

### 7.1 Chomsky Normal Form

(개요) Context-Free 문법의 가장 간결한 형태(Normal Form),  
Context-free 언어를 위한 Pumping Lemma 증명에 요긴하게 사용된다.

(정의 Chomsky Normal Form; CNF)

Context-free 문법  $G = (N, T, P, S)$ 의 문법 규칙  $P$ 가

$S \rightarrow \epsilon$ 이외에는 모두

$A \rightarrow BC$  또는  $A \rightarrow a$ 인 형태만을 가지면(단  $A, B, C \in N$ 이고  $a \in T$ 이다)

이 문법  $G$ 를 Chomsky Normal Form(CNF)이라 부른다.

즉 Chomsky Normal Form의 형태에 문법은 파스나무의 가지(branch)가 둘인 이진나무(binary tree)이고  $S \rightarrow \epsilon$ 이외에는  $\epsilon$ 를 가지지 않는다<sup>1)</sup>( $\epsilon$ -free) context-free 문법의 가장 간결한 형태(normal form)이다.

(정리 7. 15) 임의의 context-free 문법  $G = (N, T, P, S)$ 는 이와 같은 언어를 만드는,  $L(G) = L(G')$ , CNF 문법  $G' = (N', T, P', S')$ 가 있고 문법  $G$ 를 CNF 문법  $G'$ 로 바꾸는데  $O(|N|^2)$ 인 optimal 알고리즘이 있다.

(증명) 교과서나 TP 참조.

(정리 7. 17) 임의의 CNF 문법  $G = (N, T, P, S)$ 의 문장  $z \in L(G)$ 의 파스나무(parse tree)의 가장 긴 경로(path)의 길이<sup>2)</sup>가  $n \geq 1$ <sup>3)</sup> 이면  $|z| \geq 2^{n-1}$ 이다.

(증명) 경로의 길이, 자연수( $n \geq 1$ )에 관한 수학적 귀납법.

(기본)  $n = 1$ 이면,  $z \in T$ ,  $|z| = 1 \geq 2^{1-1} = 2^0 = 1$ .

(반복)  $n > 1$ 이면,

$S \rightarrow AB$ 가 파스나무(뿌리나무)의 뿌리이다. (1)

$A$ 와  $B$ 를 새로운 뿌리로 하는 두<sup>4)</sup> 나뭇가지(sub-tree)를 생각해 보자.

$A \Rightarrow^* z_A$ 이고  $B \Rightarrow^* z_B$ 라면, (2)

귀납 가정에 의하여  $|z_A| \geq 2^{n-2}$ 이고  $|z_B| \geq 2^{n-2}$ 이다.

(1), (2)를 합하면,

$S \Rightarrow AB \Rightarrow^* z_A z_B = z$ .

$\therefore |z| = |z_A| + |z_B| \geq 2^{n-2} + 2^{n-2} = 2^{n-1}$ .

1)  $S \rightarrow \epsilon$ 을 제외한 나머지 규칙들은 모두  $\epsilon$ 를 가지지 않으므로,  $\epsilon \in L(G)$ 이기 위하여 예외적으로  $S \rightarrow \epsilon$ 을 허용하였다.

2) 경로(path)에 있는 간선(edge) 수, 경로에 있는 정점(vertex) 수는 간선 수 보다 하나(1) 더 많다.

3) 나무 뿌리(root)  $S$ 는  $N$ 의 원소이고, 잎사귀(leaf)는  $T$ 의 원소이므로 경로의 길이는 1이상이다.

4) CNF이므로 가지가 2개 이다.

## 7.2 Pumping Lemma

(개요) 어떤 언어가 Context-Free 언어가 **아니라고** 증명하는데 쓰는 Lemma

(정리 7.18) Context-free 언어에 관한 Pumping Lemma

[가정] 언어  $L$ 이 context-free라고 하자.

[결론] 언어  $L = L(G)$ 인 CNF 문법  $G = (N, T, P, S)$ 가 있다.

$n = 2^{|N|}$ 이라고, 길이  $|z| \geq n$ 인 문장  $z \in L$ 의 파스나무를 생각해 보자.

가장 긴 경로(path)의 길이가  $k+1$ 이라하자.

$$n = 2^{|N|} \leq |z| \leq 2^{(k+1)-1} = 2^k \quad (\text{정리 7. 17}).$$

$$\therefore |N| \leq k.$$

길이가  $k+1$ 인 경로를  $(A_0, A_1, \dots, A_k, a)$ 라 하자.

$$\therefore 0 \leq \exists i < \exists j \leq k, A_i = A_j = A. \quad (\because |N| \leq k)$$

$S \Rightarrow^* uA_iy \Rightarrow^* wA_jxy \Rightarrow^* wuxy$ 라면,

$S \Rightarrow^* uAy$ 이고  $A_i = A_j = A$ 이어서,  $A \Rightarrow^* w$ 나  $A \Rightarrow^* vAx$ 이므로,

$$S \Rightarrow^* uAy \Rightarrow^* wxy.$$

$$S \Rightarrow^* uAy \Rightarrow^* uAxxy \Rightarrow^* uvwxy.$$

$$S \Rightarrow^* uAy \Rightarrow^* uAxxy \Rightarrow^* uvvAxxxy \Rightarrow^* w^2uw^2y.$$

$$S \Rightarrow^* uAy \Rightarrow^* uAxxy \Rightarrow^* uvvAxxxy \Rightarrow^* uvvvAxxxxy \Rightarrow^* w^3uw^3y.$$

...

$$\therefore \forall k \geq 0: S \Rightarrow^* w^k u w^k y \in L.$$

위를 수식으로 정리하여 표현하면

[가정] 언어  $L$ 은 context-free이다.

[결론] (a)  $\exists n \geq 0$ :

(b)  $\forall w \in L: |w| \geq n$ ,

(c)  $\exists u, v, w, x, y \in \Sigma^*: z = uvwxy, vx \neq \epsilon, |vwx| \leq n$ ,

(d)  $\forall k \geq 0: w^k u w^k y \in L$ .

언어  $L$ 이 context-free이면, (a) 길이가 **어떤**( $\exists$ ) 자연수  $n$  이상( $|z| \geq n$ )인 (b) **모든**( $\forall$ ) 문장 ( $z \in L$ )에 (c) **어떤**( $\exists$ ) substring  $v$ 와  $x$ 가 **어떤**( $\exists$ ) substring  $u, w, y$ 의 사이사이에서 (d) **항상**( $\forall$ ) **같은 수**( $k$ ) 만큼 반복(pumping)하여 문장이 된( $w^k u w^k y \in L$ )다.

Pumping Lemma의 **대우**(contra-verse) 명제

[결론의 부정]

(a)  $\forall n \geq 0$ :

(b)  $\exists w \in L: |w| \geq n$ ,

(c)  $\forall u, v, w, x, y \in \Sigma^*: z = uvwxy, vx \neq \epsilon, |vwx| \leq n$ ,

5) 이때  $n$ 은 언어  $L$ 을 받아들이는 CNF 형태의 문법의  $n$  터미널의 개수 가  $|N|$ 이라면  $2^{|N|}$ 이다.

(d)  $\exists k \geq 0: w^k u x^k y \notin L$ .

[가정의 부정] 언어  $L$ 이 context-free가 아니다.

어떤 언어  $L$ 이 context-free가 아니라는 증명을 하려면,

(a) 길이가 모든( $\forall$ ) 자연수  $n$  이상,  $|w| \geq n$ , 인,

(b) 어떤 (무한) 문장,  $w \in L$ 은,

(c) 모든( $\forall$ ) substring 쌍  $v$ 와  $x$ 가 쌍으로 반복(pumping)하면서,

(a) 단 반복할 substring  $v$ 와  $x$ 는 빈 문자열은 아니고( $vx \neq \epsilon$ ; **non-empty pumping**),

(b) 첫 번째 반복( $v^1 u x^1$ ; **first pump**)는 CNF 문법의 모든  $n$  터미널을 유도(derive) 하기 이전( $|vux| \leq n$ )을 우선 생각한다.

(d) 문장이 되지 않는 경우가, **있**( $\exists k$ )다( $w^k u x^k y \notin L$ )고 증명하면 된다.

(예 1)  $L = \{0^n 1^n 2^n \mid n \geq 0\}$ 는 context-free가 아니다.

(증명) (a)  $\forall n \geq 0: (b) \exists w = 0^n 1^n 2^n \in L$ ,

(c)  $\forall u, v, w, x, y: z = wuxy, w \neq \epsilon, |wud| \leq n$ ,

(c.1)  $0 \leq \forall i \leq n: u = 0^{n-i}, vux = 0^i 1^{n-i}, y = 1^i 2^n \wedge$

(c.2)  $0 \leq \forall i \leq n: u = 0^n 1^{n-i}, vux = 1^i 2^{n-i}, y = 2^i$ .

(d)  $\forall k \neq 1: w^0 u x^0 y \notin L$ .

(d.1)  $v = 0^j (j > 0), x = 1^m (m > 0)$ 이면  $w^0 u x^0 y = w y = 0^{n-kj} 1^{n-km} 2^n \notin L \wedge$

(d.2)  $v = 1^j (j > 0), x = 2^m (m > 0)$ 이면  $w^0 u x^0 y = w y = 0^n 1^{n-kj} 2^{n-km} \notin L$ .