & 1 + \max\limits_{v_1, v_2, s_1, s_2, i} height(v_i, s_i), & \wedge\, (s = s_1, s_2) \,\wedge\, (i \in \{1, 2\}) \\
 & & & \wedge\, (v_1 \overset{*}{\Rightarrow} s_1) \,\wedge\, (v_2 \overset{*}{\Rightarrow} s_2)
\end{array}
$$

  - If there is more than one way for α to derive s, our definition defines the height to be the “tallest” one. This doesn’t really matter for the way we’ll use $height$ when proving the pumping lemma.
  - Because G is CNF, $S_0 \overset{|s|+1}{\Rightarrow} s$ for any string of length two or longer. In particular, there are no unbounded derivations. Thus, the $height$ is bounded and well-defined.

Properties of $height$ (1 of 2)

Choose any $v \in V$ and $s \in \Sigma^*$ such that $v \stackrel{*}{\Rightarrow} s$ and $height(v, s) = h$.

- Property 1: $|s| \leq 2^{h-1}$. Equivalently, $h \geq \lceil \log_2 |s| \rceil + 1$.

  Proof: By induction on h.

  It's a consequence of the fact that parse-trees for CNF grammars are binary trees (with degree one nodes for the unit rules that produce the terminals at the end).

Properties of *height* (2 of 2)

As before, choose any $v \in V$ and $s \in \Sigma^*$ such that $v \stackrel{*}{\Rightarrow} s$ and $height(v, s) = h$.

- Property 2 (Paths): There are strings $\alpha_1 \ldots \alpha_h$ and $\beta_1 \ldots \beta_h$ in Ψ^*, and variables $v_1 \ldots v_h \in V$ such that
  - $\alpha_1 = \beta_1 = \epsilon$ and $v_1 = v$.
  - For each $i \in 2 \ldots h$ there is a rule $v_{i-1} \Rightarrow u_1 u_2$ in R such that either
    - $v_i = u_1$, $\alpha_i = \alpha_{i-1}$ and $\beta_i = u_2 \beta i - 1$; or
    - $v_i = u_2$, $\alpha_i = \alpha_{i-1} u_1$ and $\beta_i = \beta i - 1$.
  - For $1 \leq i < j \leq h$, there are strings $x, y \in \Psi+$ such that

    $$\alpha_i v_i \beta_i \stackrel{j-i}{\Rightarrow} \alpha_j v_j \beta_j$$

    with $\alpha_j = \alpha_i x$ and $\beta_j = y \beta_i$.
  - $\alpha_h v_h \beta_h \stackrel{*}{\Rightarrow} s$.
  - Proof: by induction on h, but you could say that "By the definition of $height$, there is a path of length h in the parse tree from the root, v, to a leaf." and get full-credit.

Proof strategy (refined)

- Let $p = 2^{|V|}$. Let s be any string in A with $|s| \geq p$.
- From the Property 1 of $height$, $height(S_0.s) \geq |V| + 1$.
- From the Property 2 of $height$, there must be a path from S_0 to a terminal in s that passes throgh at least $|V| + 1$ variables.
- By the pigeon-hole principle, two of these parse-tree nodes must be labeled with the same variables.
- The two subtrees rooted at these nodes are interchangeable. We're ready to pump. ☺

Proof of the pumping lemma

Let $p = 2^{|V|}$. Let s be any string in A with $|s| \geq p$.

- From the Property 1 of $height$, $height(S_0.s) \geq |V| + 1$. Let $h = |V| + 1$.
- Need to show strings u, w, x, y and z such that: $s = uwxyz$; $|wxy| \leq p$; $|wy| \geq 1$; and $\forall i \in \mathbb{Z}^{\geq 0}.\ uw^ixy^iz \in A$.
- Let $\alpha_1 \ldots \alpha_h$ and $\beta_1 \ldots \beta_h$ be strings in Ψ^*, and $v_1 \ldots v_h$ be variables in V according to Property 2 of $height$.
- By the pigeon hole principle, we can find $i, j \in 1 \ldots h$ with $i \neq j$ and $v_i = v_j$.
- We now have

$$\begin{array}{rcl} S_0 & \overset{i-1}{\Rightarrow} & \alpha_i v_i \beta_i \\ & \overset{j-i}{\Rightarrow} & \alpha_i w v_j y \beta_j \\ & \overset{*}{\Rightarrow} & s \end{array}$$

- Let u, x and z be strings such that $\alpha_i \overset{*}{\Rightarrow} u$, $v_j \overset{*}{\Rightarrow} x$, $\beta_i \overset{*}{\Rightarrow} z$, and $s = xyz$.
- We'll show on the next slide that for all u, w, x, y and z as defined above satisfy the conditions of the pumping lemma.

Example: ww is not a CFL

Game with an adversary

Example: $a^n b^{n^2}$ is not a CFL

Closing remarks