

LÝ THUYẾT TÍNH TOÁN

BÀI 7: Ôtômat đẩy xuống

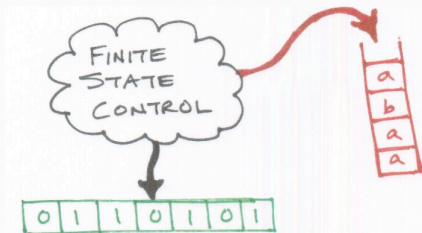
Phạm Xuân Cường
Khoa Công nghệ thông tin
cuongpx@tlu.edu.vn

1. Khái niệm
2. Định nghĩa hình thức
3. Sự tương đương với CFG
4. Ngôn ngữ không phi ngữ cảnh

Khái niệm

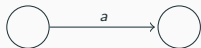
Khái niệm

- Ôtômat đẩy xuống = Push down automata (PDA)
- **PDA**: Là một mô hình tính toán, giống với NFA ngoại trừ một thành phần mở rộng được gọi là **ngăn xếp**
- **Ngăn xếp**: Là một cấu trúc dữ liệu hoạt động theo cơ chế **LIFO**
 - Các phương thức: read + push / ignored, pop/ignored
- PDA \Leftrightarrow CFG về sức mạnh \rightarrow Thêm công cụ hữu ích khi đoán nhận một ngôn ngữ phi ngữ cảnh



Biểu diễn hình học của PDA

FSM



PDA



Trong đó:

- a là ký tự vào
- b là ký tự nằm ở đỉnh ngăn xếp, ký tự này sẽ được lấy ra (pop)
- c là ký tự được đẩy (push) vào trong ngăn xếp

a, b, c đều có thể nhận ký tự ϵ

- Nếu $b = \epsilon \rightarrow$ ngăn xếp đang rỗng hoặc chưa được đọc
- Nếu $c = \epsilon \rightarrow$ không có gì được đẩy vào ngăn xếp

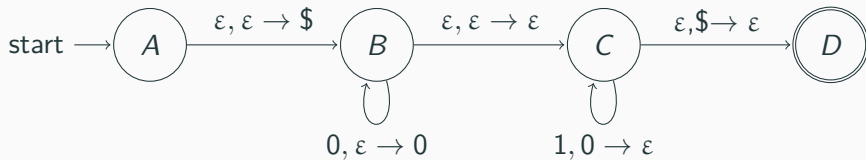
Ví dụ PDA

Xét ngôn ngữ $A = \{0^n 1^n \mid n \geq 0\}$

$\Sigma = \{0, 1\} \rightarrow$ Bộ chữ đầu vào

$\Gamma = \{\$, \epsilon\} \rightarrow$ Bộ chữ ngăn xếp

Ký tự $\$$ dùng để xác định đáy của ngăn xếp

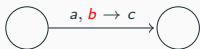


PDA hoạt động như thế nào?

Khi nào 1 chuỗi được chấp thuận:

- Tồn tại 1 đường đi từ trạng thái bắt đầu đến trạng thái chấp thuận trong bộ điều khiển trạng thái
- Đến cuối xâu sẽ đọc hết các ký tự và không còn ký tự nào trong ngăn xếp

Chú ý:



Lấy b từ ngăn xếp ra



Không lấy gì từ ngăn xếp ra

Định nghĩa hình thức

- Ôtômat đầy xuống \equiv bộ 6 (hay 6 chiều)

$$M = (Q, \Sigma, \Gamma, \delta, q_0, F)$$

Trong đó:

- **Q**: Tập trạng thái (hữu hạn)
- Σ_ϵ : Bộ chữ đầu vào, tập hữu hạn các ký tự
- Γ : Bộ chữ của ngăn xếp, tập hữu hạn các ký tự
- δ : Hàm dịch chuyển

$$\delta: Q \times \Sigma_\epsilon \times \Gamma_\epsilon \rightarrow P(Q \times \Gamma_\epsilon)$$

- **q₀**: Trạng thái bắt đầu ($q_0 \in Q$)
- **F**: Là tập các trạng thái kết thúc ($F \subseteq Q$)

Ví dụ PDA theo định nghĩa hình thức



- δ :

- $Q = \{A, B, C, D\}$
- $\Sigma = \{0, 1\}$
- $\Gamma = \{0, \$\}$
- $F = \{D\}$

Σ	0			1			ϵ		
Γ	0	\$	ϵ	0	\$	ϵ	0	\$	ϵ
A									$\{B, \$\}$
B			$\{B, 0\}$						$\{C, \epsilon\}$
C				$\{C, \epsilon\}$				$\{D, \epsilon\}$	
D									

Tất cả các ô trống đều biểu thị \emptyset

Sự tương đương với CFG

Sự tương đương với CFG

- Nhắc lại: Một ngôn ngữ phi ngữ cảnh là ngôn ngữ được biểu diễn bởi một CFG nào đó

Định lý 1

Một ngôn ngữ là phi ngữ cảnh **nếu và chỉ nếu** có một PDA nào đó đoán nhận nó

⇔ Định lý này có 2 chiều. Ta phát biểu nó thành từng bổ đề sau

Bổ đề 1.1

Nếu một ngôn ngữ là phi ngữ cảnh, thì tồn tại một PDA đoán nhận nó

Bổ đề 1.2

Nếu một PDA đoán nhận một ngôn ngữ nào đó, thì ngôn ngữ đó là phi ngữ cảnh

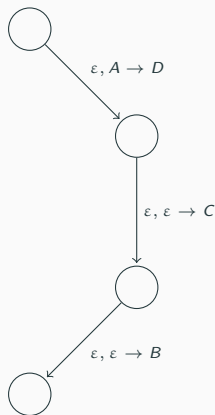
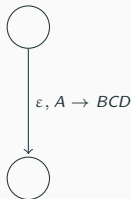
Chứng minh bổ đề 1.1

Ý TƯỞNG: Xây dựng PDA $P = (Q, \Sigma, \Gamma, \delta, q_0, F)$ đoán nhận cùng ngôn ngữ với CFG

Quy tắc xây dựng

1. Đặt ký hiệu đánh dấu \$ và biến ban đầu vào trong ngăn xếp
2. Lặp các bước sau:
 - a Nếu đỉnh của ngăn xếp là 1 ký hiệu biến $A \rightarrow$ Chọn một quy tắc ứng với A và thay thế bởi phần bên phải của quy tắc đó
 - b Nếu đỉnh của ngăn xếp là 1 ký hiệu kết thúc $a \rightarrow$ Đọc ký hiệu tiếp theo từ dữ liệu vào và so sánh. Nếu giống nhau thì lặp lại, khác nhau thì bỏ qua nhánh này
 - c Nếu đỉnh của ngăn xếp là ký hiệu \$, chuyển vào trạng thái chấp thuận. Nếu tất cả dữ liệu vào đã được đọc \rightarrow Chấp thuận

Chú ý: Để đưa nhiều ký hiệu vào ngăn xếp ta cần thêm 1 số bước trung gian



Biểu đồ trạng thái của PDA P sẽ có dạng sau:

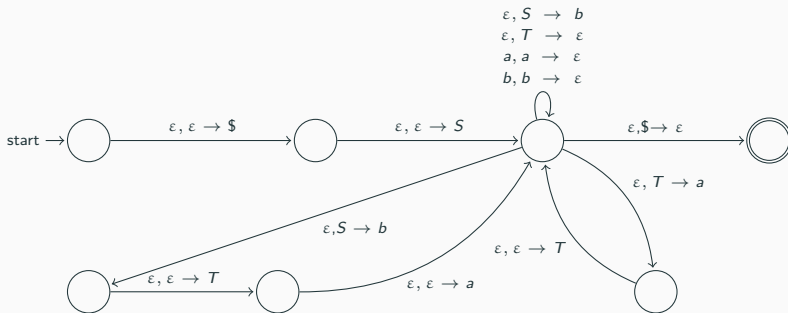


Ví dụ

Cho CFG sau:

$$S \rightarrow aTb \mid b$$

$$T \rightarrow Ta \mid \varepsilon$$

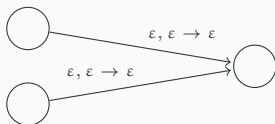


Ý TƯỞNG: Xây dựng CFG từ PDA đã có

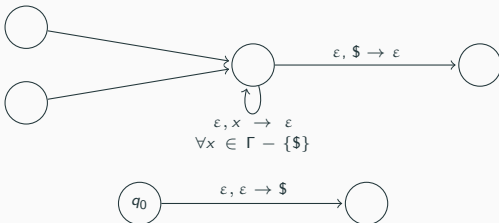
- Bước 1: Đơn giản hóa PDA sao cho có 3 đặc điểm sau:
 - Có duy nhất 1 trạng thái chấp thuận q_{accept}
 - Nó làm rỗng ngăn xếp trước khi chấp thuận
 - Không thực hiện push và pop các ký hiệu vào ngăn xếp cùng 1 lúc
- Bước 2: Xây dựng CFG

Ví dụ đơn giản hóa PDA

- PDA có duy nhất 1 trạng thái chấp thuận q_{accept}



- Nó làm rỗng ngăn xếp trước khi chấp thuận



Ví dụ đơn giản hóa PDA

- Không thực hiện push và pop các ký hiệu vào ngăn xếp cùng 1 lúc



Quy tắc xây dựng CFG

Cho $P = (Q, \Sigma, \Gamma, \delta, q_0, \{q_{accept}\})$ ta xây dựng CFG G với các biến là $\{A_{pq} | p, q \in Q\}$ biến ban đầu là $A_{q_0, q_{accept}}$

- Với mỗi $p, q, r, s \in Q$, $t \in \Gamma$ và $a, b \in \Sigma_\epsilon$ có chứa (r, t) và $\delta(p, a, \epsilon)$ có chứa (q, ϵ) thì ta đưa quy tắc $A_{pq} \rightarrow aA_{rs}b$ vào trong G
- Với mỗi $p, q, r \in Q$ đưa quy tắc $A_{pq} \rightarrow A_{pr}A_{rq}$ vào trong G
- Với mỗi $p \in Q$ đưa quy tắc $A_{pp} \rightarrow \epsilon$ vào trong G

→ Kết thúc chứng minh

Ngôn ngữ không phi ngữ cảnh

- Mọi ngôn ngữ phi ngữ cảnh đều có 1 giá trị đặc biệt được gọi là **độ dài dẫn xuất**
- Có một cách chứng minh ngôn ngữ không phi ngữ cảnh tương tự ngôn ngữ không chính quy

Bổ đề Bơm (Pumping Lemma) cho ngôn ngữ phi ngữ cảnh

Nếu A là một ngôn ngữ phi ngữ cảnh, thì tồn tại một số p sao cho nếu s là một chuỗi bất kỳ thuộc A có độ dài ít nhất là p , thì s có thể được chia ra làm 5 phần $s=uvxyz$ thỏa mãn các điều kiện sau:

1. $uv^i xy^i z \in A \ \forall i \geq 0$
2. $|vy| > 0$
3. $|vxy| \leq p$

- Sử dụng bổ đề Bơm để chứng minh một ngôn ngữ A là không chính quy

Ý TƯỞNG: (Chứng minh bằng phản chứng)

- Giả sử A là chính quy
- Nó có một độ dài dẫn xuất p
- Tất cả các xâu trong A có độ dài lớn hơn p ($|s| \geq p$) có thể chia làm 5 đoạn $s = uvxyz$
- Chọn 1 xâu như vậy trong A
- Chia nó làm 5 đoạn $uvxyz$
- Chỉ ra rằng $uv^ixy^iz \notin A$ bằng cách
 - Xét tất cả các trường hợp mà s có thể chia thành 5 đoạn
 - Chỉ ra rằng không có trường hợp nào thỏa mãn 3 điều kiện của bổ đề Bơm

→ Mâu thuẫn, do đó kết luận A không phải là chính quy

Sử dụng bổ đề Bơm để chứng tỏ rằng $B = \{a^n b^n c^n | n \geq 0\}$ là không phi ngữ cảnh

- Giả sử B là ngôn ngữ phi ngữ cảnh (CFL) $\rightarrow B$ có một độ dài dẫn xuất p
- Xây chúng ta lựa chọn để chỉ ra phản chứng là: $s = a^p b^p c^p$
- Xét các trường hợp có thể chia s thành 5 đoạn $uvxyz$
 - v và y chỉ chứa 1 loại ký tự
 - v và y chứa nhiều ký tự

- Xét TH 1: $a\textcolor{red}{a}abbbbbc\textcolor{green}{c}cc \rightarrow uv^2xy^2z = a\textcolor{red}{a}a|a\textcolor{red}{a}abbbbbc|c\textcolor{green}{c}c$
Xâu $uv^2xy^2z \notin B$
- Tương tự, TH 2: $a\textcolor{red}{a}ab\textcolor{green}{b}bcc \rightarrow uv^2xy^2z = a\textcolor{red}{a}ab|a\textcolor{red}{a}b|b|b\textcolor{green}{c}cc$
 $\notin B$
- Ngoài ra theo điều kiện 3:
 - TH1: $|vxy| = |a\textcolor{red}{a}abbbbbc| = 9 \leq p = 4 \rightarrow \text{False}$
 - TH2: $|vxy| = |a\textcolor{red}{a}b\textcolor{green}{b}| = 5 \leq p = 3 \rightarrow \text{False}$
- Có các mâu thuẫn nên giả thiết là sai $\rightarrow B$ là ngôn ngữ không phi ngữ cảnh

Questions?