Chap. 8
Properties of
Context-Free Languages

Agenda of Chapter 8

- Two pumping lemmas
  - A Pumping lemma for context-free languages
  - A Pumping lemma for linear languages

- Closure properties and decision algorithms for CFL
  - Closure of context-free languages
  - Some decidable properties of context-free languages

Pumping Lemma for regular language – logical expression

If L is an infinite regular language, then

∃ m >0 such that

∀ w∈L (|w| ≥ m)

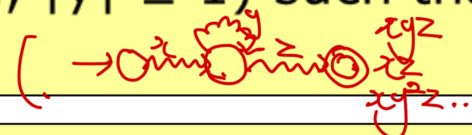
∃ a decomposition w=xyz (|xy| ≤ m, |y| ≥ 1) such that

∀ i=1,2,.. $w_i=xy^iz \in L$

(state의 수 : finite. w의 길이 : 무한.) → 무한 dfa에서 cycle.

길이가 긴 w에는 pumping되는 substring y가 존재.

→ 어떤 state는 반복될 수 밖에 없다.



Theorem 8.2] Pumping Lemma for linear languages

– L : an infinite linear language.

Then ∃ positive integers m such that

∀w∈L with |w| ≥ m, ∃ a decomposition of w,

w=uvxyz, with |uvyz|≤m and |vy|≥1, such that

∀ i=0,1,2,... $w_i=uv^ixy^iz$ is also in L.

≤m

위치 : 가운데 (중간 부분)

pumping 되는 부분이 양쪽

(uxz∈L, uvxyz∈L, uv²xy²z∈L)

linear-language → pumping 되는 위치가 다름.

위치 → 양쪽 끝

Theorem 8.1] Pumping Lemma for CFL

– L : an infinite CFL

Then ∃ positive integers m such that

∀w∈L with |w| ≥ m,

∃ a decomposition of w,

w=uvxyz with |vxy|≤m and |vy|≥1, such that

∀ i=0,1,2,... $w_i=uv^ixy^iz$ is also in L

Pumping Lemma for CFL (1/4)

Theorem 8.1] Pumping Lemma for CFL

- L : an infinite CFL

  Then ∃ positive integers m such that

  ∀w∈L with |w| ≥ m,

  ∃ a decomposition of w,

  w=u$\boxed{vxy}$z with |vxy|≤m and |vy|≥1, such that

  ∀ i=0,1,2,… $w_i=uv^ixy^iz$ is also in L

Proof) Consider L-{λ} with a CFG

having no λ−production and unit-production.

Note]

A→x.

오른쪽 string의 길이는 유한.

- the length of string on the right side of any production is bounded.
- L is infinite
- There exist arbitrary long derivation → w가 무한.
- The number of variable in G is finite. → variable의 개수는 유한.

Pumping Lemma for CFL (2/4)

$S \overset{*}{\Rightarrow} uAy \overset{*}{\Rightarrow} uvAyz \overset{*}{\Rightarrow} uvxyz$

Proof continued)

This indicates the existence of some variable A that repeats on this path, such as

$$S \overset{*}{\Rightarrow} uAz \overset{*}{\Rightarrow} uvAyz \overset{*}{\Rightarrow} uvxyz,$$

$uvvAyyz \Rightarrow uv^2Ay^2z \Rightarrow uv^3Ay^3z \cdots \Rightarrow uv^ixy^iz.$

where u,v,x,y,z are all strings of terminals.

$S \overset{*}{\Rightarrow} aSb$

From $A \overset{*}{\Rightarrow} vAy$ and $A \overset{*}{\Rightarrow} x$, we can see that all the string uv^ixy^iz ($i=0,1,\ldots$) can be generate by G.

$S \overset{*}{\Rightarrow} ab.$

pumping이 존재

계속 생성될 수 있다.

a^nb^n

$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaaaSbbbb$

$uvxyz.$

$\Rightarrow aaaaabbbbb.$

$|uvyz| \le m.$

bound 가능

[poof of $|vxy| \le m$]

Assuming that A is the only repeating variable in $A \overset{*}{\Rightarrow} vAy$, $A \overset{*}{\Rightarrow} x$, we get $|vxy| \le m$.

유한 유한

[poof of $|vy| \ge 1$]

From the fact that G has no λ–production and unit-production, we get $|vy| \ge 1$.

Pumping Lemma for CFL (3/4)

Ex 8.1] Show that L={$a^nb^nc^n$ | n≥0} is not context-free.

- For any given m, choose w= $a^mb^mc^m$. Assume. L is CFL.
- Consider all possible choice of vxy. →PL is false →모순. →L is not CFL.
  - For the case vxy = $a^{k1}xa^{k2}$, set i=0 then w_0=uxz ∉ L.
  - For the case vxy = $a^{k1}xb^{k2}$, set i=0 then w_0=uxz ∉ L.

Assume L is CFL.

∀m>0, we choose $w=a^mb^mc^m \in L$ ($|w|\ge m$, $w\in L$)

∀possible decomp of $w=uvxyz$ ($|vxy|\le m$, $|vy|\ge 1$)

① $vxy = a^{k1}xa^{k2}$ ($1\le k_1+k_2\le m$)

② $vxy = b^{k1}xb^{k2}$ (")

③ $vxy = c^{k1}xc^{k2}$ (")

④ $vxy = a^{k1}xb^{k2}$

⑤ $vxy = b^{k1}xc^{k2}$

$w = \overbrace{a\cdots a}^{m}\overbrace{b\cdots b}^{m}\overbrace{c\cdots c}^{m}$ (vxy vxy vxy vxy vxy)

set i=0

$w_0 = uxz = a^{m-k_1-k_2}b^mc^m \notin L$ ($m-k_1-k_2\neq m$).

set i=0 $w_0=uxz=a^{m-k_1}b^{m-k_2}c^m \notin L$.

- Cf) L={a^nb^n | n≥0} is context-free language.

Hyeyoung Park, Dept. Comp. Sci.

Pumping Lemma for CFL (3/4)

Ex 8.2] $L=\{ww \mid w \in \{a,b\}^*\}$

- For any given m, choose $w= a^m b^m a^m b^m$.
- Consider all possible choice of vxy.

Pumping Lemma for CFL (4/4)

Ex 8.3] $L=\{a^{n!} \mid n \geq 0\}$ is not context-free

- For any given m, choose $w= a^{m!}$.
- Consider all possible choice of vxy.

Ex 8.4] $L=\{a^n b^j \mid n=j^2\}$ is not context-free

- For any given m, choose $w= a^{m^2} b^m$.
- Consider all possible choice of vxy.

Pumping Lemma for Linear Language (1/3)

Definition] Linear languages

CFL which can be generated by a linear CFG.

변수 1개

$S \to aSb \mid ab$. (linear)

$S \to SS \mid aSb \mid bSa \mid ab \mid ba$. (not linear)

Ex 8.5]

- Is the language $L=\{a^n b^n \mid n \geq 0\}$ linear? O

$A \to xBy$

terminal

linear grammar.

$S \to aSb \mid ab$

- Is the language $L=\{w \mid n_a(w)=n_b(w)\}$ linear? X

pumping lemma.

Theorem 8.2] Pumping Lemma for linear languages

- L : an infinite linear language.

  Then ∃ positive integers m such that

  $\forall w \in L$ with $|w| \geq m$, ∃ a decomposition of w,

  $w=uvxyz$, with $\boxed{|uvyz| \leq m}$ and $|vy| \geq 1$, such that

  $\forall\ i=0,1,2,\ldots\ w_i=uv^ixy^iz$ is also in L.

Pumping Lemma for Linear Language (2/3)

Proof) Consider L-{λ} with a LG without λ−production and unit-production.

Note]

- The length of string on the right side of any production is bounded
- L is infinite
- There exist arbitrary long derivation
- The number of variable in G is finite.

CFG $S \to SS$ $S \Rightarrow SS \Rightarrow SSS$

$\Rightarrow aSbSSS$

This indicates the existence of variable A repeating on this path, such as

$S \overset{*}{\Rightarrow} uAz \overset{*}{\Rightarrow} uvAyz \overset{*}{\Rightarrow} uvxyz$, where u,v,x,y,z are all strings of terminals.

From $A \overset{*}{\Rightarrow} vAy$ and $A \overset{*}{\Rightarrow} x$, we can see that

all the string uv^ixy^iz (i=0,1,…) can be generated by G.

Variable들

uvxyz

We can also assume that

- there only finite number of variables in $S \Rightarrow uAz$
- these variables generates finite number of terminals → $|uz|$ is bounded.

Similarly, $|vy|$ is also bounded, and $|uvyz| \le m$.

Since G has no λ−production and unit-production, $|vy| \ge 1$.

Hyeyoung Park, Dept. Comp. Sci.

Pumping Lemma for Linear Language (3/3)

Ex8.6] L={w| $n_a(w)=n_b(w)$} is not linear.

- For any given m, choose w= $a^m b^{2m} a^m$
- Consider all possible choice of uvyz. *

Assume L is Linear.

$\forall m>0$, we choose $w = a^m b^{2m} a^m \in L$ ($|w| \ge m$).

$\forall$ decomposition of $w = uvxyz$ ($|uvyz| \le m$, $|vy| \ge 1$) = $a^{k_1} a^{k_2} x a^{k_3} a^{k_4}$ ($k_1+k_2+k_3+k_4 \le m$, $k_2+k_3 \ge 1$)

$a^{k_1} a^{k_2} = \overbrace{a \dots a}^{m} \overbrace{b \dots \dots b}^{2m} \overbrace{a \dots a}^{m}$

uv yz

set i=0

$w = a^{m-k_2} b^{2m} a^{m-k_3} \notin L$

($m-k_2-k_3 \neq 2m$)

∴ pumping lemma for linear language is false

→ L is not Linear Language.

- $F_{LL} \subset F_{CFL}$
  - The family of linear languages is a proper subset of the family of context-free languages.

Closure of CFL (1/5)

$\{a^n b^n c\} \cup \{a^n b^{2n} c\} \rightarrow$ CFL

start variable 만 조정. 추가.

Theorem 8.3] The family of CFL is closed under
union, **concatenation**, and **star-closure**.

$S_4 \rightarrow S_1 S_2$.

$S_5 \rightarrow S_1 S_5 \mid \lambda$.

Proof) Let $L_1 = L(G_1)$, $G_1=(V_1,T_1,S_1,P_1)$ is a CFG

Let $L_2 = L(G_2)$, $G_2=(V_2,T_2,S_2,P_2)$ is a CFG

Assume V_1 and V_2 are disjoint.

1. Closure under union.

Consider $L(G_3)$,

추가.

$G_3=(V_1 \cup V_2 \cup \{S_3\}, T_1 \cup T_2, S_3, P_3)$, $P_3=P_1 \cup P_2 \cup \{S_3 \rightarrow S_1 \mid S_2\}$.

Then, $L(G_3)$ is a CFL and $L(G_3) = L_1 \cup L_2$

i) For any $w \in (L_1 \cup L_2)$,

either $S_3 \Rightarrow S_1 \Rightarrow w$ or $S_3 \Rightarrow S_2 \Rightarrow w$ is possible derivation in G_3. Thus, $w \in L(G_3)$.

ii) For any $w \in L(G_3)$

either $S_3 \Rightarrow S_1$ or $S_3 \Rightarrow S_2$ must be the first step of the derivation.

Case of $S_3 \Rightarrow S_1$: the rest $S_1 \Rightarrow w$ involve productions in P_1 only. Thus, $w \in L_1$.

Case of $S_3 \Rightarrow S_2$: the rest $S_2 \Rightarrow w$ involve productions in P_2 only. Thus, $w \in L_2$.

Closure of CFL (2/5)

Proof continued)

2. **Closure under concatenation.**

Consider $G_4=(V_1 \cup V_2 \cup \{S_4\}, T_1 \cup T_2, S_4, P_4)$,

$P_4 = P_1 \cup P_2 \cup \{S_4 \rightarrow S_1 S_2\}$.

Then, $L(G_4)$ is a CFL and $L(G_4) = L(G_1)L(G_2)$.

3. **Closure under star-closure**

Consider $G_5=(V_1 \cup \{S_5\}, T_1, S_5, P_5)$,

$P_5 = P_1 \cup \{S_5 \rightarrow S_1 S_5 | \lambda\}$.

Then, $L(G_5)$ is a CFL and $L(G_5) = L(G_1)^*$.

하나라도 안 된게 있으면 → closed x.

Closure of CFL (3/5)

Theorem 8.4] The family of CFL is not closed under **intersection** and **complementation**.

Proof)

1. **Intersection** (Proof by a counter example)

   CF CF

   Let $L_1=\{a^nb^nc^m \mid n,m\geq 0\}$, $L_2=\{a^nb^mc^m \mid n,m\geq 0\}$.

   L_1 and L_2 are context-free languages. (by using Theorem 8.3)

   But, $L_1 \cap L_2 = \{a^nb^nc^n \mid n\geq 0\}$ is not context-free language.

   → pumping lemma로 증명 했음.

2. **Complementation** (Proof by contradiction)

   Assume a family of CFL is closed under complementation.

   Then for two CFL L_1 and L_2, $(L_1' \cup L_2')'$ is also a CFL.

   → closed x.

   But, since $(L_1' \cup L_2')' = L_1 \cap L_2$, this is a contradiction.

   Thus, the family of CFL is not closed under complementation.

Hyeyoung Park, Dept. Comp. Sci.

Closure of CFL (4/5)

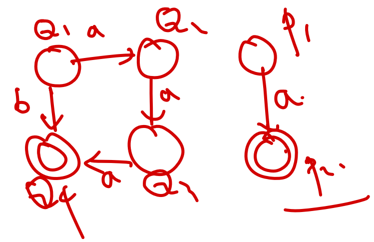
Theorem 8.5] Closure under regular intersection.

L_1: a CFL, L_2: a regular language. Then $L_1 \cap L_2$ is a CFL

Proof) Construct a ndpa M′ accepting $L_1 \cap L_2$.

Let $M_1 = (Q, \Sigma, \Gamma, \delta_1, q_0, z, F_1)$ be an ndpa, $L(M_1) = L_1$.

Let $M_2 = (P, \Sigma, \delta_2, p_0, F_2)$ be a dfa, $L(M_2) = L_2$.



Construction of a new ndpa]

$M' = (Q', \Sigma, \Gamma, \delta', q_0', z, F')$ simulating the parallel action of M_1 and M_2.

Let $Q' = Q \times P$, $q_0' = (q_0, p_0)$, $F' = F_1 \times F_2$

When $(q_k, x) \in \delta_1(q_i, a, b)$, $\delta_2(p_j, a) = p_l$, Let $\delta'((q_i, p_j), a, b) \ni ((q_k, p_l), x)$

Using M′ and induction method, we can obtain

$((q_0, p_0), w, z) \vdash^*_{M'} ((q_r, p_s), \lambda, x)$ with $q_r \in F_1$, $p_s \in F_2$

iff $(q_0, w, z) \vdash^*_{M1} (q_r, \lambda, x)$ and $\delta_2^*(p_0, w) = p_s$.

Thus, w is accepted by M′ iff w is accepted by M_1 and M_2.

Closure of CFL (5/5)

Ex 8.7] Show that "$L=\{a^n b^n \mid n \geq 0,\ n \neq 100\}$" is CFL.

- Use the fact that $L_1=\{a^n b^n \mid n \geq 0\}$ is context-free.

$L = \underbrace{L_1}_{CF} \cap \underbrace{\overline{\{a^{100} b^{100}\}}}_{reg}$. $\{a^{100} b^{100}\}^c$: reg

∴ L is CFL

Ex 8.8] Show $L=\{w \in \{a,b,c\}^* \mid n_a(w)=n_b(w)=n_c(w)\}$ is not a CFL.

- Use the fact that $L_1=\{a^n b^n c^n \mid n \geq 0\}$ is not context-free

Assume. L is "CFL"

$\underbrace{\underbrace{L}_{CFL} \cap \underbrace{\{a^n b^m c^l\}}_{Reg \to a^*b^*c^*}}_{CFL} = L_1$ ← not CFL.

∴ 가정에 모순. L is not a CFL.

Some decidable properties of CFL

Theorem 8.6] Given a CFG G=(V,T,S,P),
there exist an algorithm for deciding whether or not L(G) is empty.

Proof)

가는 힘든지.

Assume that λ is not in L(G).

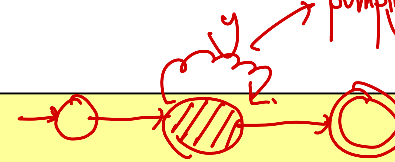
chapter 6.

Use the algorithm for removing useless symbols and productions.

S is useless iff L(G) is empty.

terminal은 못 만드는지

pumping 이 되면 infinite.



Theorem 8.7] Given a CFG G=(V,T,S,P),
there exist an algorithm for deciding whether or not L(G) is infinite.

Proof)

Need only to determine whether G has some repeating variable.
Draw a $\boxed{\text{dependency graph}}$ with an edge (A,B) whenever there is A → xBy.
Any variable at the base of a cycle is a repeating one.

A B