Simplifications of Context-Free Grammars

Course Instructor: Hashim Ayub

Book: Prof. Sipser-MIT
Slides: Prof. Busch - LSU

Simplification of Context Free Grammar

Reduction of CFG

In CFG, sometimes all the production rules and symbols are not needed for the derivation of strings. Besides this, there may also be some NULL Productions and UNIT Productions. Elimination of these productions and symbols is called Simplification of CFG.

Simplification consists of the following steps:

1) Reduction of CFG
2) Removal of Unit Productions
3) Removal of Null Productions

REDUCTION OF CFG

CFG are reduced in two phases

Phase 1: Derivation of an equivalent grammar G' , from the CFG, G, such that each variable derives some terminal string

Derivation Procedure:

Step 1: Include all Symbols W_1 ,that derives some terminal and initialize i = 1

Step 2: Include symbols W_{i+1} , that derives W_i

Step 3: Increment i and repeat Step 2, until $W_{i+1} = W_i$

Step 4: Include all production rules that have W_i in it

Phase 2: Derivation of an equivalent grammar G'' , from the CFG, G', such that each symbol appears in a sentential form

Derivation Procudure:

Step 1: Include the Start Symbol in Y_1 and initialize i =1

Step 2: Include all symbols Y_{i+1} , that can be derived from Y_i and include all production rules that have been applied

Step 3: Increment i and repeat Step 2, until $Y_{i+1} = Y_i$

Example: Find a reduced grammar equivalent to the grammar G, having production rules

P: $S \to AC|B$, $A \to a$, $C \to c|BC$, $E \to aA|e$

Phase 1

$T = \{a, c, e\}$

$W_1 = \{A, C, E\}$

$W_2 = \{A, C, E, S\}$

$W_3 = \{A, C, E, S\}$

$G' = \{(A, C, E, S), \{a, c, e\}, P, (S)\}$

P: $S \to AC$, $A \to a$, $C \to c$, $E \to aA|e$

$W_2 = \{A, C, E, S\}$

$W_3 = \{A, C, E, S\}$

$G' = \{(A, C, E, S), \{a, c, e\}, P, (S)\}$

P: $S \to AC$, $A \to a$, $C \to c$, $E \to aA \mid e$

Phase 2 : $Y_1 = \{S\}$

$Y_2 = \{S, A, C\}$

$Y_3 = \{S, A, C, a, c\}$

$Y_4 = \{S, A, C, a, c\}$

$G'' = \{(A, C, S), \{a, c\}, P, \{S\}\}$

P: $S \to AC$, $A \to a$, $C \to c$

Simplification of Context Free Grammar

Removal of Unit Productions

Any Production Rule of the form $A \to B$ where $A, B \in$ Non Terminals is called Unit Production

Procedure for Removal

Step 1: To remove $A \to B$, add production $A \to x$ to the grammar rule whenever $B \to x$ occurs in the grammar. [$x \in$ Terminal, x can be Null]

Step 2: Delete $A \to B$ from the grammar.

Step 3: Repeat from Step 1 until all Unit Productions are removed.

Example: Remove Unit Productions from the Grammar whose production rule is given by

P: $S \to XY$, $X \to a$, $Y \to Z|b$, $Z \to M$, $M \to N$, $N \to a$

$Y \to Z$, $Z \to M$, $M \to N$

1) Since $N \to a$, we add $M \to a$

P: $S \to XY$, $X \to a$, $Y \to Z|b$, $Z \to M$, $M \to a$, $N \to a$

2) Since $M \to a$, we add $Z \to a$

P: $S \to XY$, $X \to a$, $Y \to Z|b$, $Z \to a$, $M \to a$, $N \to a$

3) Since $Z \to a$, we add $Y \to a$

P: $S \to XY$, $X \to a$, $Y \to a|b$, $Z \to a$, $M \to a$, $N \to a$

Remove the Unreachable symbols

P: $S \to XY$, $X \to a$, $Y \to a|b$

Simplification of Context Free Grammar

Removal of Null Productions

In a CFG, a Non-Terminal Symbol 'A' is a nullable variable if there is a production $A \rightarrow \epsilon$ or there is a derivation that starts at 'A' and leads to ϵ. (Like $A \rightarrow \ldots\ldots\ldots\ldots \rightarrow \epsilon$)

Procedure for Removal:

Step 1: To remove $A \rightarrow \epsilon$, look for all productions whose right side contains A

Step2: Replace each occurences of 'A' in each of these productions with ϵ

Step 3: Add the resultant productions to the Grammar

Example: Remove Null Productions from the following Grammar

$S \to ABAC, \quad A \to aA \mid \epsilon, \quad B \to bB \mid \epsilon, \quad C \to c$

$A \to \epsilon \quad , \quad B \to \epsilon$

1) To eliminate $A \to \epsilon$

$S \to \underline{A}B\underline{A}C$

$S \to ABC \mid BAC \mid BC$

$A \to aA$

$A \to a$

New production: $S \to ABAC \mid ABC \mid BAC \mid BC$

$A \to aA \mid a \quad , \quad B \to bB \mid \epsilon \quad , \quad C \to c$

2) To eliminate $B \rightarrow \epsilon$

$S \rightarrow AAC \mid AC \mid C$, $B \rightarrow b$

New production: $S \rightarrow ABAC \mid ABC \mid BAC \mid BC \mid AAC \mid AC \mid C$

$A \rightarrow aA \mid a$

$B \rightarrow bB \mid b$

$C \rightarrow c$

A Substitution Rule

$$S \rightarrow aB$$
$$A \rightarrow aaA$$
$$A \rightarrow abBc$$
$$B \rightarrow aA$$
$$B \rightarrow b$$

Substitute

$$B \rightarrow b$$

Equivalent grammar

$$S \rightarrow aB \mid ab$$
$$A \rightarrow aaA$$
$$A \rightarrow abBc \mid abbc$$
$$B \rightarrow aA$$

$$S \to aB \mid ab$$

$$A \to aaA$$

$$A \to abBc \mid abbc$$

$$B \to aA$$

Substitute

$$B \to aA$$

$$S \to \cancel{aB} \mid ab \mid aaA$$

$$A \to aaA$$

$$A \to \cancel{abBc} \mid abbc \mid abaAc$$

Equivalent grammar

Now, following are equivalent grammars

$$S \rightarrow aB$$
$$A \rightarrow aaA$$
$$A \rightarrow abBc$$
$$B \rightarrow aA$$
$$B \rightarrow b$$

$$\equiv$$

$$S \rightarrow ab \mid aaA$$
$$A \rightarrow aaA$$
$$A \rightarrow abbc \mid abaAc$$

The new equivalent grammar involves no 'B'

In general: $A \to xBz$

$B \to y_1$

Substitute

$B \to y_1$

$A \to xBz \mid xy_1z$

equivalent grammar

Nullable Variables

$\lambda-\text{production}:$ $X \rightarrow \lambda$

Nullable Variable: $Y \Rightarrow \ldots \Rightarrow \lambda$

Example:

$$S \rightarrow aMb$$

$$M \rightarrow aMb$$

$$M \rightarrow \lambda$$

Nullable variable

$\lambda-\text{production}$

Removing $\lambda-$productions

$$S \to aMb$$

$$M \to aMb$$

$$\cancel{M \to \lambda}$$

Substitute $M \to \lambda$

$$S \to aMb \mid ab$$

$$M \to aMb \mid ab$$

After we remove all the $\lambda-$productions

We have more concise grammar.

Unit-Productions

Unit Production: $X \to Y$

(a single variable in both sides)

Example:

$$S \to aA$$

$$A \to a$$

$$\boxed{A \to B}$$

$$\boxed{B \to A}$$

$$B \to bb$$

Unit Productions

Removal of unit productions:

$$S \to aA$$

$$A \to a$$

$$\cancel{A \to B}$$

$$B \to A$$

$$B \to bb$$

Substitute

$$A \to B$$

$$S \to aA \mid aB$$

$$A \to a$$

$$B \to A \mid B$$

$$B \to bb$$

Unit productions of form $X \to X$ can be removed immediately

$$S \to aA \mid aB$$
$$A \to a$$
$$B \to A \mid \cancel{B}$$
$$B \to bb$$

Remove
$B \to B$

$$S \to aA \mid aB$$
$$A \to a$$
$$B \to A$$
$$B \to bb$$

$$S \to aA \mid aB$$

$$A \to a$$

$$\cancel{B \to A}$$

$$B \to bb$$

Substitute

$$B \to A$$

$$S \to aA \mid aB \mid aA$$

$$A \to a$$

$$B \to bb$$

Remove repeated productions

$$S \to aA \mid aB \mid \cancel{aA}$$

$$A \to a$$

$$B \to bb$$

Final grammar

$$S \to aA \mid aB$$

$$A \to a$$

$$B \to bb$$

Useless Productions

$$S \to aSb$$

$$S \to \lambda$$

$$S \to A$$

$$A \to aA$$ Useless Production

Some derivations never terminate...

$$S \Rightarrow A \Rightarrow aA \Rightarrow aaA \Rightarrow \ldots \Rightarrow aa\ldots aA \Rightarrow \ldots$$

Another grammar:

$$S \to A$$

$$A \to aA$$

$$A \to \lambda$$

$$B \to bA$$ Useless Production

Not reachable from S

In general:

If there is a derivation

$$S \Rightarrow \ldots \Rightarrow xAy \Rightarrow \ldots \Rightarrow w \quad \in L(G)$$

consists of terminals

Then variable A is useful

Otherwise, variable A is useless

A production $A \to x$ is useless
if any of its variables is useless

Variables		Productions
	$S \to aSb$	
	$S \to \lambda$	
	$S \to A$	useless
useless	$A \to aA$	useless
useless	$B \to C$	useless
useless	$C \to D$	useless

Removing Useless Variables and Productions

Example Grammar:

$$S \to aS \mid A \mid C$$

$$A \to a$$

$$B \to aa$$

$$C \to aCb$$

First: find all variables that can produce strings with only terminals or λ (possible useful variables)

$$S \to aS \mid A \mid C$$
$$A \to a$$
$$B \to aa$$
$$C \to aCb$$

Round 1: $\{A, B\}$

(the right hand side of production that has only terminals)

Round 2: $\{A, B, S\}$

(the right hand side of a production has terminals and variables of previous round)

This process can be generalized

Then, remove productions that use variables other than $\{A, B, S\}$

$$S \to aS \mid A \mid \cancel{C}$$

$$A \to a$$

$$B \to aa$$

$$\cancel{C \to aCb}$$

$$\Rightarrow$$

$$S \to aS \mid A$$

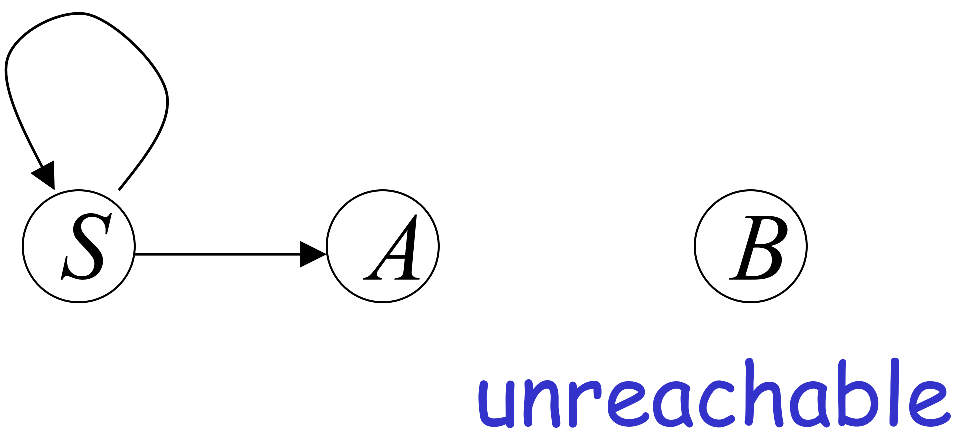
$$A \to a$$

$$B \to aa$$

Second: Find all variables reachable from S

Use a Dependency Graph where nodes are variables

$$S \to aS \mid A$$
$$A \to a$$
$$B \to aa$$



Keep only the variables reachable from S

$$S \to aS \mid A$$

$$A \to a$$

$$\cancel{B \to aa}$$

Final Grammar

$$S \to aS \mid A$$

$$A \to a$$

Contains only useful variables

Summary:

Step 1: Remove Nullable Variables

Step 2: Remove Unit-Productions

Step 3: Remove Useless Variables

This sequence guarantees that unwanted variables and productions are removed

Any Questions?

Thank You!