CS:4330 Theory of Computation
Spring 2018

Context-Free Languages

Non-Context-Free Languages

Haniel Barbosa

The University of Iowa

Readings for this lecture

Chapter 2 of [Sipser 1996], 3rd edition. Section 2.3.

Proving context-freeness

- We will use a similar mechanism as with regular languages
- Pumping lemma for context-free languages states that every CFL has a specif value, called *pumping length*, such that all longer strings in the language can be pumped
- However, the meaning of pumping is a bit more complex than in the case of regular languages
- Here pumping means that a string can be divided into five parts so that the second and fourth parts may be repeated any number of times and the resulting string is in the language

Pumping lemma for CFLs

Theorem

If A is a context-free language, then there is a number p (the pumping length) such that if $s \in A$ and $|s| \geq p$, then s may be divided into five pieces, $s = uvxyz$, satisfying the conditions:

1. *For each $i \geq 0$, $uv^ixy^iz \in A$*
2. $|vy| \geq 0$
3. $|vxy| \leq p$

Interpretation

- ▷ Condition 1 states that the length of strings of A can be unlimited but have a fixed structure $uv^ixy^iz,\ i \geq 0$
- ▷ Condition 2 says that in the structure uv^ixy^iz either v or y is not empty, otherwise the theorem would be trivially true
- ▷ Condition 3 states that pieces v, x, y together have length at most p

Proof idea

Let A be a CFL and G be a CFG generating A. We must show that any sufficiently long $s \in A$ can be pumped and remain in A. The idea behind this proof is:

1. Because $s \in A$, it is derivable from G and has a derivation tree D_s
2. Since s is *very long*, the derivation tree D_s of s must be *very tall*
3. This means that D_s contains some long path from the start nonterminal of G at the root to one of the terminal symbols at a leaf.
4. On this long path, some nonterminal symbol X must be repeated due to the pigeonhole principle

Note

- ▷ The repetition of X in D_s allows us to replace subtrees under X and still get a legal derivation tree.
- ▷ Hence, we may cut s into five pieces and we may repeat the second and fourth pieces and obtain a string $uv^ixy^iz \in A$, for any $i \geq 0$.

Proof

Let $G = (V, \Sigma, R, S)$ be a CFG that generates A. Let $b \geq 2$ be the maximum number of symbols in the *rhs* of rules in R.

- ▷ In any derivation tree D using G we know that a node cannot have more than b children, i.e. at most b leaves in 1 step from S; at most b^2 leaves in 2 steps; at most b^h in h steps
- ▷ If the height of D is at most h the length of the string generated is at most b^h
- ▷ Conversely, if a string is at least b^{h+1} long, each of its derivation trees must be at least $h+1$ high
- ▷ With $|V|$ being the number of nonterminals in G, then we set $p = b^{|V|+1}$

Consequently, a derivation tree of any string from A of length at least p requires a height at least $|V| + 1$.

Proof, continuation

Suppose $s \in A$, $|s| \geq p$. We will show how to pump s

- ▷ Let D_s be the derivation tree of s; if s has several derivation trees we choose D_s to be the tree with the smallest number of nodes.
- ▷ Since $|s| \geq p$ we know that D_s has height at least $|V| + 1$, so the longest path in D_s has length at least $|V| + 1$
- ▷ The longest path in D_s has at least $|V| + 1$ nodes labeled by nonterminals, since only the leaf is a terminal
- ▷ Since there are only $|V|$ variables in V, some variable X appears more than once on the longest path in D_s; we may assume that X is the closest nonterminal to the leaf that is repeated on this path

Proof, continuation

Now we divide s into $uvxyz$

- ▷ Each occurrence of X has a subtree under it, generating a portion of s
- ▷ The upper occurrence of X has the larger subtree and generates vxy, whereas the lower occurrence generates just x
- ▷ Since both of the trees mentioned above are generated by X, we may substitute one for the other and still obtain a valid derivation tree
- ▷ Replacing the smaller tree by the larger repeatedly, gives derivation trees for the strings uv^ixy^iz for each $i > 1$.
- ▷ Replacing the larger tree by the smaller one generates the string uxz. This establishes condition 1 of the lemma.

Proof, continuation

To prove condition 2 we must ensure that not both v and y are ϵ:

- ▷ If both v and y were ϵ, the derivation tree obtained by substituting the smaller tree for the larger tree would have fewer nodes than D_s, and would still generate s
- ▷ This however is a contradiction with out assumption that D_s is the smallest derivation tree generating s

Proof, continuation

To prove condition 3, we must ensure that $|vxy| \leq p$:

- ▷ In the derivation tree D_s the upper occurrence of X generates vxy
- ▷ We choose X so that both occurrences fall within the bottom $|V|+1$ nonterminals on the path and we chose the longest path in D_s
- ▷ Hence, the subtree where X generates vxy is at most $|V|+1$ high.
- ▷ A tree of this height can generate a string of length at most $b^{|V|+1} = p$

Example

We will use the pumping lemma to show that the language $B = \{\mathrm{a}^n\mathrm{b}^n\mathrm{c}^n \mid n \geq 0\}$ is not context-free by assuming it is and deriving a contradiction.

Proof

- Let p be the pumping length for B. Consider the string $s = \mathrm{a}^p\mathrm{b}^p\mathrm{c}^p$, for which it holds that $s \in B$ and $|s| \geq p$.
- By condition 2 of pumping lemma, in any division $s = uvxyz$ either v or y is not empty

Example

We will use the pumping lemma to show that the language $B = \{\mathrm{a}^n\mathrm{b}^n\mathrm{c}^n \mid n \geq 0\}$ is not context-free by assuming it is and deriving a contradiction.

Proof

- Let p be the pumping length for B. Consider the string $s = \mathrm{a}^p\mathrm{b}^p\mathrm{c}^p$, for which it holds that $s \in B$ and $|s| \geq p$.
- By condition 2 of pumping lemma, in any division $s = uvxyz$ either v or y is not empty
  - When both v and y contain only one type of symbol $(\mathrm{a}, \mathrm{b}, \mathrm{c})$, v does not contain both a's and b's or both b's and c's; the same holds for y.
    In this case uv^2xy^2z cannot contain an equal number of $\mathrm{a}, \mathrm{b}, \mathrm{c}$.

Example

We will use the pumping lemma to show that the language $B = \{\mathrm{a}^n\mathrm{b}^n\mathrm{c}^n \mid n \geq 0\}$ is not context-free by assuming it is and deriving a contradiction.

Proof

- Let p be the pumping length for B. Consider the string $s = \mathrm{a}^p\mathrm{b}^p\mathrm{c}^p$, for which it holds that $s \in B$ and $|s| \geq p$.
- By condition 2 of pumping lemma, in any division $s = uvxyz$ either v or y is not empty
  - When both v and y contain only one type of symbol $(\mathrm{a}, \mathrm{b}, \mathrm{c})$, v does not contain both a's and b's or both b's and c's; the same holds for y. In this case uv^2xy^2z cannot contain an equal number of $\mathrm{a}, \mathrm{b}, \mathrm{c}$.
  - When either v or y contain more than one type of symbols $\mathrm{a}, \mathrm{b}, \mathrm{c}$, uv^2xy^2z may contain equal number of $\mathrm{a}, \mathrm{b}, \mathrm{c}$, but they do not occur in the right order.

Example

We will use the pumping lemma to show that the language $B = \{\mathrm{a}^n\mathrm{b}^n\mathrm{c}^n \mid n \geq 0\}$ is not context-free by assuming it is and deriving a contradiction.

Proof

- Let p be the pumping length for B. Consider the string $s = \mathrm{a}^p\mathrm{b}^p\mathrm{c}^p$, for which it holds that $s \in B$ and $|s| \geq p$.
- By condition 2 of pumping lemma, in any division $s = uvxyz$ either v or y is not empty
  - When both v and y contain only one type of symbol $(\mathrm{a}, \mathrm{b}, \mathrm{c})$, v does not contain both a's and b's or both b's and c's; the same holds for y. In this case uv^2xy^2z cannot contain an equal number of $\mathrm{a}, \mathrm{b}, \mathrm{c}$.
  - When either v or y contain more than one type of symbols $\mathrm{a}, \mathrm{b}, \mathrm{c}$, uv^2xy^2z may contain equal number of $\mathrm{a}, \mathrm{b}, \mathrm{c}$, but they do not occur in the right order.
- Since one of these cases must occur, a contradiction is unavoidable, and therefore B is not a CFL.

Another example

We will show that $D = \{ww \mid w \in \{0,1\}^*\}$ is not a CFL by assuming it is and deriving a contradiction.

Proof

- ▷ Let p be the pumping length for D. Choosing a counter-example s is less obvious.

Another example

We will show that $D = \{ww \mid w \in \{0,1\}^*\}$ is not a CFL by assuming it is and deriving a contradiction.

Proof

- ▷ Let p be the pumping length for D. Choosing a counter-example s is less obvious.
- ▷ We will consider $s = 0^p 1^p 0^p 1^p$. Can we pump it?
- ▷ In dividing $s = uvxyz$, $|vxy|$ must be at most p, and v or y is not empty.

Another example

We will show that $D = \{ww \mid w \in \{0,1\}^*\}$ is not a CFL by assuming it is and deriving a contradiction.

Proof

- ▷ Let p be the pumping length for D. Choosing a counter-example s is less obvious.
- ▷ We will consider $s = 0^p 1^p 0^p 1^p$. Can we pump it?
- ▷ In dividing $s = uvxyz$, $|vxy|$ must be at most p, and v or y is not empty.
- ▷ If vxy would occur only in the first half, pumping uv^2xy^2z moves a 1 into the first position of the second half of s so it cannot be of the form ww

Another example

We will show that $D = \{ww \mid w \in \{0,1\}^*\}$ is not a CFL by assuming it is and deriving a contradiction.

Proof

- ▷ Let p be the pumping length for D. Choosing a counter-example s is less obvious.
- ▷ We will consider $s = 0^p 1^p 0^p 1^p$. Can we pump it?
- ▷ In dividing $s = uvxyz$, $|vxy|$ must be at most p, and v or y is not empty.
- ▷ If vxy would occur only in the first half, pumping uv^2xy^2z moves a 1 into the first position of the second half of s so it cannot be of the form ww
- ▷ Similarly if vxy is in the second part of s, pumping uv^2xy^2z moves a 0 into the last position of the first half of s so it cannot be of the form ww

Another example

We will show that $D = \{ww \mid w \in \{0,1\}^*\}$ is not a CFL by assuming it is and deriving a contradiction.

Proof

- ▷ Let p be the pumping length for D. Choosing a counter-example s is less obvious.
- ▷ We will consider $s = 0^p 1^p 0^p 1^p$. Can we pump it?
- ▷ In dividing $s = uvxyz$, $|vxy|$ must be at most p, and v or y is not empty.
- ▷ If vxy would occur only in the first half, pumping uv^2xy^2z moves a 1 into the first position of the second half of s so it cannot be of the form ww
- ▷ Similarly if vxy is in the second part of s, pumping uv^2xy^2z moves a 0 into the last position of the first half of s so it cannot be of the form ww
- ▷ If vxy contains the midpoint of s, however, when we pump s down to uxz, it has the form $0^p 1^k 0^l 1^p$, where k and l cannot both be p. Therefore this string is not of the form ww either.

Another example

We will show that $D = \{ww \mid w \in \{0,1\}^*\}$ is not a CFL by assuming it is and deriving a contradiction.

Proof

- ▷ Let p be the pumping length for D. Choosing a counter-example s is less obvious.
- ▷ We will consider $s = 0^p 1^p 0^p 1^p$. Can we pump it?
- ▷ In dividing $s = uvxyz$, $|vxy|$ must be at most p, and v or y is not empty.
- ▷ If vxy would occur only in the first half, pumping uv^2xy^2z moves a 1 into the first position of the second half of s so it cannot be of the form ww
- ▷ Similarly if vxy is in the second part of s, pumping uv^2xy^2z moves a 0 into the last position of the first half of s so it cannot be of the form ww
- ▷ If vxy contains the midpoint of s, however, when we pump s down to uxz, it has the form $0^p 1^k 0^l 1^p$, where k and l cannot both be p. Therefore this string is not of the form ww either.
- ▷ Therefore, s cannot be pumped and D is not a CFL.