

C Y K Algorithm

John Cocke 1970

Daniel Younger 1967

Yasuo Kasami 1965

Goal: Given a CFG G in Chomsky Normal Form,
and a string w , determine if $w \in L(G)$.

We know that if $|w|=n$, it requires exactly 2^{n-1} derivations.

Naive idea: Try out all derivations of 2^{n-1} steps.
This is not time efficient.

CYK algorithm is based on dynamic programming.
This runs in $O(n^3)$ time.

$w_{24} = 111$
 $w = 011100$

Break down the problem into
similar sub problems.

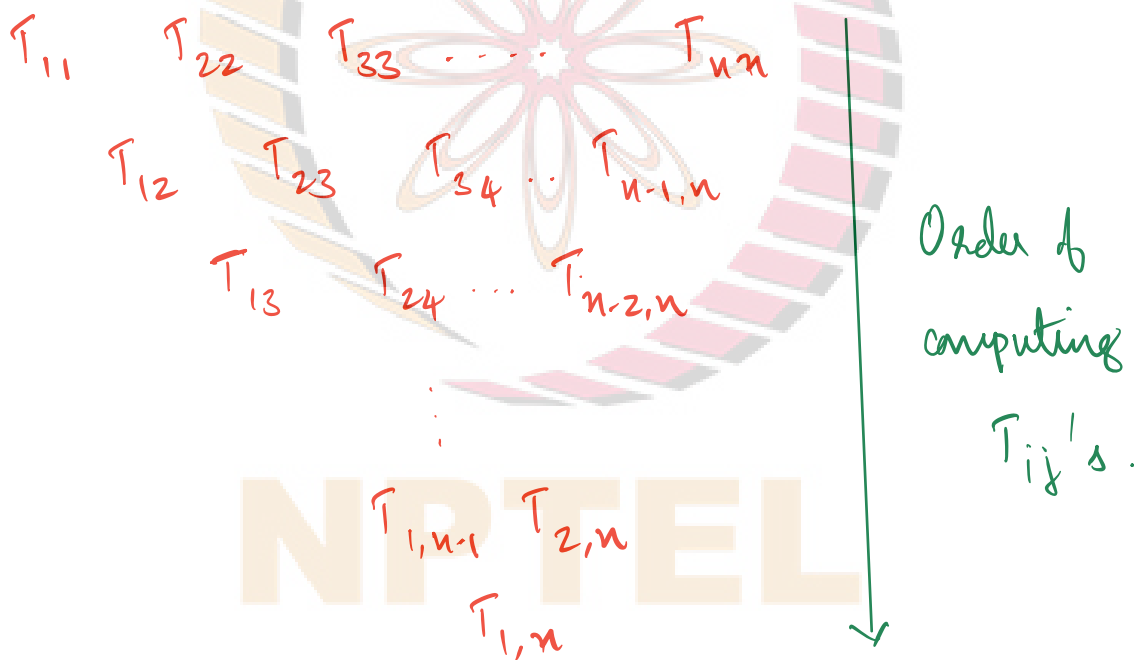
Let $w = a_1 a_2 \dots a_n$ where each $a_i \in \Sigma$.

Define $w_{ij} = a_i a_{i+1} \dots a_j$ for all $1 \leq i \leq j \leq n$.

CYK builds T_{ij} for all $1 \leq i \leq j \leq n$, such that

$$T_{ij} = \{A \in V \mid A \xRightarrow{*} w_{ij}\}$$

leading up to T_{1n} . Finally we check if the start variable $S \in T_{1n}$. That is, if $S \xRightarrow{*} w_{1n} = w$.



A algorithm outline

k+1 long substring
 \uparrow

For $k=0$ to $n-1$, compute all $T_{i,i+k}$!

First we have $k=0$. $T_{i,i} = \{A \mid A \xRightarrow{*} w_{i,i} = a_i\}$

All the rules $A \rightarrow a_i$

$w_{i,i+k} = a_i a_{i+1} \dots a_{i+j}$
 $w_{i,i+j}$

$a_{i+j+1} \dots a_{i+k}$
 $w_{i+j+1,i+k}$

For $k > 0$,

$A \in T_{i,i+k}$

$$\iff \begin{cases} B \in T_{i,i+j} \\ C \in T_{i+j+1,i+k} \\ A \rightarrow BC \text{ is a rule} \end{cases}$$

Algorithm

If $w = \epsilon$, accept if $\underline{S \rightarrow \epsilon}$ is a rule

For $i=1$ to n

$A \in T_{i,i} \iff A \rightarrow a_i$ is a rule

For $k=1$ to $n-1$

For $i=1$ to $n-k$

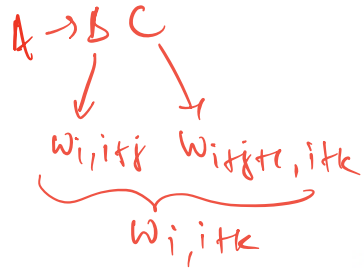
For $j=0$ to $k-1$

Check all the rules $A \rightarrow BC$

If $T_{i,i+j}$ contains B , and

$T_{i+j+1, k}$ contains C ,

then $T_{i,i+k} = T_{i,i+k} \cup \{A\}$



If $S \in T_{1,n}$, we say that $w \in L(G)$

Else $w \notin L(G)$.

Running Time: $O(n^3 r)$ where r is the number of rules. For a fixed grammar, r is considered to be a constant.

$\rightarrow O(n^3)$
algorithm

Correctness: Is evident from the algorithm.

Example:

$S \rightarrow AB \mid BC$

$A \rightarrow BA \mid a$

$B \rightarrow CC \mid b$

$C \rightarrow AB \mid a$

$w = \underline{baaba}$

$w_{11} = b \quad T_{11} = \{B\}$

$w_{22} = a \quad T_{22} = \{A, C\}$

$T_{15} = \{S, A, C\}$

Since $S \in T_{15}$, we say that w is in $L(G)$.

$T_{i,j}$ - Contents of $T_{i,j}$

5	SAC	SAC	B	SA	AC
4	-	B	SC	B	
3	-	B	AC		
2	SA	AC			
1	B				
	1	2	3	4	5

$\uparrow j$

\xrightarrow{i}

baaba

NPTEL