The Pumping Lemma

The Lemma & Decision/Closure Properties

Before We Start

- Any questions?

Languages

- Future Exam Question
  - What is a language?
  - What is a class of languages?

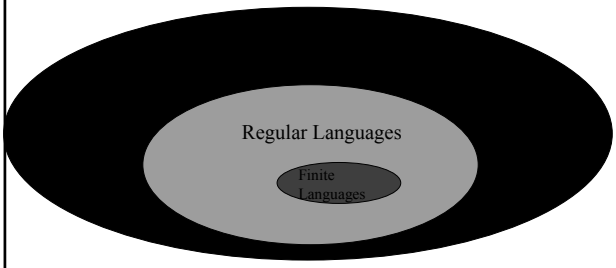
Context Free Languages

- Context Free Languages(CFL) is the next class of languages outside of Regular Languages:
  - Means for defining: Context Free Grammar
  - Machine for accepting: Pushdown Automata

Plan for today

- The Return of the Pumping Lemma
- Closure Properties and Decision Properties for CFLs

Now our picture looks like



Is there anything out here?

Just when you thought it was safe

- Return of the Pumping Lemma
  - But before we start that!
  - When we last left our CFGs…

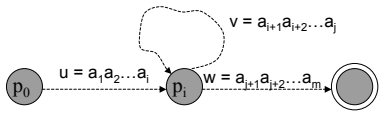
Chomsky Normal Form

- Chomsky Normal Form
  - A context free grammar is in Chomsky Normal Form (CNF) if every production is of the form:
    - $A \rightarrow BC$
    - $A \rightarrow a$
    - Where A,B, and C are variables and a is a terminal.

The Pumping Lemma for RL

- Statement of the pumping lemma for RL
  - Let L be a regular language.
  - Then there exists a constant n (which varies for different languages), such that for every string $x \in L$ with $|x| \geq n$, x can be expressed as x = uvw such that:
    1. $|v| > 0$
    2. $|uv| \leq n$
    3. For all $k \geq 0$, the string uv^kw is also in L.

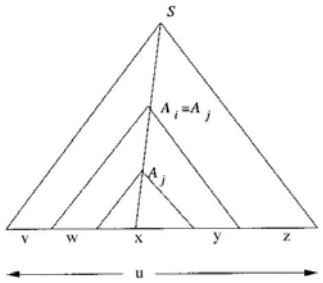
The Pumping Lemma for RL



The Pumping Lemma for CFLs

- With CFLs
  - strings are distinguished by their derivation (or parse trees) based on the productions of a CFG.
  - The idea behind the Pumping Lemma for CFLs:
    - If a string is long enough, then at least one variable in it's derivation will have to be repeated.
    - We can repeatedly reapply productions for the repeated variable ("pump you up") and the resultant string will also be in the language

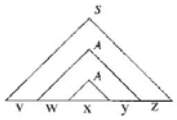
The Pumping Lemma for CFLs

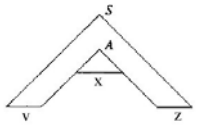


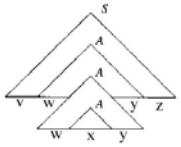
The Pumping Lemma for CFLs

- $S \Rightarrow^* vAz \Rightarrow^* vwAyz \Rightarrow^* vwxyz$
  - So $A \Rightarrow^* x$ but also $A \Rightarrow^* wAy$
  - We can then write instead:
    - $S \Rightarrow^* vAz \Rightarrow^* vwAyz \Rightarrow^* vw^2Ay^2z \Rightarrow^* vw^3Ay^3z$
    - And so on…

The Pumping Lemma for CFLs



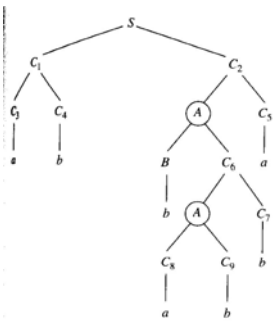




The Pumping Lemma for CFLs

- How long is long enough?
  - Recall Chomsky Normal Form
    - A context free grammar is in Chomsky Normal Form (CNF) if every production is of the form:
      - $A \rightarrow BC$
      - $A \rightarrow a$
    - Where A,B, and C are variables and a is a terminal.

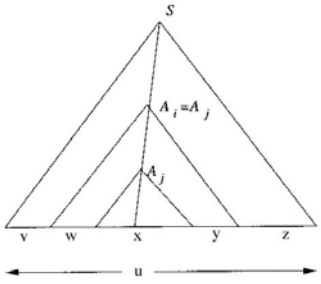
The Pumping Lemma for CFLs



The Pumping Lemma for CFLs

- The parse tree for a grammar in CNF will be a binary tree
  - A binary tree having more than 2^{k-1} leaf nodes must have a height (longest path) > k.
  - If we let k be the number of number of variables in our grammar, then
    - For any string x, where $|x| > 2^k$
    - At least one variable in the longest path will be repeated.

The Pumping Lemma for CFLs



The Pumping Lemma for CFLs

- Let L be a CFL. Then there is an integer n so that for all strings u, where $|u| \geq n$, u can be expressed as u = vwxyz where
  - $|wy| > 0$
  - $|wxy| \leq n$
  - For any $m \geq 0$, $vw^m xy^m z \in L$
- $n = 2^{p+1}$ where p = number of variables

The Pumping Lemma for CFLs

- The real strength of the pumping lemma is proving that languages are not context free
  - Proof by contradiction
    - Assume that the language to be tested is a CFL
    - Use the pumping lemma to come to a contradiction
    - Original assumption about the language being a CFL is false
- You cannot prove a language to be a CFL using the Pumping Lemma!!!!

The Pumping Lemma for CFLs

- The Pumping Lemma game
  - To show that a language L is not a CFL
    - Assume L is context free
    - Choose an "appropriate" string x in L
    - Express x = uvwxy following rules of pumping lemma
    - Show that $uv^k wx^k z$ is not in L, for some k
    - The above contradicts the Pumping Lemma
    - Our assumption that L is context free is wrong
    - L must not be context free

The Pumping Lemma for CFLs

- Example:
  - $L = \{ a^i b^i c^i \mid i \geq 1 \}$
  - Strings of the form abc where number of a's, b's and c's are equal
  - Let's play!
  - Assume that L is context free. Then by the pumping lemma all strings u with $|u| \geq n$ can be expressed as u = vwxyz and
    - $|wy| > 0$
    - $|wxy| \leq n$
    - For any $m \geq 0$, $vw^m xy^m z \in L$

The Pumping Lemma for CFLs

- Example
  - $L = \{ a^i b^i c^i \mid i \geq 1 \}$
  - Choose an appropriate $u = a^n b^n c^n = vwxyz$
  - Since $|wxy| \leq n$ then wxy must consists of
    - All a's or all b's or all c's
    - Some a's and some b's
    - Some b's and some c's

The Pumping Lemma for CFLs

- In all three cases
  - vw^2xy^2z will not have an equal number of a's b's and c's.
  - Pumping Lemma says $vw^2xy^2z \in L$
  - Can't contradict the pumping lemma!
  - Our original assumption must be wrong.
  - L is not context-free.

The Pumping Lemma for CFLs

- By the same argument (same choice of u), we can show that:
  - $L = \{ x \in \{a,b,c\}^* \mid n_a(x) = n_b(x) = n_c(x) \}$
- Is not context free

The Pumping Lemma for CFLs

- Another Example:
  - $L = \{ a^i b^j c^k \mid i < j \text{ and } i < k \}$
  - Number of a's is less than the number of b's and the number of c's
  - Let's play!
  - Assume that L is context free. Then by the pumping lemma all strings u with $|u| \geq n$ can be expressed as u = vwxyz and
    - $|wy| > 0$
    - $|wxy| \leq n$
    - For any $m \geq 0$, $vw^m xy^m z \in L$

The Pumping Lemma for CFLs

- Example
  - $L = \{ a^i b^j c^k \mid i < j \text{ and } i < k \}$
  - Choose an appropriate $u = a^n b^{n+1} c^{n+1} = vwxyz$
  - Since $|wxy| \leq n$ then wxy must consists of
    - Case 1: All a's or all b's or all c's
    - Case 2: Some a's and some b's
    - Case 3: Some b's and some c's

The Pumping Lemma for CFLs

- Let's consider each case individually:
  - Case 1: All a's or all b's or all c's
    - If wxy consists of all a's then there will be k such that when we pump w and y k times, the number of a's will be greater than n+1
    - If wxy consists of all b's then vw^0xy^0z will contain the same number or less b's than a's
    - If wxy consists of all c's then vw^0xy^0z will contain the same number or less c's than a's

The Pumping Lemma for CFLs

- Let's consider each case individually:
  - Case 2: Some a's and some b's
    - If wxy consists of only a's and b's then there will be k such that when we pump w and y k times, the number of a's will be greater than n+1 (# of c's)
    - Relationship between a's and b's might be maintained, but not the relationship between a's and c's

The Pumping Lemma for CFLs

- Let's consider each case individually:
  - Case 2: Some b's and some c's
    - If wxy consists of only b's and c's then vw^0xy^0z will contain the same number or less c's or b's than a's

The Pumping Lemma for CFLs

- In all cases
  - We found a "pumped" (or unpumped) string that the pumping lemma said should be in the langauge but did not maintain the relationship of a's to b's and c's as specified in the language.
  - Can't contradict the pumping lemma!
  - Our original assumption must be wrong.
  - L is not context-free.

The Pumping Lemma for CFLs

- By the same argument (same choice of u), we can show that:
  - $L = \{ x \in \{a,b,c\}^* \mid n_a(x) < n_b(x) \text{ and } n_a(x) < n_c(x) \}$
- Is not context free

The Pumping Lemma for CFLs

- Questions?

Closure Properties

- We already seen that CFLs are closed under:
  - Union
  - Concatenation
  - Kleene Star
- Regular Languages are also closed under
  - Intersection
  - Complementation
  - Difference
- What about Context Free Languages?

Closure Properties

- Sorry, Charlie
  - CFLs are <u>not</u> closed under intersection
  - Meaning:
    - If L_1 and L_2 are CFLs then $L_1 \cap L_2$ is not necessarily a CFL.

Closure Properties

- CFLs are <u>not</u> closed under intersection
  - Example:
    - $L_1 = \{a^i b^j c^k \mid i < j \}$
    - $L_2 = \{a^i b^j c^k \mid i < k \}$
    - Are both CFLs

Closure Properties

- CFLs are not closed under intersection

$L_1 = \{a^i b^j c^k \mid i < j\}$	$L_2 = \{a^i b^j c^k \mid i < k\}$
$S \rightarrow ABC$	$S \rightarrow AC$
$A \rightarrow aAb \mid \varepsilon$	$A \rightarrow aAc \mid B$
$B \rightarrow bB \mid b$	$B \rightarrow bB \mid \varepsilon$
$C \rightarrow cC \mid \varepsilon$	$C \rightarrow cC \mid c$

Closure Properties

- CFLs are not closed under intersection
  - $L_1 \cap L_2 = \{a^i b^j c^k \mid i < j \text{ and } i < k\}$
  - Which we just showed to be non-context free.

Closure Properties

- Sorry, Charlie
  - CFLs are not closed under complement
  - Why?
    - $L_1 \cap L_2 = (L_1' \cup L_2')'$

Closure Properties

- Sorry, Charlie
  - CFLs are not closed under difference
  - Why?
    - $L' = \Sigma^* - L$
    - We know Σ^* is regular, and as such is also a CFL.
    - If CFLs were closed under difference, then $\Sigma^* - L = L'$ would always be a CFL
    - But we showed that CFLs are not closed under complement

Closure Properties

- What went wrong?
  - Can't we apply the same construction as we did for the complement of RLs?
    - Reverse the accepting / non-accepting states
    - PDAs can "crash".
      - I.e Fail by having no place to go.
      - PDAs can "crash" in accepting or non-accepting state
      - Making non-accepting states accepting will not handle crashes.

Closure Properties

- What went wrong?
  - Can't we apply the same construction as we did for the intersection of RLs?
    - The states of M are an ordered pair (p, q) where $p \in Q_1$ and $q \in Q_2$
    - Informally, the states of M will represent the current states of M_1 and M_2 at each simultaneous move of the machines.

Closure Properties

- What went wrong?
  - Can't we apply the same construction as we did for the intersection of RLs?
    - The problem is the stack.
    - Although we could try the same thing for PDAs and have a combined machine keep track of where both PDAs are at any one time.
    - We can't keep track of what's on both stacks at any given tine.

Closure Properties

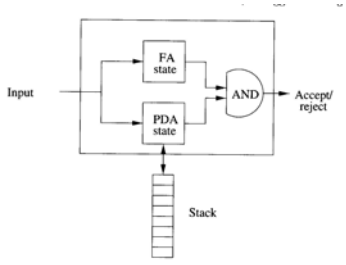
- However, if one of the CFLs does not use the stack (I.e. it is an FA), then we can build a PDA that accepts $L_1 \cap L_2$.
- In other words:
  - If L_1 is a context free language and L_2 is a regular language, then $L_1 \cap L_2$ is context free.

Closure Properties

- Basic idea:
  - Like with the FA construction, let the states of the new machine keep track of the states of the PDA accepting L_1 (M_1) and the FA accepting L_2 (M_2).
  - Our single stack of the new machine will operate the same as the stack of the PDA accepting L_1
  - Accepting states will be all states that contain both an accepting state from M_1 and M_2.

Closure Properties

- Basic idea



Closure Properties

- Summary
  - CFLs are closed under
    - Union, Concatenation, Kleene Star
  - CFLs are NOT closed under
    - Intersection, Difference, Complement
  - But
    - The intersection of a CFL with a RL is a CFL

Decision Properties

- Questions we can ask about context free languages and how we answer such questions.

Decision Properties

- Given regular languages, specified in any one of the four means, can we develop algorithms that answer the following questions:
  1. Is a given string in the language?
  2. Is the language empty?
  3. Is the language finite?

Decision Properties

- Membership
  - Unlike FAs, we can't just run the string through the machine and see where it goes since PDAs are non-deterministic.
    - Must consider all possible paths

Decision Properties

- Membership
  - Instead, start with your grammar in CNF.
    - The proof of the pumping lemma states that the longest derivation path of a string of size n will be $2n - 1$.
    - Systematically generate all derivations with one step, then two steps, …, then $2n - 1$ steps where the length of the string tested = n. If one of the derivations derive x, return true, else return false.

Decision Properties

- Emptiness
  - By the proof of the pumping lemma, if a grammar in CNF has p states, the longest string, not subject to the pumping lemma will have length $n = 2^{p+1}$.
    - Systematically generate all strings with lengths less than n.
    - Test each one using membership algorithm
    - If all fail and $\varepsilon \notin L$, then L is empty
    - Else L is not empty.

Decision Properties

- Finiteness
  - Just as with RLs, a language is infinite if there is a string x with length between n and 2n
    - With RLs n = number of states in an FA
    - With CFLs $n = 2^{p+1}$ where p is the number of variables in the CFG
    - Systematically generate all strings with lengths between n and 2n
    - Run through membership algorithm
    - If one passes, L is infinite, if all fail, L is finite

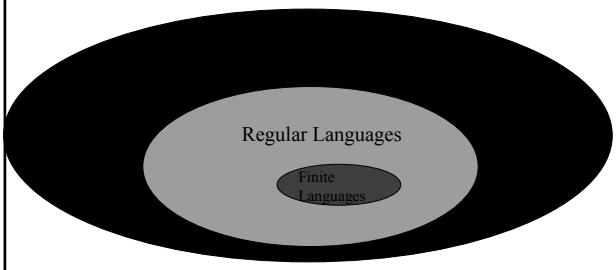
Decision Properties

- Questions?

Summary

- Pumping Lemma for CFLs
- Closure Properties
- Decision Properties

Now our picture looks like



Is there anything out here? YES

Next Time

- Next classes of languages
- However,
  - We start with the machine rather than the language
  - Move beyond simple language acceptance into the realm of computation.
- Enter…The Turing Machine!!!