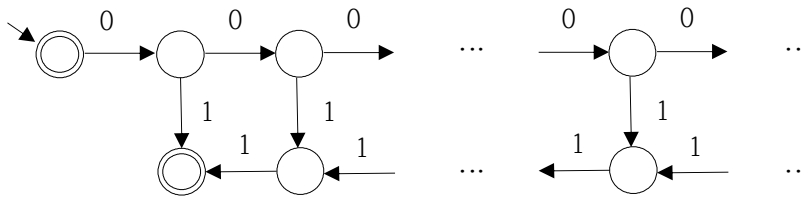


4장에서는 regular(type 3) 언어의 성질을 공부한다. 첫 째로 4.1에서는 regular언어의 성질을 알기위하여 regular가 아닌 언어를 살펴보고 4.2에서 언어가 regular가 아님을 formal하게 증명하기 위한 pumping lemma를 공부하고, 마지막으로 4.3에서 regular 언어의 닫혀(closure)있는 성질을 공부한다.

4.1 Regular가 아닌 언어

$$L = \{0^n1^n \mid n \geq 0\}$$

위의 언어를 받아들이는 오토마타를 설계해 보자.



상태수가 무한하게 된다.

이번에는 regular expression을 생각하여 보자.

0^*1^* 는 0^n1^n 의 슈퍼집합이다.

우리는 두 방법 모두 실패했다. 그러나 이것이 L 이 regular아 아니라는 증명은 아니다. 왜냐하면 우리가 실패했지, 다른 사람들도 모두 실패할 것이라는 확신은 없기 때문이다.

그래서 다음 섹션의 regular 언어를 위한 pumping lemma를 생각한다.

4.2 Pumping Lemma

(개요) Regular 언어가 아님을 증명하는데 쓰는 Lemma

Pumping Lemma

[가정] 언어 L 이 regular하다고 하자.

[결론] 어떤 자연수 n 이 존재하여 언어 L 을 받아들이는 상태가 n 개인 DFA가 있을 것이다. 길이가 자연수 n 보다 크거나 같은 문장(sentence) $w \in L$ 을 생각하자.

$$w = a_1a_2 \cdots a_m \quad (0 \leq \forall l < m: a_l \in \Sigma) \text{으로 표현 한다면 } m \geq n \text{이다.}$$

한편 문장 w 의 길이가 m 이므로 $m+1$ 개의 상태를 방문할 것이다.

이 때 방문한 상태를 (q_0, q_1, \dots, q_m) 라고 하면,

$$0 \leq \forall l < m: \delta(q_l, a_{l+1}) = q_{l+1} \text{일 것이다.}$$

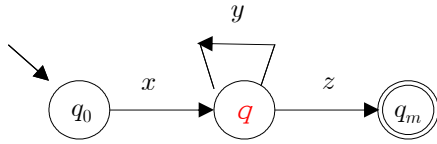
방문한 상태 수 $m+1$ 이 전제 상태 수 n 보다 크므로, ($m \geq n$ 이므로 $m+1 > n$)

같은 상태를 두 번 이상 방문하는 경우가 반드시 생긴다.

가장 먼저 방문한 같은 상태가 $q_i = q_j$ 라 하면 $1 \leq i < j \leq m$ 이고,

1) 가장 먼저 도착한 같은 상태가 $q_i = q_j$ 이므로 $i < j \leq m$ 이다.

$\delta^i(q_0, a_1 a_2 \cdots a_i) = q_i$, $\delta^{j-i}(q_i, a_{i+1} a_{i+2} \cdots a_j) = q_j$, $\delta^{m-j}(q_j, a_{j+1} a_{j+2} \cdots a_m) = q_m$ 이다,
 짧게 써서 $x = a_1 a_2 \cdots a_i$, $y = a_{i+1} a_{i+2} \cdots a_j$, $z = a_{j+1} a_{j+2} \cdots a_m$ 라고, $q_i = q_j = q$ 라면
 $w = xyz$ 이고 $|y| \geq 1$, $|xy| \leq n$ 이며, $\delta^{|x|}(q_0, x) = q$, $\delta^{|y|}(q, y) = q$, $\delta^{|z|}(q, z) = q_m$ 이다.



$$\delta^{|z|}(\delta^{|x|}(q_0, x), z) = \delta^{|xz|}(q_0, xz) = \delta^*(q_0, xz) = q_m,$$

$$\delta^{|z|}(\delta^{|y|}(\delta^{|x|}(q_0, x), y), z) = \delta^{|xyz|}(q_0, xyz) = \delta^*(q_0, xyz) = q_m,$$

$$\delta^{|z|}(\delta^{|y|}(\delta^{|y|}(\delta^{|x|}(q_0, x), y), y), z) = \delta^{|xy^2z|}(q_0, xy^2z) = \delta^*(q_0, xy^2z) = q_m,$$

...

$$\delta^{|z|}(\delta^{|y|} \dots (\delta^{|y|}(\delta^{|x|}(q_0, x), y), \dots y), z) = \delta^{|xy^kz|}(q_0, xy^kz) = \delta^*(q_0, xy^kz) = q_m,$$

...

$q_m \in F$ 이므로 $\forall k \geq 0 : xy^kz \in L$ 이다.

위를 수식으로 정리하여 표현하면

[가정] 언어 L 이 regular하다.

[결론] (a) $\exists n \geq 0$:

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

(c) $\exists x, y, z \in \Sigma^* : w = xyz, y \neq \epsilon, |xy| \leq n$,

(d) $\forall k \geq 0 : xy^kz \in L$.

언어 L 이 regular하면, (a) 길이가 어떤(\exists) 자연수 n^2 이상($|w| \geq n$)인 (b) 모든(\forall) 문장 ($w \in L$)이 가진 (c) 어떤(\exists) substring(y)이 (d) 항상($\forall k$) 반복(pumping)하여 문장이 된 ($xy^kz \in L$)다.

Pumping Lemma의 대우(contra-verse) 명제

[결론의 부정]

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

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

(c) $\forall x, y, z \in \Sigma^* : w = xyz, y \neq \epsilon, |xy| \leq n$,

(d) $\exists k \geq 0 : xy^kz \notin L$.

[가정의 부정] 언어 L 이 regular하지 않다.

명제 $p \Rightarrow q$ 가 참이면, 대우명제 $\neg q \Rightarrow \neg p$ 는 역시 참이다. 따라서 어떤 언어 L 이 regular 하지 않다는 증명을 하려면, 본 pumping lemma의 결론 부분의 부정을 증명하면 된다.

2) 이때 n 은 언어 L 을 받아들이는 DFA중 상태 수가 가장 작은(minimal DFA)의 상태수이다.

어떤 언어 L 이 regular하지 않다는 증명을 하려면,

- (a) 길이가 모든(\forall) 자연수 n 이상, $|w| \geq n$, 인³⁾,
- (b) 어떤(\exists) (무한) 문장, $w \in L$ 은,
- (c) 모든(\forall) substring($y \neq \epsilon$)이 반복(pumping)하여,
 - (a) 단, 반복할 substring y 는 빈 문자열은 아니고($y \neq \epsilon$; non-empty pumping),
 - (b) 첫 번째 반복(xy^1 ; first pump)은 minimal DFA의 모든 상태를 방문하기 이전 ($|xy| \leq n$)을 우선 생각한다.
- (d) 문장이 되지 않는 경우가 있($\exists k$)다($xy^kz \notin L$)고 증명하면 된다.

두 개의 모든(\forall) 중에 (a)는 무한문자열을 뜻하므로 주의하여야하고, (c)는 모든 substring으로의 나눔이므로 더욱 주의하여야 한다.

(예 1) $L = \{0^n 1^n \mid n \geq 0\}$ 는 regular하지 않다.

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

(c) $\forall x, y, z$: $w = xyz$, $y \neq \epsilon$, $|xy| \leq n$,

$$(c.1) 0 \leq \forall i, j \leq n: x = 0^i, y = 0^j, z = 0^{n-i-j} 1^n, j \geq 1, i+j \leq n \wedge$$

$$(c.2) 0 \leq \forall i, j \leq n: x = 0^n 1^i, y = 1^j, z = 1^{n-i-j}, j \geq 1, i+j \leq n.$$

$$(c'.1) \forall i, j: 0 \leq i < n, 0 < j \leq n, i+j \leq n: x = 0^i, y = 0^j, z = 0^{n-i-j} 1^n \wedge$$

$$(c'.2) \forall i, j: 0 \leq i < n, 0 < j \leq n, i+j \leq n: x = 0^n 1^i, y = 1^j, z = 1^{n-i-j}.$$

(d) $\exists k \geq 0$: $xy^kz \notin L$.

$$(d.1) \exists k \geq 0, k \neq 1: xy^kz = 0^{i+kj+n-i-j} 1^n = 0^{n+(k-1)j} 1^n \notin L \wedge$$

$$(d.2) \exists k \geq 0, k \neq 1: xy^kz = 0^n 1^{i+kj+n-i-j} = 0^n 1^{n+(k-1)j} \notin L.$$

$$(d'.1) \exists k=0: xy^0z = 0^{n-j} 1^n \notin L \wedge$$

$$(d'.2) \exists k=0: xy^0z = 0^n 1^{n-j} \notin L.$$

$$(d''.1) \exists k=2: xy^2z = 0^{n+j} 1^n \notin L \wedge$$

$$(d''.2) \exists k=2: xy^2z = 0^n 1^{n+j} \notin L.$$

(c)는 \forall 에 관한 조건이므로, $((c'.1) \wedge (c'.2)) \equiv ((c''.1) \wedge (c''.2)) \equiv$ (3)과 동치인 조건을 보아야하지만, (d)는 \exists 에 관한 조건이므로, $((d'.1) \wedge (d'.2)), ((d''.1) \wedge (d''.2)) \supseteq ((d.1) \wedge (d.2)) \equiv$ (d)의 부분집합인 조건만 확인하여도 충분하다. (a), (c)의 \forall 부분(adversary pick, 모든 경우)이 Pumping Lemma를 이용한 증명에서 어려운 부분이다⁴⁾.

$(c'.1) \wedge (c'.2)$, $(c''.1) \wedge (c''.2)$, $(d'.1) \wedge (d'.2)$, $(d''.1) \wedge (d''.2)$ 을 모두 증명하여야하나, 식의 형태로 보아 증명이 너무 쉬우므로(trivial) 생략하는 것에 지나지 않음에 유의하라.

3) 이 조건은 무한(infinite)문자열을 의미한다. 유한문자열은 pumping lemma의 대상이 아니고, 이미 regular이다.

4) 교과서 130p를 참고하시오.

(예 2) $L = \{w \in 1^* \mid |w| \text{는 소수다}\}$ 는 regular하지 않다.

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

(c) $\forall x, y, z$: $w = xyz$, $|y| \geq 1$, $|xy| \leq n$,

(c') $\forall m$, $m \geq 1$, $m \leq n$: $w = xyz = 1^p$, $y = 1^m$, $xz = 1^{p-m}$.

(d) $\exists k = p - m$ ⁵⁾: $|xy^{p-m}z| = (p-m) + (p-m)m = (m+1)(p-m)$.

$m+1 \geq 2 \wedge p-m \geq 2 (\because p \geq n+2 \wedge m \leq n)$

$\therefore |xy^{p-m}z| = (m+1)(p-m)$ 는 소수가 아니다. $xy^{p-m}z \notin L$.

참고로 (예 2)의 언어를 받아들이는 오토마타를 한번 생각해보고, 프로그램도 생각해보자.

5) $k = p - m$ 인 경우를 생각하는 것은 증명이 깔끔하기 때문이다.