CHAPTER 2

Context-Free Languages

Contents

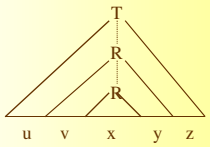
- Context-Free Grammars
  - definitions, examples, designing, ambiguity, Chomsky normal form
- Pushdown Automata
  - definitions, examples, equivalence with context-free grammars
- **Non-Context-Free Languages**
  - **the pumping lemma for context-free languages**

Non-Context-Free Languages

- Is the language $L = \{a^m b^n c^k : \ n = m = k\}$ context-free?
- First attempts at constructing a PDA for this language seem to fail
  - One can read the *a*'s, push them on the stack and match them with the *b*'s while popping them off, but then there is no way to check that the number of *c*'s equals the number of *b*'s.
  - Alternatively, one could push two *a*'s on the stack upon reading each *a*, and match the *b*'s and finally the *c*'s while popping off the *a*'s. But this method can only check that the total number of *b*'s and *c*'s is exactly $2n$; it cannot check that the number of *b*'s is exactly n.
- Of course, just because these simple ideas do not work, we cannot conclude that the language is not context-free.
- We will present a technique for proving that certain languages are not context-free.
- We will present a ***Pumping Lemma for context-free languages*** (which is similar to the Pumping Lemma for regular languages).
- It states that every context-free language has a special value called ***pumping length*** such that all longer strings in the language can be "pumped".
- That is the string can be divided into five parts so that the second and the fourth parts may be repeated together any number of times and the resulting string still remains in the language.

Pumping Lemma

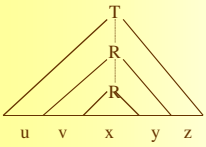
Lemma. **For any** context-free language L,
there exists a number $p \geq 1$ such that
for every word $w \in L$ with at least p letters
there exist u, v, x, y, z with $w = uvxyz$ and
$|vy|>0$ and $|vxy| \leq p$ such that
for every number $i \geq 0$, $uv^i xy^i z \in L$.



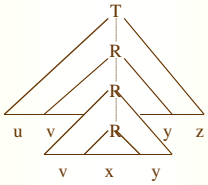
- We call p **the pumping length** of L, and $uvxyz$ **the pumping decomposition** of w.

Proof

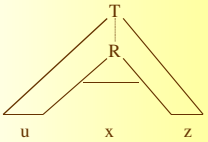
- Let G be a CFG in Chomsky N.F. for CFL L and $|V|$ be the number of variables in G.
- Then all parse trees are binary trees.
- Set p to be $2^{|V|+1}$.
- Now consider a word w in L with at least p letters.
- The parse tree of w has to have at least height $|V|+1$.
- We choose a parse tree that have the smallest number of nodes. (!!!)
- Some branch of the parse tree of w must be a sequence of at least $|V|+1$ nonterminals.
- Hence some variable R must occur on the path from the root to a leaf more than once.
- We select R to be a variable that repeats among the lowest $|V|+1$ variables on this path. (!!!!!!)
- We divide w into $uvxyz$ according to figure.



$uvxyz$



uv^2xy^2z



uxz

- Replacing the smaller subtree by the larger one repeatedly gives parse trees for the strings $uv^i xy^i z$ (i>1).
- Replacing the larger by smaller generates the string uxz.
- Why $|vy|>0$ and $|vxy| \leq p$? (see !!! and !!!!!!)

How to prove that a language is not context-free?

- Suppose we want to prove that a language L is not context-free.
- We can do this by showing that the pumping lemma does not hold for L; that is, we prove the negation of the pumping lemma:

for any number $p \geq 1$
there exists a word $w \in L$ with at least p letters such that
for all u, v, x, y, z with $w = uvxyz$ and $|vy|>0$ and $|vxy| \leq p$
there exists a number $i \geq 0$ such that $uv^i xy^i z \notin L$.

- We have to consider all possibilities for the pumping number p,
- All possibilities for the pumping decomposition u,v,x,y,z (often by case analysis).
- But we are free to choose a single word w,
- And a single iteration number i.
- Choosing a suitable w is usually the crux of the proof (one needs a bit of creative thinking)
- For i, we can typically choose $i=0$ or $i=2$.

Examples

- ***Example 1:*** $L = \{a^m b^n c^k : \ n = m = k\}$ is not context-free.

- Choose any pumping number p (we know only that $p \geq 1$).
- Choose $w = a^p b^p c^p$.
- Consider any pumping decomposition $w=uvxyz$ ($|vy|>0$ and $|vxy| \leq p$).
- Since $|vxy| \leq p$ there are three possibilities:

  (a) vxy contains no a's;
  (b) vxy contains no b's;
  (c) vxy contains no c's.

- Choose $i=2$. We need to show that uv^2xy^2z is not in L.

  (a) uv^2xy^2z contains either more b's than a's, or more c's than a's;
  (b) uv^2xy^2z contains either more a's than b's, or more c's than b's;
  (c) uv^2xy^2z contains either more a's than c's, or more b's than c's.

- ***Example 2:*** $L' = \{ww : w \in \{0,1\}^*\}$ is not context-free.
- Choose $w = 0^p 1^p 0^p 1^p$.