Ngôn ngữ không chính quy

Trần Vĩnh Đức

Trường Đại Học Bách Khoa Hà Nội

Ngày 19 tháng 2 năm 2019

Giới thiệu

- Để hiểu khả năng của otomat hữu hạn, ta cũng cần phải hiểu giới hạn của nó.
- Ta sẽ xem xét một phương pháp để chứng minh nhiều ngôn ngữ không đoán nhận bởi otomat hữu hạn.
- Ví dụ, ngôn ngữ
$$B = \{0^n1^n \mid n \geq 0\}$$
không chính quy.

Chính quy hay không?

Ngôn ngữ nào dưới đây không chính quy?

Ví dụ

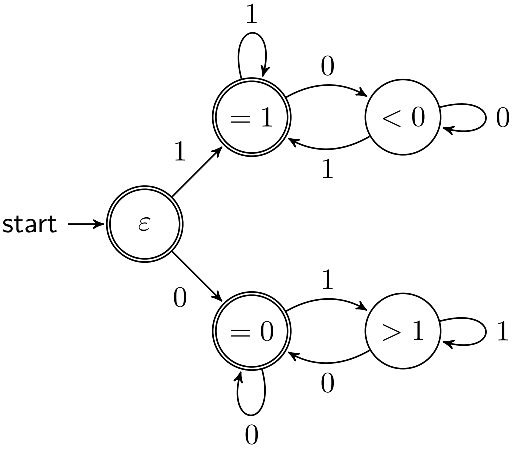
Xét hai ngôn ngữ sau trên bảng chữ $\{0, 1\}$.

- $C = \{w \mid w$ có số ký hiệu 0 bằng số ký hiệu $1\}$
- $H = \{w \mid w$ có số xâu con 01 bằng số xâu con $10\}$

"Nhầm lẫn" của trực giác

$$H = \{w \mid w \text{ có số xâu con } 01 \text{ bằng số xâu con } 10\}$$

là chính quy.



Bổ đề bơm

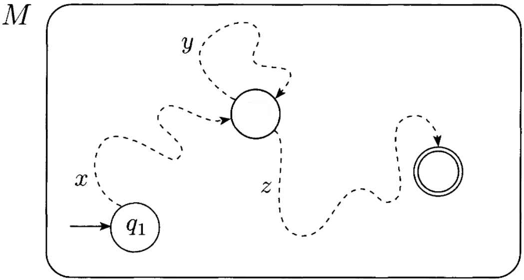
Định lý (Bar-Hillel)

- *Giả thiết:* L *là ngôn ngữ chính quy.*
- *Kết luận:* *Tồn tại một số nguyên dương* p *sao cho mọi xâu* $s \in L$ *và* $|s| \geq p$ *đều có thể viết dưới dạng* $s = xyz$ *thỏa mãn ba điều kiện sau:*
  1. *với mọi số nguyên dương* $i \geq 0$*, xâu* $xy^iz \in L$*,*
  2. $|y| > 0$*,*
  3. $|xy| \leq p$*.*

Có nghĩa rằng, nếu ngôn ngữ L chính quy thì với mọi xâu đủ dài của L, có một xâu con y có thể *bơm* (lũy thừa) lên.

Bổ đề bơm: Ý tưởng chứng minh

- Xét M là DFA đoán nhận L.
- Chọn p bằng số trạng thái của M.
- Với mọi $s \in L$ và $|s| \geq p$, theo nguyên lý Dirichlet, ta có hình sau.



Hình: xâu $s = xyz \in L$ thì $xy^n z \in L$

Ví dụ 1

Xét ngôn ngữ

$$B = \{0^n 1^n \mid n \geq 0\}$$

Ta sẽ dùng bổ đề bơm để chỉ ra B không là chính quy.

- Giả thiết phản chứng B chính quy. Theo bổ đề bơm, tồn tại số nguyên p thỏa mãn điều kiện của bổ đề.
- Chọn $s = 0^p 1^p \in B$, và $|s| = 2p > p$. Vậy ta viết được $s = xyz$ thỏa mãn điều kiện 1 của bổ đề:

$$\forall i \geq 0, \quad xy^i z \in B$$

  Có ba khả năng xảy ra:

1. Xâu y thuộc đoạn chỉ gồm số 0. Vậy $xy^2z \notin B$ vì có nhiều số 0 hơn số 1. Mâu thuẫn.
2. Xâu y thuộc đoạn chỉ gồm toàn số 1. Cũng mâu thuẫn.
3. Xâu y gồm cả 0 và 1. Vậy $xy^2z \notin B$ vì có số 1 đứng trước số 0. Mâu thuẫn.

Ví dụ 2

Xét ngôn ngữ

$$C = \{w \mid w \text{ có số ký hiệu } 0 \text{ bằng số ký hiệu } 1\}$$

Ta sẽ dùng bổ đề bơm để chỉ ra C không là chính quy.

- Giả thiết phản chứng C chính quy. Theo bổ đề bơm, tồn tại số nguyên p thỏa điều kiện của bổ đề.
- Chọn $s = 0^p 1^p \in C$, và $|s| = 2p > p$. Vậy ta viết được $s = xyz$ thỏa mãn

$$\begin{aligned} \forall i \geq 0, \quad & xy^i z \in L && \text{(điều kiện 1)} \\ & |xy| \leq p && \text{(điều kiện 3)} \end{aligned}$$

- Xâu y thuộc đoạn chỉ gồm số 0. Vậy $xy^2z \notin C$ vì có nhiều số 0 hơn số 1. Mâu thuẫn.

Ví dụ 2: Một chứng minh khác

Xét ngôn ngữ

$$C = \{w \mid w \text{ có số ký hiệu } 0 \text{ bằng số ký hiệu } 1\}$$

- Ta biết rằng: "Giao của hai ngôn ngữ chính quy là một ngôn ngữ chính quy".
- Nếu C là chính quy vậy

  $$C \cap 0^*1^* = \{0^n1^n \mid n \geq 0\} = B$$

  là chính quy. Nhưng ta đã chứng minh rằng B *không* là chính quy.
- Vậy C không chính quy.

Ví dụ 3

Xét ngôn ngữ

$$F = \{ww \mid w \in \{0,1\}^*\}$$

Hãy dùng bổ đề bơm để chỉ ra rằng F không chính quy.

- Giả thiết phản chứng F chính quy. Theo bổ đề bơm, tồn tại số nguyên p thỏa điều kiện của bổ đề.
- Chọn $s = 0^p10^p1 \in F$, và $|s| = 2p + 2 > p$.
- Vậy ta viết được $s = xyz$ thỏa mãn

$$\begin{aligned} \forall i \geq 0, \quad & xy^iz \in F && \text{(điều kiện 1)} \\ & |xy| \leq p && \text{(điều kiện 3)} \end{aligned}$$

- Xâu y thuộc đoạn chỉ gồm số 0. Vậy $xy^2z \notin F$. Mâu thuẫn.

Ví dụ 4: Ngôn ngữ trên bảng chữ một chữ

Xét ngôn ngữ

$$D = \{1^{n^2} \mid n \geq 0\}$$

Hãy dùng bổ đề bơm để chỉ ra rằng D không chính quy.

- Giả thiết phản chứng D chính quy. Theo bổ đề bơm, tồn tại số nguyên p thỏa điều kiện của bổ đề.
- Chọn $s = 1^{p^2} \in D$, và $|s| = p^2 > p$.
- Vậy ta viết được $s = xyz$ thỏa mãn

$$\forall i \geq 0, \quad xy^iz \in D \qquad \text{(điều kiện 1)}$$
$$0 < |y| \leq p \qquad \text{(từ điều kiện 2 và 3)}$$

- Ta có

$$p^2 < |xy^2z| = |xyz| + |y| \leq p^2 + p < (p+1)^2$$

Vậy $xy^2z \notin D$. Mâu thuẫn.

Ví dụ 5: Bơm xuống

Xét ngôn ngữ

$$E = \{0^i 1^j \mid i > j\}$$

Hãy dùng bổ đề bơm để chỉ ra rằng E không chính quy.

- Giả thiết phản chứng E chính quy. Theo bổ đề bơm, tồn tại số nguyên p thỏa điều kiện của bổ đề.
- Chọn $s = 0^{p+1}1^p \in E$, và $|s| = 2p + 1 > p$.
- Vậy ta viết được $s = xyz$ thỏa mãn

$$\begin{aligned} \forall i \geq 0, \quad & xy^i z \in E && \text{(điều kiện 1)} \\ & |y| > 0 && \text{(điều kiện 2)} \\ & |xy| \leq p && \text{(điều kiện 3)} \end{aligned}$$

- Vậy y chỉ gồm các số 0.
- Xâu $xy^0 z = xz \notin E$ vì $|xz|_0 \leq |xz|_1 = p$. Mâu thuẫn.

Chú ý

Bổ đề bơm cho ta một *điều kiện cần*, nhưng không phải *đủ*.

Ví dụ

Xét $L \subset 1^*$ là một ngôn ngữ không chính quy trên bảng chữ một chữ. Hai ngôn ngữ *không chính quy* sau đây thỏa mãn điều kiện của bổ đề bơm.

- $0^+L \cup 1^*$
- $1^* \cup 0L \cup 00^+\{0,1\}^*$

Bài tập

- Xâu $w \in \Sigma^+$ gọi là *nguyên thủy* nếu w không là lũy thừa của một xâu khác. Có nghĩa rằng

  $$\text{nếu } w = u^n \text{ với } n \geq 1, \text{ thì } n = 1.$$

- Ta ký hiệu Q là tập mọi xâu nguyên thủy trên bộ chữ $\Sigma = \{0, 1\}$.

  $$Q = \{0, 1, 01, 10, 001, 010, 011, 100, 101, 110, 0001, \dots\}$$

- Hãy chứng minh rằng ngôn ngữ Q không chính quy.