

形式语言与自动机导论

内容

- 语言、文法、自动机的基本概念
- 有穷自动机
- 正则语言、正则表达式、正则文法
- 正则语言的性质
- 上下文无关语言
- 下推自动机
- 上下文无关语言的性质
- 图灵机与短语结构文法
- 线性有界自动机与上下文有关文法
- 计算理论的几个相关概念



语言、文法、自动机的基本概念

- 字母表、字母、句子、字母表上的语言、语言的基本概念
- 文法的直观意义与形式定义；文法的推导与归约；文法产生的语言、句子、句型；文法的等价；文法的构造。
- 自动机模型、自动机工作过程、自动机接受的语言



语言、文法、自动机的基本概念

- 文法是定义语言的一个数学模型，而自动机可看作是语言的识别系统。
- 通过对一些定理的证明，说明对于一个文法产生的语言，可以构造相应自动机接受该语言；一个自动机接受的语言，可以构造对应的文法产生该语言。一定类型的自动机和一定类型的文法具有等价性。



语言、文法、自动机的基本概念

正则语言	正则文法	有穷自动机
上下文无关语言	上下文无关文法	下推自动机
上下文相关语言	上下文相关文法	线性有界自动机
短语结构语言	短语结构文法	图灵机



有穷自动机

主要内容:

- 确定的有穷状态自动机(**DFA**)
 - ∞ 作为对实际问题的抽象、直观物理模型、形式定义,
DFA接受的句子、语言, 状态转移图。
- 不确定的有穷状态自动机(**NFA**)
 - ∞ 定义;
 - ∞ **NFA**与**DFA**的等价性;
- 带空移动的有穷状态自动机(ϵ -**NFA**)
 - ∞ 定义;
 - ∞ ϵ -**NFA**与**DFA**的等价性。
- **FA**是正则语言的识别器



有穷自动机

本章讨论正则语言的识别器——FA。
包括DFA、NFA、 ε -NFA。

- (1) FA M 是一个五元组, $M=(Q, \Sigma, \delta, q_0, F)$, 它可以用状态转换图表示。
- (2) DFA M 接受的语言为 $\{x \mid x \in \Sigma^* \text{ 且 } \delta(q, x) \in F\}$ 。如果 $L(M_1)=L(M_2)$, 则称 M_1 与 M_2 等价。
- (3) FA的状态具有有穷的存储功能。这一特性可以用来构造接受一个给定语言的FA。

有穷自动机

(4) **NFA**允许**M**在一个状态下读入一个字符时选择地进入某一个状态，对于 $\forall x \in \Sigma^*$ ，如果 $\delta(q_0, x) \cap F \neq \emptyset$ ，则称**x**被**M**接受，如果 $\delta(q_0, x) \cap F = \emptyset$ ，则称**M**不接受**x**。**M**接受的语言 $L(M) = \{x \mid x \in \Sigma^* \text{ 且 } \delta(q_0, x) \cap F \neq \emptyset\}$ 。



有穷自动机

- (5) ϵ -NFA是在NFA的基础上，允许直接根据当前状态变换到新的状态而不考虑输入带上的符号。对 $\forall q \in Q$ ， $\delta(q, \epsilon) = \{p_1, p_2, \dots, p_m\}$ 表示M在状态q不读入任何字符，可以选择地将状态变成 p_1 、 p_2 、...或者 p_m 。这叫做M在状态q做一个空移动。
- (6) NFA与DFA等价， ϵ -NFA与NFA等价，统称它们为FA。
- (7) FA是正则语言的识别模型。



正则语言、正则表达式、正则文法

■ 主要内容

- ∞ 正则表达式**RE**。

- ∞ 典型**RE**的构造。

- ∞ 与**RE**等价**FA**的构造。

- ∞ 与**DFA**等价的**RE**的构造。

- ∞ 正则文法(**RG**)、左线性文法、右线性文法

- ∞ **RG**与**FA**的等价性、相互转换方法。



正则语言、正则表达式、正则文法

本章讨论了RL及其与FA的等价性、RG及其与FA的等价性。

(1) 字母表 Σ 上的RE用来表示 Σ 上的RL。 Φ 、 ε 、 a ($a \in \Sigma$)，是 Σ 上的最基本的RE，它们分别表示语言 Φ 、 $\{\varepsilon\}$ 、 $\{a\}$ ，以此为基础，如果 r 和 s 分别是 Σ 上的表示语言 R 和 S 的RE，则 $r+s$ 、 rs 、 r^* 分别是 Σ 上的表示语言 $R \cup S$ 、 RS 、 R^* 的RE。如果 $L(r)=L(s)$ ，则称 r 与 s 等价。



正则语言、正则表达式、正则文法

(2) RE对乘、加满足结合律；乘对加满足左、右分配律；加满足交换律和幂等律。

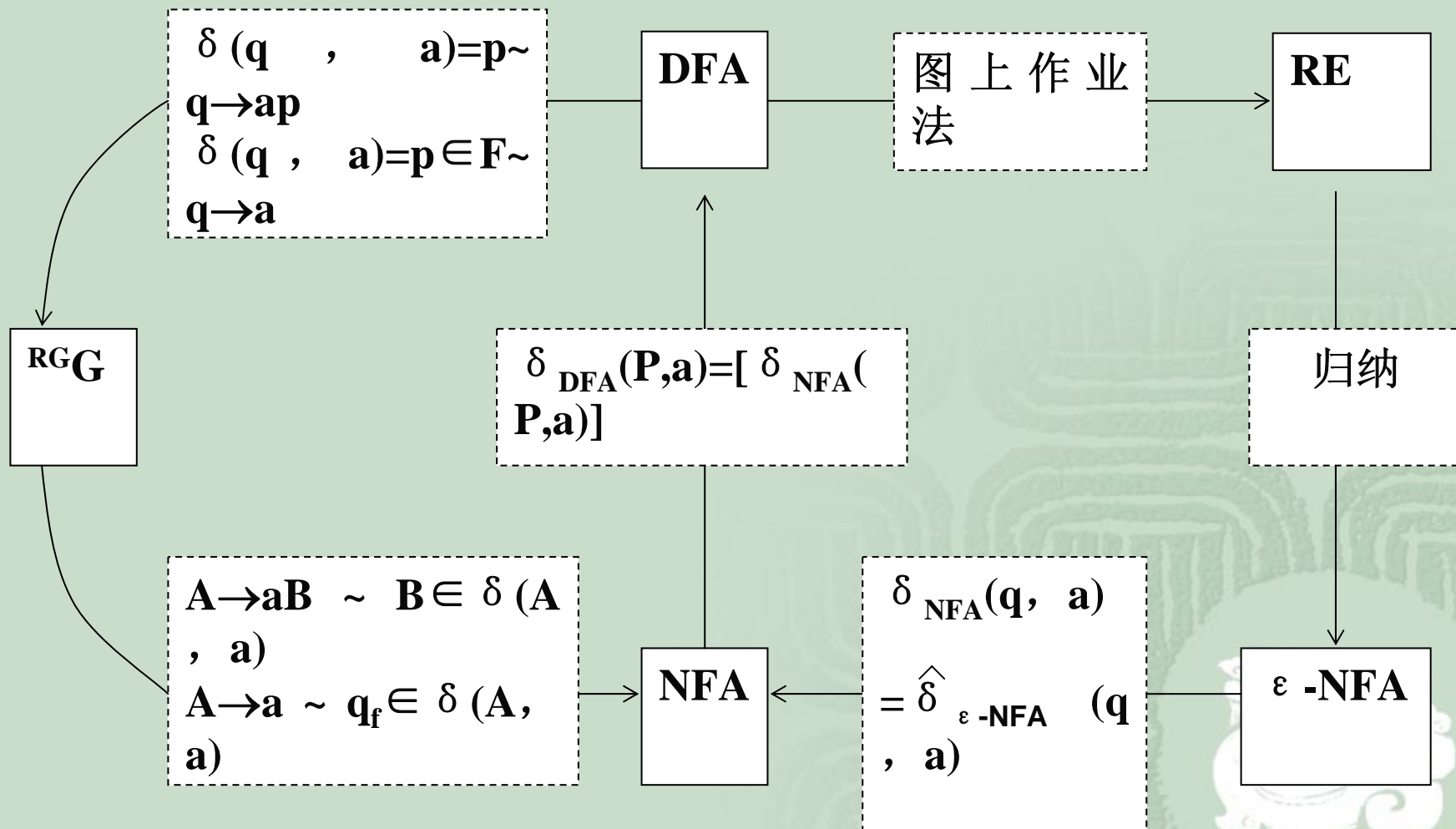
(3) RE是RL的一种描述。容易根据RE构造出与它等价的FA。反过来，可以用图上作业法构造出与给定的DFA等价的RE。

(4) 分别按照对推导和归约的模拟，可以证明FA和左线性文法、右线性文法等价。

(5) RL的5种等价描述模型转换图。



正则语言等价模型的总结



正则语言的性质

■ RL性质

- ∞ 泵引理及其应用

- ∞ 并、乘积、闭包、补、交的封闭性

- ∞ 正则代换、同态、逆同态的封闭性

■ 从RL固有特征寻求表示的一致性

- ∞ Myhill-Nerode定理

- ∞ FA的极小化

■ RL的几个判定问题

- ∞ 空否、有穷否、两个DFA等价否、成员关系



正则语言的性质

本章讨论了**RL**的性质。包括：**RL** 的泵引理，**RL** 关于并、乘积、闭包、补、交、正则代换、同态、逆同态等运算的封闭性。**Myhill-Nerode**定理与**FA**的极小化。**RL**的几个判定问题。

- (1) 泵引理。泵引理是用**RL**的必要条件来用来证明一个语言不是 **RL** 的。它不能用来证明一个语言是 **RL**，而且是采用反证法。

正则语言的性质

- (2) **RL** 对有关运算的封闭性。**RL** 在并、乘、闭包、补、交、正则代换、同态映射运算下是有效封闭的。**RL** 的同态原像是 **RL**。
- (3) 设 $L_1, L_2 \subseteq \Sigma^*$ ，如果 L_1 是 **RL**，则 L_1/L_2 也是 **RL**。
- (4) R_M ， R_L ，Myhill-Nerode定理。



正则语言的性质

- (5) 如果 L 是 RL ，则根据 R_L 确定的 Σ^* 的等价类可以构造出接受 L 的最小 DFA 。更方便的方法是通过确定给定 DFA 状态的可区分性构造出等价的最小 DFA 。
- (6) 存在判定 $L(M)$ 是否非空、 $L(M)$ 是否无穷、 M_1 与 M_2 是否等价、 x 是不是 RL L 的句子的算法。



上下文无关语言

■ 主要内容

∞ 关于CFL的分析

- 派生和归约、派生树

∞ CFG的化简

- 无用符、单一产生式、空产生式

∞ CFG的范式

- 乔姆斯基范式CNF
- 格雷巴赫范式GNF

∞ CFG的自嵌套特性



上下文无关语言

本章讨论了CFG的派生与归约，派生树，A子树，最左派生与最右派生，派生与派生树的关系，二义性文法与固有二义性语言，句子的自顶向下分析和自底向上分析；无用符号的消去算法，空产生式的消除，单一产生式的消除。CFG的CNF和GNF。自嵌套文法。

(1) $S \Rightarrow^* \alpha$ 的充分必要条件为G有一棵结果为 α 的派生树。

(2) 如果 α 是CFG G的一个句子，则G中存在 α 的最左派生和最右派生。

上下文无关语言

(3) 文法可能是二义性的，但语言只可能是固有二义性的，且这种语言是存在的。

(4) 对于任意CFG G ， $\varepsilon \notin L(G)$ ，存在等价的CFG G_1 ， G_1 不含无用符号、 ε -产生式和单一产生式。

(5) 对于任意CFG G ， $\varepsilon \notin L(G)$ ，存在等价的 CNF G_2 。

(6) 对于任意CFG G ， $\varepsilon \notin L(G)$ ，存在等价的GNF G_3 。

(7) CFG $G=(V, T, P, S)$ 是化简后的文法，如果 G 中存在有形如 $A \Rightarrow^+ \alpha A \beta$ 的派生，其中 $\alpha, \beta \in (V \cup T)^*$ ，且 α 和 β 中至少有一个不为 ε ，则称 G 为自嵌套文法。

下推自动机

■ 主要内容

- ∞ PDA的基本概念。

- ∞ PDA的构造举例。

- ∞ 用终态接受语言和用空栈接受语言的等价性。

- ∞ PDA是CFL的接受器。



下推自动机

- **PDA M**是一个七元组： $M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$
- 它是**CFL**的识别模型，它比**FA**多了栈符号，这些符号和状态一起用来记录相关的语法信息。
- 在决定移动时，它将栈顶符号作为考虑的因素之一。



下推自动机

- **PDA**可以用终态接受语言，也可以用空栈接受语言。
- 与**DFA**不同， $\forall (q, a, Z) \in Q \times \Sigma \times \Gamma$ ，**DPDA**仅要求 $|\delta(q, a, Z)| + |\delta(q, \varepsilon, Z)| \leq 1$



下推自动机

关于CFG和PDA主要有如下结论:

- (1) 对于任意PDA M_1 , 存在PDA M_2 , 使得 $N(M_2) = L(M_1)$;
- (2) 对于任意PDA M_1 , 存在PDA M_2 , 使得 $L(M_2) = N(M_1)$;
- (3) 对于任意CFG G , 存在PDA M , 使得 $N(M) = L(G)$;
- (4) 对于任意的PDA M , 存在CFG G 使得 $L(G) = N(M)$ 。

上下文无关语言的性质

■ 主要内容

∞ CFL的泵引理及其应用

∞ CFL的封闭性

■ 封闭运算：并、乘、闭包、代换、同态映射、逆同态映射

■ 不封闭运算：交、补

∞ CFL的判定算法。

■ 判定CFG产生的语言是否为空、是否有穷、一个给定的符号串是否为该文法产生的语言的一个句子等问题。

上下文无关语言的性质

本章讨论了CFL 的性质和CFL的一些判定问题。

(1) 泵引理：与RL的泵引理类似，CFL的泵引理用来证明一个语言不是 CFL。它不能证明一个语言是 CFL。



上下文无关语言的性质

- (2) **CFL** 在并、乘、闭包、代换、同态映射、逆同态映射等运算下是封闭的。
- (3) **CFL** 在交、补运算下是不封闭的。但 **CFL** 与 **RL** 的交是 **CFL**。
- (4) 存在判定 **CFG** 产生的语言是否为空、是否有穷，以及一个给定的符号串是否为该文法产生的语言的一个句子的算法。



图灵机

■ 主要内容

- ∞ TM作为一个计算模型，它的基本定义，即时描述，TM接受的语言；
- ∞ TM的构造技术；
- ∞ TM的变形；
- ∞ 通用TM；
- ∞ TM与PSG的等价性。



图灵机

TM是一个计算模型，用**TM**可以完成的计算被称为是图灵可计算的。

(1) **TM**的基本概念：形式定义、递归可枚举语言、递归语言、部分递归函数、完全递归函数。

(2) 构造技术：状态的有穷存储功能的利用、多道技术、子程序技术。



图灵机

(3) TM的变形：双向无穷带TM、多带TM、不确定的TM，它们都与基本TM等价。

(4) 通用TM可以实现对所有TM的模拟。

(5) 对于任一PSG $G=(V, T, P, S)$ ，存在TM M ，使得 $L(M)=L(G)$ ；

(6) 对于任一TM M ，存在PSG $G=(V, T, P, S)$ ，使得 $L(G) = L(M)$ ；



线性有界自动机与上下文有关文法

■ 主要内容

- ∞ 线性界限自动机(LBA)。

- ∞ LBA作为CSL的识别器。



线性有界自动机与上下文有关文法

本章介绍了识别CSL的装置——LBA。

- (1) LBA是一种非确定的TM，它的输入串被用符号 ϕ 和 $\$$ 括起来，而且读头只能在 ϕ 和 $\$$ 之间移动；
- (2) 如果 L 是 CSL， $\varepsilon \notin L$ ，则存在 LBA M ，使得 $L=L(M)$ ；
- (3) 对于任意 L ， $\varepsilon \notin L$ ，存在 LBA M ，使得 $L=L(M)$ ，则 L 是 CSL。



计算理论的几个相关概念

- 丘奇图灵论题
任何可以机械方式完成的计算都可以用图灵机完成。
- 可计算性和可判定性
机械计算的能力是有限的。
- **P**与**NP**问题

什么是计算机能计算的？

什么是计算机能真实计算的？

