第六讲

正则语言的判定性质

正则语言的判定性质

- 基本问题
- 泵引理
- 非正则语言的判定

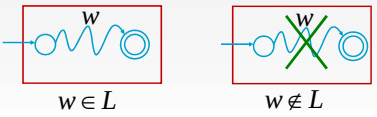
正则语言的判定性质

- 基本问题
- 泵引理
- 非正则语言的判定

字符串与语言

问题: 给出正则语言 L 和字符串 w，如何判定 $w \in L$ ？

方法: 取接受语言 L 的DFA M，判断字符串 w 是否被M接受。



$w \in L$ $w \notin L$

判定算法

- 以 DFA 表示正则语言：
  - 判定算法

    从初态开始，输入字符串 w ，若可以到达某一终态，则该正则语言中接收 w，否则不接收 w。
  - 算法复杂度

    设输入字符串 w 的长度 $|w|=n$，上述判定算法的复杂度为 O(n).

判定算法

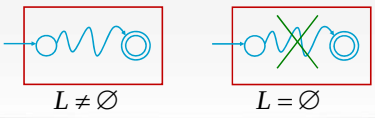
- 以 NFA 表示正则语言：
  - 可将其转化为等价的 DFA，然后执行上述过程；也可直接模拟处理字符串 w 的过程。
  - 判定算法的复杂度为 $O(ns^2)$, 其中n为 w 的长度，s为 NFA 的状态数目。
- 以正则表达式表示正则语言：

  将其转化为等价的 NFA ，然后执行上述过程。

正则语言为空

问题： 给定一正则语言L，怎样判断L是否为空，$L=\varnothing$？

方法： 取接受语言L的DFA M，判断M是否有从初态到终态的路径.



$L\neq\varnothing$ $L=\varnothing$

判定算法

- DFA表示的正则语言

  若某一终态可达，则该语言非空

– 判定算法：由如下步骤递归地计算可达状态集合：
  - ✓ 基础：初态是可达的；
  - ✓ 归纳：设状态 q 是可达的，若对于某个输入符号或 ε，q 可转移到 p，则 p 也是可达的；

– 算法复杂度：设有限自动机的状态数目为 n，上述判定算法的复杂度为 $O(n^2)$。

判定算法

- 正则表达式表示正则语言

– 判定算法：归纳正则表示的语言是否为空。

基础：$L(\varnothing)$ 为空语言，$L(\varepsilon)$ 和 $L(a)$ 不为空；

归纳：设 $R=R_1+R_2$，$L(R)$ 为空 *iff* $L(R_1)$ 和 $L(R_2)$ 都为空；
设 $R=R_1R_2$，$L(R)$ 为空 *iff* $L(R_1)$ 或 $L(R_2)$ 为空；
设 $R=R_1^*$，$L(R)$ 非空（至少包含 ε）；
设 $R=(R_1)$，$L(R)$为空 *iff* $L(R_1)$ 为空.

– 算法复杂度：设正则表示包含的符号数目为 n，上述判定算法的复杂度为 O(n)。

正则语言相等

问题: 给出两个正则语言L_1和L_2，判定$L_1=L_2$？

方法: 判断是否有：

$$(L_1\cap\overline{L_2})\cup(\overline{L_1}\cap L_2)=\varnothing$$

或者

$$L_1\subseteq L_2 \quad \text{且} \quad L_2\subseteq L_1$$

判定算法

- 判定算法采取如下步骤

将两个正则语言转化为 DFA，证明其DFA是否等价
- 适当重命名，使两个DFA没有重名的状态；
- 将两个DFA相并，构造新的DFA M；
- 对M运用填表算法，证明原两个初态是否不可区别.

- 算法复杂度
  - 上述算法的复杂度，其上限为$O(n^4)$；
  - 适当设计填表算法的数据结构，其复杂度降为 $O(n^2)$.

无限正则语言

问题: 给定正则语言L，判断L是否无限

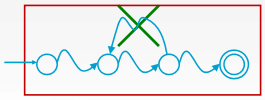
解答: 取接受语言L的DFA，判断初态和终态之间是否有环路。

正则语言无限

DFA

L 是无限的

DFA



L 是有限的

非正则语言

问题: 怎样判定语言L是非正则的？

证明没有DFA可以接受L？

不可能！

非正则语言

问题: 怎样判定语言L是非正则的？

证明没有DFA可以接受L？

方法: 泵引理！

正则语言的判定性质

- 基本问题
- 泵引理
- 非正则语言的判定

非正则语言

正则语言

$a*b$

$b*c+a$

$b+c(a+b)*$

$etc...$

非正则语言

$\{a^n b^n \mid n \geq 0\}$

$\{vv^R \mid v \in \{a,b\}^*\}$

非正则语言

- 鸽巢原理
- 泵引理

鸽巢原理

n 只鸽子:

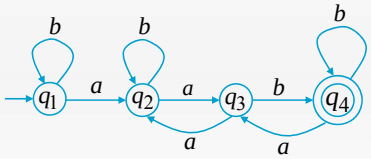
m 个鸽巢:

$n > m$

一个鸽巢中至少有两只鸽子

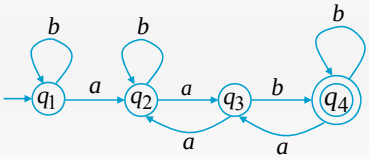
鸽巢原理与 DFA

有4个状态的DFA



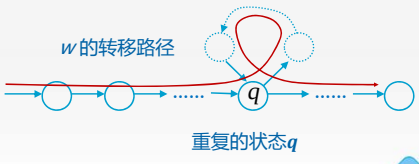
鸽巢原理与 DFA

若$|w| \geq 4$ ⇨ 状态一定会出现重复



鸽巢原理与 DFA

一般地，若$|w|$ ≥状态数 ⇨ *w* 的路径中一定有重复的状态

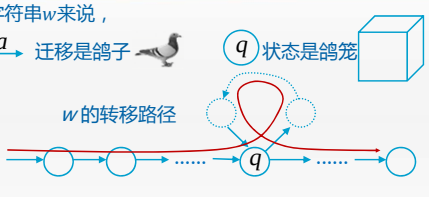


鸽巢原理与 DFA

对字符串*w*来说，

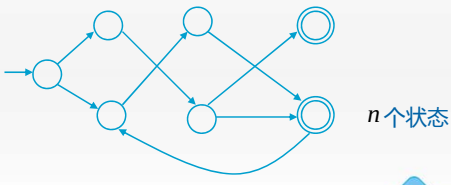
a 迁移是鸽子

q 状态是鸽笼



泵引理

取一个无限的正则语言 *L*，存在接受*L*的DFA



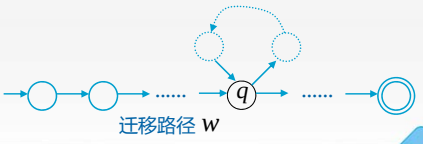
泵引理

取字符串 $w \in L$，
存在一条转移路径 w：

泵引理

若 $|w| \geq n$（n为DFA中状态的数量）。根据鸽巢原理，一定有状态在的路径 w 中重复出现.

设 q 为 w 的迁移路径中第一个重复出现的状态。

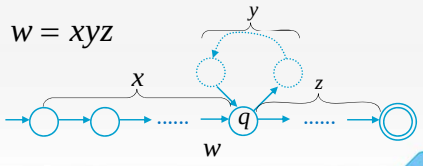


泵引理

根据设计 $|xy| \leq n$，$|y| \geq 1$. 接受的字符串：

$xz \in L \quad xyyz \in L \quad xyyyz \in L$ .....

$w = xyz$

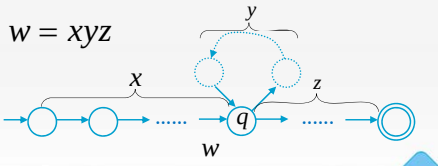


泵引理

一般地：

$xy^i z \in L \qquad i = 0, 1, 2, \ldots$

$w = xyz$



泵引理

泵引理：（正则语言的必要条件）

给定一无限正则语言 L，存在一正整数 m.
对$\forall w \in L$，$|w| \geq m$，$\exists$ x、y、z，$w=xyz$，满足：

$|xy| \leq m$ 且 $|y| \geq 1$

有

$w_i = xy^i z \in L, \quad i=0,1,2,\ldots$

正则语言的判定性质

- 基本问题
- 泵引理
- 非正则语言的判定

泵引理

泵引理：（正则语言的必要条件）

给定一无限正则语言L，存在一正整数 m.
对$\forall w \in L$，$|w| \geq m$，$\exists$ x、y、z，$w=xyz$，满足：

$|xy| \leq m$ 且 $|y| \geq 1$

有

$w_i = xy^i z \in L$, $i=0,1,2,...$

泵引理的应用

- 证明一个语言不是正则语言

Pumping 引理的条件可形式表示为：

$\exists m \forall w \exists x \exists y \exists z \forall k (w \in L \wedge |w| \geq m \rightarrow w=xyz \wedge y \neq \varepsilon \wedge |xy| \leq m \wedge (k \geq 0 \rightarrow xy^k z \in L))$

该命题的否定形式为：

$\forall m \exists w \forall x \forall y \forall z \exists k (w \in L \wedge |w| \geq m \wedge (w=xyz \wedge y \neq \varepsilon \wedge |xy| \leq m \rightarrow k \geq 0 \wedge xykz \notin L))$

泵引理的应用

- 证明步骤

$\forall m \exists w \forall x \forall y \forall z \exists k (w \in L \wedge |w| \geq m \wedge (w=xyz \wedge y \neq \varepsilon \wedge |xy| \leq m \rightarrow k \geq 0 \wedge xy^k z \notin L))$

1. 选任意的m.
2. 找到一个长度至少为m的串$w \in L$.
3. 选满足$w=xyz \wedge y \neq \varepsilon \wedge |xy| \leq m$的$x, y, z$
4. 找到一个 $k \geq 0$，使 $xy^k z \notin L$.

非正则语言的判定

例：证明语言

$L_{01} = \{ 0^n 1^n \mid n \geq 1 \}$

不是正则的。

证明：选任意的m.

取$w = 0^m 1^m$.

选x, y, z，满足条件

$w=xyz \wedge y \neq \varepsilon \wedge |xy| \leq m$

若取 $k=0$, 则有 $xy^k z = xz \notin L_{01}$

因此语言L_{01}不是正则的

非正则语言的判定

非正则语言 $L = \{vv^R \mid v \in \Sigma^*\}$

非正则语言的判定

语言 $L = \{vv^R \mid v \in \Sigma^*\}$ $\Sigma = \{a, b\}$ 不是正则的

证明：

L 是无限的，
可以利用泵引理

非正则语言的判定

$$L=\{vv^R \mid v\in\Sigma^*\}$$

对任何正整数 m，设字符串 w 满足：

$$w\in L \quad 且 \quad |w|\ge m$$

取： $w=a^m b^m b^m a^m$

w可以写成： $a^m b^m b^m a^m = x\,y\,z$

满足： $|x\,y|\le m,\ \ |y|\ge 1$

非正则语言的判定

因此 $y=a^k,\ \ k\ge 1$

$$xyz=\overbrace{\underbrace{a\ldots a}_{x}\underbrace{a\ldots a}_{y}\ldots a}^{m}\underbrace{\overbrace{ab\ldots b}^{m}\overbrace{b\ldots b}^{m}\overbrace{a\ldots a}^{m}}_{z}$$

取 i=2

$$xy^2z=\overbrace{\underbrace{a\ldots a}_{x}\underbrace{a\ldots a}_{y}\underbrace{a\ldots a}_{y}\ldots a}^{m+k}\underbrace{\overbrace{ab\ldots b}^{m}\overbrace{b\ldots b}^{m}\overbrace{a\ldots a}^{m}}_{z}$$

非正则语言的判定

$$xy^2z=a^{m+k}b^m b^m a^m$$

而: $L=\{vv^R \mid v\in\Sigma^*\}$

$$\Downarrow$$

$$a^{m+k}b^m b^m a^m \notin L$$

结论: L 不是正则语言

例 1

非正则语言 $L=\{a^n b^l c^{n+l} \mid n,l\ge 0\}$

正则语言

例 2

证明: $L=\{a^n b^l c^{n+l} \mid n,l\ge 0\}$

对任何正整数 m，取一字符串 w满足：

$$w\in L \quad 且 \quad |w|\ge m$$

设： $w=a^m b^m c^{2m}$

w 可以写成：$a^m b^m c^{2m}=x\,y\,z$

满足： $|x\,y|\le m,\ \ |y|\ge 1$

例 1

因此: $y=a^k,\ \ k\ge 1$

$$xyz=\overbrace{\underbrace{a\ldots a}_{x}\underbrace{a\ldots a}_{y}\ldots a}^{m}\underbrace{\overbrace{ab\ldots b}^{m}\overbrace{c\ldots c}^{2m}\ldots c}_{z}$$

取i=0

$$xz=xy^0z=\overbrace{\underbrace{a\ldots a}_{x}\underbrace{a\ldots a}_{y}\ldots a}^{m-k}\underbrace{\overbrace{ab\ldots b}^{m}\overbrace{c\ldots c}^{2m}\ldots c}_{z}$$

例 1

$xz = xy^0z = a^{m-k}b^mc^{2m}$

但是: $L = \{a^n b^l c^{n+l} \mid n, l \geq 0\}$

$a^{m-k}b^mc^{2m} \notin L$

结论: L 不是正则语言

例 2

非正则语言 $L = \{a^{n!} : n \geq 0\}$

正则语言

例 4

证明： $L = \{a^{n!} : n \geq 0\}$

对任何正整数 m ，取字符串 w 满足：

$w \in L$ 且 $|w| \geq m$

取： $w = a^{m!}$

w 可以写成： $a^{m!} = xyz$

满足： $|xy| \leq m, \quad |y| \geq 1$

例 2

因此: $y = a^k, \quad 1 \leq k \leq m$

$xyz = a^{m!} = \overbrace{a\ldots aa\ldots aa}^{m}\overbrace{\ldots aa\ldots a}^{m!-m}$ （x | y | z）

取i=2

$xy^2z = \overbrace{a\ldots aa\ldots aa}^{m+k}\overbrace{\ldots aa\ldots aa\ldots a}^{m!-m}$ （x | y | y | z）

例 2

$xy^2z = a^{m!+k}$

由于 $L = \{a^{n!} \mid n \geq 0\}$

如果有 $a^{m!+k} \in L$

那么要求存在 p 满足 $m!+k = p!$

例 2

然而: 对于 $m > 1, \quad k \leq m$

$$\begin{aligned} m!+k &\leq m!+m \leq m!+m! \\ &< m!m + m! \\ &= m!(m+1) = (m+1)! \end{aligned}$$

$\Rightarrow m!+k < (m+1)!$

$\Rightarrow m!+k \neq p!$ 对于任意 p 均成立

结论： L 不是正则语言

例 3

非正则语言 $L=\{0^p \mid p \text{ 为素数}\}$

例 3

证明: $L=\{0^p \mid p \text{ 为素数}\}$

任取正整数 m

因为素数无界，取素数 $p > m$, $w = 0^p$，则 $w \in L$，且 $|w| > m$。

设 $w = xyz, |xy| \le m, y = 0^j, 1 \le j \le m$

例 3

选取 $k = p - j$

$$\begin{aligned} xy^k z &= xy^{p-j}z \\ &= xyy^{p-j-1}z = 0^{p+(p-j-1)j} \\ &= 0^{(j+1)(p-j)} \notin L \end{aligned}$$

因此语言 $\{0^p | p$为素数$\}$ 不是正则语言

说明

Pumping 引理不是正则语言的充分条件

反例

a, *b*, *c* 串构成的语言

$L = \{a^i b^j c^k \mid i, j, k \ge 0$,若 $i = 1$ 则 $j = k\}$

课后练习

✧ **必做题:**

- **P-129 Ex.4.1.1 (d), (e)**
- **P-130 Ex.4.1.2 (e), (g), (f)**

✧ **思考题:**

- **P-130 *Ex.4.1.2*(c)**
- **P-130 *Ex.4.1.3*(a)**

Thank you