Context Free Grammars

Shakir Ullah Shah

Outline

- Removing Λ-Productions
- Removing Unit-Productions
- Removing Useless Variables
- Normal Form of Context Free Grammars
  - Chomsky Normal Form (CNF)

Killing Λ-Productions

Λ-Productions:

In a given CFG, we call a non-terminal N null able

if there is a production N → Λ, or there is a

derivation that starts at N and lead to a Λ.

N → ……… → Λ

- Λ-Productions are undesirable.
- We can replace Λ-production with appropriate non-Λ productions.

Theorem 23

- If L is CFL generated by a CFG having Λ-productions, then there is a different CFG that has no Λ-production and still generates either the whole language L (if L does not include Λ) or else generate the language of all the words in L other than Λ.
- Replacement Rule.
  1. Delete all Λ-Productions.
  2. Add the following productions:
- For every production of the X → old string

Add new production of the form X → .., where right side will account for every modification of the old string that can be formed by deleting all possible subsets of null-able Non-Terminals, except that we do not allow X → Λ, to be formed if all the character in old string are null-able

Example

Consider the CFG

$S \to Xa$

$X \to aX \mid bX \mid \Lambda$

Old nullable Production	New Production
$S \to Xa$	$S \to a$
$X \to aX$	$X \to a$
$X \to bX$	$X \to b$

So the new CFG is

$S \to Xa \mid a$

$X \to aX \mid bX \mid a \mid b$

Example

$S \rightarrow XY$
$X \rightarrow Zb$
$Y \rightarrow bW$
$Z \rightarrow AB$
$W \rightarrow Z$
$A \rightarrow aA \mid bA \mid \Lambda$
$B \rightarrow Ba \mid Bb \mid \Lambda$

- Null-able Non-terminals are?
- A, B, Z and W

S → XY
X → Zb
Y → bW
Z → AB
W → Z
A → aA | bA | Λ
B → Ba | Bb | Λ

Example Contd.

Old nullable Production	New Production
X → Zb	X → b
Y → bW	Y → b
Z → AB	Z → A and Z → B
W → Z	Nothing new
A → aA	A → a
A → bA	A → b
B → Ba	B →a
B → Bb	B → b

important

So the new CFG is
S → XY
X → Zb | b
Y → bW | b
Z → AB | A | B
W → Z
A → aA | bA | a | b
B → Ba | Ba | a | b

Killing unit-productions

- **Definition:** A production of the form
- Nonterminal one Nonterminal

is called a **unit production**.

- The following theorem allows us to get rid of unit productions:

Theorem 24:

If there is a CFG for the language L that has no

Λ-productions, then there is also a CFG for L with

no Λ-productions and **no unit productions**.

Proof of Theorem 24

- This is another proof by constructive algorithm.
- **Algorithm:** For every pair of nonterminals A and B, if the CFG has a unit production $A \rightarrow B$, or if there is a chain

  $A \rightarrow X_1 \rightarrow X_2 \rightarrow \ldots \rightarrow B$

where X_1, X_2, ... are nonterminals, create new productions

as follows:

- If the non-unit productions from B are

  $B \rightarrow s_1 \mid s_2 \mid \ldots$

where s_1, s_2, ... are strings, we create the productions

Example

- Consider the CFG

  S → A| bb

  A → B | b

  B → S | a
- The non-unit productions are

  S → bb A → b B → a
- And unit productions are

  S → A

  A → B

  B → S

Example contd.

- Let's list all unit productions and their sequences and create new productions:

  S → A gives S → b

  S → A → B gives S → a

  A → B gives A → a

  A → B → S gives A → bb

  B → S gives B → bb

  B → S → A gives B → b

- Eliminating all unit productions, the new CFG is

  S → bb | b | a

  A → b | a | bb

  B → a | bb | b

- This CFG generates a finite language since there are no nonterminals in any strings produced from S.

Consider the CFG

S → A| bb

A → B | b

B → S | a

unit productions are

S → A

A → B

B → S

Normal Form of Context Free Grammars

Normal Form

- To convert a grammar in normal form means “standardizing the productions without changing the resulting language”.

THE CHOMSKY NORMAL FORM

A context-free grammar is in Chomsky normal form if every rule is of the form:

A → BC	**B and C are not start variables**
A → a	**a is a terminal**
S → ε	**S is the start variable and** if **ε is in the language**

Any variable A that is not the start variable can only generate strings of length > 0

THE CHOMSKY NORMAL FORM

A context-free grammar is in Chomsky normal form if every rule is of the form:

A → BC	**B and C are not start variable**
A → a	**a is a terminal**
S → ε	**S is the start variable**

S → 0S1
S → TT
T → ε

T → 0
U → SV
S → TU | ε
V → 1

- $S \rightarrow AS|a$
- $A \rightarrow SA|b$

Is in Chomsky Normal Form. The Grammar

- $S \rightarrow AS|ASS$
- $A \rightarrow SA|aa$
- Is not; both production $S \rightarrow AS|ASS$ and $A \rightarrow SA|aa$ violate the conditions of definition.

Theorem: Any context-free language can be generated by a context-free grammar in Chomsky normal form

“Can transform any CFG into Chomsky normal form”

Chomsky Normal Form

There are 5 steps to follow in order to transform a grammar into CNF:

1. Add the a new start variable S_0 and the production rule $S_0 \longrightarrow S$.
2. Eliminate the ε-rules.
3. Eliminate the unary productions $A \longrightarrow B$.
4. Add rules of the form $V_t \longrightarrow t$ for every terminal t and replace t with the variable V_t.
5. Transform the remaining of the rules to the form $A \longrightarrow BC$ (A, B, C variables).

1. Add a new start variable

- We have to make sure that the start variable doesn't occur to the right side of some rule.
- Thus, we add a new start variable S_0 and the rule $S_0 \longrightarrow S$, where S is the old start variable.

2. Eliminate ε-rules

- We have to eliminate all productions of the form $A \longrightarrow \varepsilon$, for A being any non-start variable.
- To do so we should remove the rule $A \longrightarrow \varepsilon$ and replace every appearance of A with ε in all other rules.

3. Eliminate unary productions

- A unary production is a production of the form $A \longrightarrow B$ (with both A, B being variables).
- There should only be productions of the form $V_1 \longrightarrow V_2V_3$ involving variables, thus we have to eliminate unary productions.
- To do so, we replace B in $A \longrightarrow B$ with the right parts of the rules involving B in the left part.

4. Add $V_t \longrightarrow t$ and replace t with V_t

- There should only be rules of the form $A \longrightarrow t$ involving terminals, thus terminals should disappear from every other rule involving more than just one single literal.
- To do so, we add a new variable V_t for every terminal t and we replace every appearance of t with V_t , except those in rules of the form $A \longrightarrow t$.

5. Transform rules to $A \rightarrow BC$

- All the rules involving only variables should be of the form $A \rightarrow BC$. Thus we should take care of all the rules involving more than 2 variables in the right part
- For the rule $V \rightarrow A_1A_2A_3\ldots A_n$,we start reducing the size of the right part by replacing every two variables with one new variable (resulting in the creation of n-2 new variables).

5. Transform rules to A $\rightarrow$ BC

$$V \rightarrow A_1A_2A_3A_4A_5A_6\ldots A_n$$

5. Transform rules to $A \longrightarrow BC$

$$V \longrightarrow B_1 A_3 A_4 A_5 A_6 \ldots A_n$$

$$B_1 \longrightarrow A_1 A_2$$

5. Transform rules to $A \longrightarrow BC$

$$V \longrightarrow B_2 A_4 A_5 A_6 \ldots A_n$$

$$B_2 \longrightarrow B_1 A_3$$

$$B_1 \longrightarrow A_1 A_2$$

5. Transform rules to A $\longrightarrow$ BC

$$V \longrightarrow B_3A_5A_6\ldots A_n$$

$$B_3 \longrightarrow B_2A_4$$

$$B_2 \longrightarrow B_1A_3$$

$$B_1 \longrightarrow A_1A_2$$

5. Transform rules to $A \longrightarrow BC$

$$V \longrightarrow B_4 A_6 \ldots A_n$$

$$B_4 \longrightarrow B_3 A_5$$

$$B_3 \longrightarrow B_2 A_4$$

$$B_2 \longrightarrow B_1 A_3$$

$$B_1 \longrightarrow A_1 A_2$$

5. Transform rules to $A \longrightarrow BC$

$$V \longrightarrow B_{n-2}A_n$$

$$B_{n-2} \longrightarrow B_{n-3}A_{n-1}$$

$$\ldots$$

$$B_4 \longrightarrow B_3A_5$$

$$B_3 \longrightarrow B_2A_4$$

$$B_2 \longrightarrow B_1A_3$$

$$B_1 \longrightarrow A_1A_2$$

Example

$S \longrightarrow CSC \mid B$

$C \longrightarrow 00 \mid \varepsilon$

$B \longrightarrow 01B \mid 1$

Example

1. Add new start variablle

$S_0 \longrightarrow S$

$S \longrightarrow CSC \mid B$

$C \longrightarrow 00 \mid \varepsilon$

$B \longrightarrow 01B \mid 1$

Example

2. Eliminate ε-moves

$S_0 \longrightarrow S$

$S \longrightarrow CSC \mid B$

$C \longrightarrow 00 \mid \cancel{\varepsilon}$

$B \longrightarrow 01B \mid 1$

Example

2. Eliminate ε-moves

$S_0 \longrightarrow S$

$S \longrightarrow CSC \mid B \mid CS \mid SC \mid S$

$C \longrightarrow 00$

$B \longrightarrow 01B \mid 1$

Example

3. Eliminate Unary Productions

$S_0 \longrightarrow S$

$S \longrightarrow CSC \mid B \mid CS \mid SC \mid \cancel{S}$

$C \longrightarrow 00$

$B \longrightarrow 01B \mid 1$

Example

3. Eliminate Unary Productions

$S_0 \longrightarrow S$

$S \longrightarrow CSC \mid B \mid CS \mid SC$ –

$C \longrightarrow 00$

$B \longrightarrow 01B \mid 1$

Example

3. Eliminate Unary Productions

$S_0 \longrightarrow S$

$S \longrightarrow CSC \mid \cancel{B} \mid CS \mid SC$

$C \longrightarrow 00$

$B \longrightarrow 01B \mid 1$

Example

3. Eliminate Unary Productions

$S_0 \longrightarrow S$

$S \longrightarrow CSC \mid 01B \mid 1 \mid CS \mid SC$

$C \longrightarrow 00$

$B \longrightarrow 01B \mid 1$

Example

3. Eliminate Unary Productions

$S_0 \longrightarrow \cancel{S}$

$S \longrightarrow CSC \mid 01B \mid 1 \mid CS \mid SC$

$C \longrightarrow 00$

$B \longrightarrow 01B \mid 1$

Example

3. Eliminate Unary Productions

$S_0 \longrightarrow CSC \mid 01B \mid 1 \mid CS \mid SC$

$S \longrightarrow CSC \mid 01B \mid 1 \mid CS \mid SC$

$C \longrightarrow 00$

$B \longrightarrow 01B \mid 1$

Example

<u>4. Create V_t for every terminal t</u>

$S_0 \longrightarrow CSC \mid 01B \mid 1 \mid CS \mid SC$

$S \longrightarrow CSC \mid 01B \mid 1 \mid CS \mid SC$

$C \longrightarrow 00$

$B \longrightarrow 01B \mid 1$

$Z \longrightarrow 0$

Example

4. Create V_t for every terminal t

$S_0 \longrightarrow CSC \mid Z1B \mid 1 \mid CS \mid SC$

$S \longrightarrow CSC \mid Z1B \mid 1 \mid CS \mid SC$

$C \longrightarrow ZZ$

$B \longrightarrow Z1B \mid 1$

$Z \longrightarrow 0$

Example

4. Create V_t for every terminal t

$S_0 \to CSC \mid Z1B \mid 1 \mid CS \mid SC$

$S \to CSC \mid Z1B \mid 1 \mid CS \mid SC$

$C \to ZZ$

$B \to Z1B \mid 1$

$Z \to 0$

$A \to 1$

Example

4. Create V_t for every terminal t

$S_0 \longrightarrow$ CSC | ZAB | 1 | CS | SC

$S \longrightarrow$ CSC | ZAB | 1 | CS | SC

$C \longrightarrow$ ZZ

$B \longrightarrow$ ZAB | 1

$Z \longrightarrow 0$

$A \longrightarrow 1$

Example

5. Take care of long rules

$S_0 \longrightarrow$ CSC | ZAB | 1 | CS | SC

$S \longrightarrow$ CSC | ZAB | 1 | CS | SC

$C \longrightarrow$ ZZ

$B \longrightarrow$ ZAB | 1

$Z \longrightarrow 0$

$A \longrightarrow 1$

$D \longrightarrow CS$

Example

5. Take care of long rules

$S_0 \longrightarrow DC \mid ZAB \mid 1 \mid CS \mid SC$

$S \longrightarrow DC \mid ZAB \mid 1 \mid CS \mid SC$

$C \longrightarrow ZZ$

$B \longrightarrow ZAB \mid 1$

$Z \longrightarrow 0$

$A \longrightarrow 1$

$D \longrightarrow CS$

Example

5. Take care of long rules

$S_0 \longrightarrow DC \mid ZAB \mid 1 \mid CS \mid SC$

$S \longrightarrow DC \mid ZAB \mid 1 \mid CS \mid SC$

$C \longrightarrow ZZ$

$B \longrightarrow ZAB \mid 1$

$Z \longrightarrow 0$

$A \longrightarrow 1$

$D \longrightarrow CS$

$E \longrightarrow ZA$

Example

5. Take care of long rules

$S_0 \longrightarrow DC \mid EB \mid 1 \mid CS \mid SC$

$S \longrightarrow DC \mid EB \mid 1 \mid CS \mid SC$

$C \longrightarrow ZZ$

$B \longrightarrow EB \mid 1$

$Z \longrightarrow 0$

$A \longrightarrow 1$

$D \longrightarrow CS$

$E \longrightarrow ZA$

Convert the following into Chomsky normal form:

$A \rightarrow BAB \mid B \mid \varepsilon$

$B \rightarrow 00 \mid \varepsilon$

$S_0 \rightarrow A$

$A \rightarrow BAB \mid B \mid \varepsilon$

$B \rightarrow 00 \mid \varepsilon$

$S_0 \rightarrow A \mid \varepsilon$

$A \rightarrow BAB \mid B \mid BB \mid AB \mid BA$

$B \rightarrow 00$

$S_0 \rightarrow BAB \mid 00 \mid BB \mid AB \mid BA \mid \varepsilon$

$A \rightarrow BAB \mid 00 \mid BB \mid AB \mid BA$

$B \rightarrow 00$

$S_0 \rightarrow BC \mid DD \mid BB \mid AB \mid BA \mid \varepsilon,\quad C \rightarrow AB,$

$A \rightarrow BC \mid DD \mid BB \mid AB \mid BA,\quad B \rightarrow DD,\quad D \rightarrow 0$