Pumping Lemma for Context-Free Languages

CSCI 3130 Formal Languages and Automata Theory

Siu On CHAN

Fall 2019

Chinese University of Hong Kong

$L_1 = \{\mathsf{a}^n\mathsf{b}^n \mid n \geqslant 0\}$

$L_2 = \{z \mid z \text{ has the same number of a's and b's}\}$

$L_3 = \{\mathsf{a}^n\mathsf{b}^n\mathsf{c}^n \mid n \geqslant 0\}$

$L_4 = \{zz^R \mid z \in \{\mathsf{a}, \mathsf{b}\}^*\}$

$L_5 = \{zz \mid z \in \{\mathsf{a}, \mathsf{b}\}^*\}$

These languages are not regular

Are they context-free?

An attempt

$$L_3 = \{\mathsf{a}^n\mathsf{b}^n\mathsf{c}^n \mid n \geqslant 0\}$$

Let's try to design a CFG or PDA

$S \to \mathsf{a}B\mathsf{c} \mid \varepsilon$

$B \to$???

read a / push x

read b / pop x

???

Suppose we could construct some CFG G for L_3

e.g.

$S \to CC \mid BC \mid \text{a}$

$B \to CS \mid \text{b}$

$C \to SB \mid \text{c}$

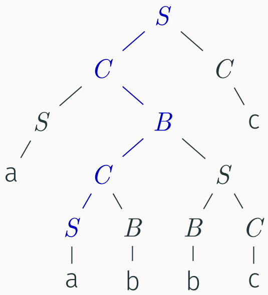
How does a long derivation look like?

$$
\begin{aligned}
S &\Rightarrow CC \\
&\Rightarrow SBC \\
&\Rightarrow SCSC \\
&\Rightarrow SSBSC \\
&\Rightarrow SSBBCC \\
&\Rightarrow \text{a}SBBCC \\
&\Rightarrow \text{aa}BBCC \\
&\Rightarrow \text{aab}BCC \\
&\Rightarrow \text{aabb}CC \\
&\Rightarrow \text{aabbc}C \\
&\Rightarrow \text{aabbcc}
\end{aligned}
$$

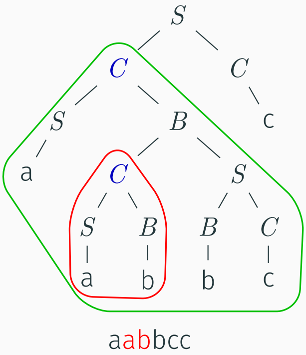
Repetition in long derivations

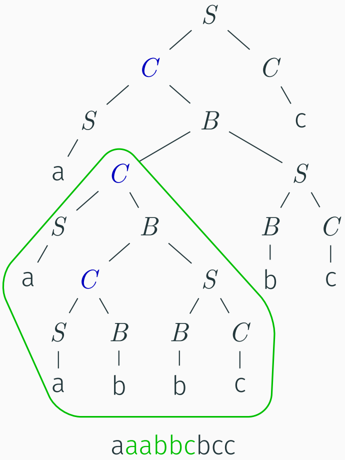
If a derivation is long enough, some variable must appear twice on the same root-to-leave path in a parse tree

$$\begin{aligned} S &\Rightarrow CC \\ &\Rightarrow SBC \\ &\Rightarrow SCSC \\ &\Rightarrow SSBSC \\ &\Rightarrow SSBBCC \\ &\Rightarrow \text{a}SBBCC \\ &\Rightarrow \text{aa}BBCC \\ &\Rightarrow \text{aab}BCC \\ &\Rightarrow \text{aabb}CC \\ &\Rightarrow \text{aabbc}C \\ &\Rightarrow \text{aabbcc} \end{aligned}$$



Then we can "cut and paste" part of parse tree





Pumping example

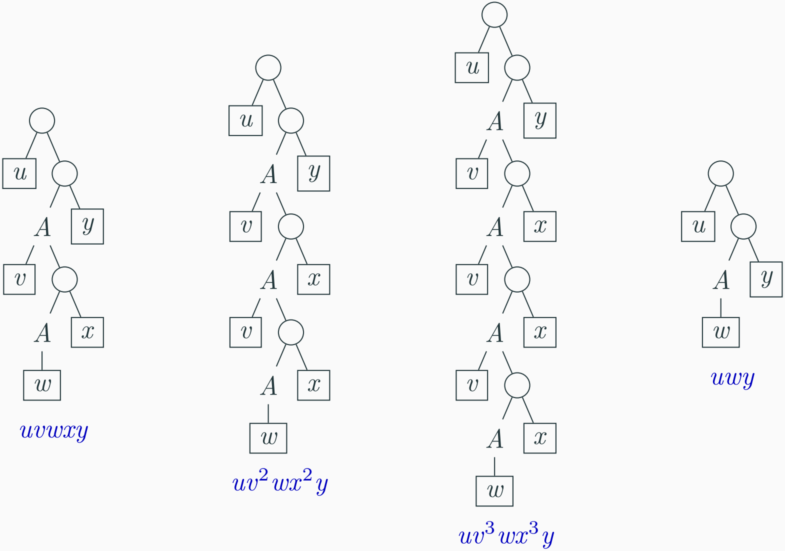
We can repeat this many times

$$\mathrm{aabbcc} \Rightarrow \mathrm{aaabbcbcc} \Rightarrow \mathrm{aaaabbcbcbcc} \Rightarrow \dots$$

$$\Rightarrow (\mathrm{a})^i\mathrm{ab}(\mathrm{bc})^i\mathrm{c}$$

Every sufficiently large derivation will have a middle part that can be repeated indefinitely

Pumping in general



Example

$$L_3 = \{\mathsf{a}^n\mathsf{b}^n\mathsf{c}^n \mid n \geqslant 0\}$$

If L_3 has a context-free grammar G, then for any sufficiently long $s \in L(G)$

s can be split into $s = uvwxy$ such that $L(G)$ also contains uv^2wx^2y, uv^3wx^3y, ...

What happens if $s = \mathsf{a}^m\mathsf{b}^m\mathsf{c}^m$

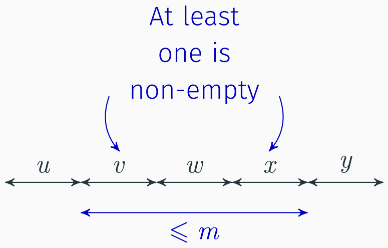
No matter how it is split, $uv^2wx^2y \notin L_3$

Pumping lemma for context-free languages

For every context-free language L

There exists a number m such that for every long string s in L ($|s| \geqslant m$), we can write $s = uvwxy$ where

1. $|vwx| \leqslant m$
2. $|vx| \geqslant 1$
3. For every $i \geqslant 0$, the string uv^iwx^iy is in L



Pumping lemma for context-free languages

To prove L is not context-free, it is enough to show that

For every m there is a long string $s \in L$, $|s| \geqslant m$, such that for every way of writing $s = uvwxy$ where

1. $|vwx| \leqslant m$
2. $|vx| \geqslant 1$

there is $i \geqslant 0$ such that uv^iwx^iy is not in L

Using the pumping lemma

$$L_3 = \{\mathsf{a}^n\mathsf{b}^n\mathsf{c}^n \mid n \geqslant 0\}$$

1. for every m
2. there is $s = \mathsf{a}^m\mathsf{b}^m\mathsf{c}^m$ (at least m symbols)
3. no matter how the pumping lemma splits s into $uvwxy$ ($|vwx| \leqslant m$, $|vx| \geqslant 1$)
4. $uv^2wx^2y \notin L_3$ (but why?)

Using the pumping lemma

Case 1: v or x contains two kinds of symbols

$$\text{aa}\underbrace{\text{aabb}}_{v}\text{bbcccc}$$

Then $uv^2wx^2y \notin L_3$ because the pattern is wrong

Case 2: v and x both contain (at most) one kind of symbol

$$\text{aaa}\underbrace{\ \text{a}\ }_{v}\text{b}\underbrace{\text{bb}}_{x}\text{bcccc}$$

Then uv^2wx^2y does not have the same number of a's, b's and c's

Conclusion: $uv^2wx^2y \notin L_3$

Which is context-free?

$L_1 = \{\mathsf{a}^n\mathsf{b}^n \mid n \geqslant 0\}$ ✓

$L_2 = \{z \mid z \text{ has the same number of a's and b's}\}$ ✓

$L_3 = \{\mathsf{a}^n\mathsf{b}^n\mathsf{c}^n \mid n \geqslant 0\}$ ✗

$L_4 = \{zz^R \mid z \in \{\mathsf{a}, \mathsf{b}\}^*\}$ ✓

$L_5 = \{zz \mid z \in \{\mathsf{a}, \mathsf{b}\}^*\}$

Example

$$L_5 = \{zz \mid z \in \{\mathsf{a}, \mathsf{b}\}^*\}$$

1. for every m
2. there is $s = \mathsf{a}^m \mathsf{b} \mathsf{a}^m \mathsf{b}$ (at least m symbols)
3. no matter how the pumping lemma splits s into $uvwxy$ ($|vwx| \leqslant m$, $|vx| \geqslant 1$)
4. Is $uv^2wx^2y \notin L_5$?

Example

$$L_5 = \{zz \mid z \in \{\mathsf{a}, \mathsf{b}\}^*\}$$

1. for every m
2. there is $s = \mathsf{a}^m\mathsf{b}\mathsf{a}^m\mathsf{b}$ (at least m symbols)
3. no matter how the pumping lemma splits s into $uvwxy$ ($|vwx| \leqslant m$, $|vx| \geqslant 1$)
4. Is $uv^2wx^2y \notin L_5$?

$$\mathsf{aaa}\underbrace{\mathsf{a}}_{v}\mathsf{aba}\underbrace{\mathsf{a}}_{x}\mathsf{aaab}$$

Example

$$L_5 = \{zz \mid z \in \{\mathsf{a}, \mathsf{b}\}^*\}$$

1. for every m
2. there is $s = \mathsf{a}^m\mathsf{b}^m\mathsf{a}^m\mathsf{b}^m$ (at least m symbols)
3. no matter how the pumping lemma splits s into $uvwxy$ ($|vwx| \leqslant m$, $|vx| \geqslant 1$)
4. Is $uv^iwx^iy \notin L_5$ for some i?

Recall that $|vwx| \leqslant m$

Example

Three cases

Case 1 aaa $\underbrace{\text{aabbb}}_{vwx}$ bbaaaaabbbbb

vwx is in the first half of a^{m}b^{m}a^{m}b^m

Case 2 aaaaabb $\underbrace{\text{bbbaa}}_{vwx}$ aaabbbbb

vwx is in the middle part of a^{m}b^{m}a^{m}b^m

Case 3 aaaaabbbbbaaa $\underbrace{\text{aabbb}}_{vwx}$ bb

vwx is in the second half of a^{m}b^{m}a^{m}b^m

Example

Apply pumping lemma with $i = 0$

Case 1 aaa $\underbrace{\text{aabbb}}_{vwx}$ bbaaaaabbbbb

uwy becomes $\mathsf{a}^j\mathsf{b}^k\mathsf{a}^m\mathsf{b}^m$, where $j < m$ or $k < m$

Case 2 aaaaabb $\underbrace{\text{bbbaa}}_{vwx}$ aaabbbbb

uwy becomes $\mathsf{a}^m\mathsf{b}^j\mathsf{a}^k\mathsf{b}^m$, where $j < m$ or $k < m$

Case 3 aaaaabbbbbaaa $\underbrace{\text{aabbb}}_{vwx}$ bb

uwy becomes $\mathsf{a}^m\mathsf{b}^m\mathsf{a}^j\mathsf{b}^k$, where $j < m$ or $k < m$

Not of the form zz

This covers all cases, so L_5 is not context-free