



下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

第七章 下推自动机

姚 刚

中国科学院信息工程研究所



目录

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

- ① 非确定型下推自动机
- ② 下推自动机与上下文无关语言
- ③ 确定型下推自动机和确定型上下文无关语言
- ④ 确定型上下文无关语言的文法



识别上下文无关语言

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

有限接受器可以识别正则语言，但是不能识别所有的上下文无关语言。那么是否存在一类自动机，能够与上下文无关语言相关联？

要识别上下文无关语言，可能需要存储没有边界的信息，例如：在描述 $\{a^n b^n : n \geq 1\}$ 时需要记录 a 的个数；在描述 $\{ww^R\}$ 时需要逆序存储并匹配一个符号序列。

这表明可以尝试用栈作为存储机制，并允许以栈类似的方式操作无限的存储。



下推自动机

下推自动机

姚刚

目录

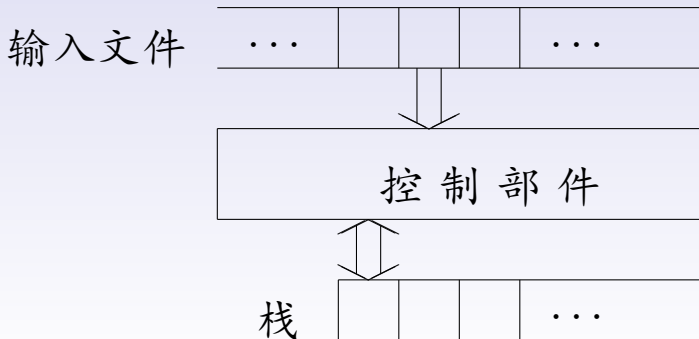
非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

下推自动机(pushdown automata, PDA)的图示:





操作过程

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

在下推自动机中，控制部件的每一次迁移都将从输入文件中读入一个字符，同时通过常规的栈操作来改变栈的内容。控制部件的每一次迁移都是由当前输入符号和当前栈顶符号同时决定的。而迁移的结果将导致控制部件状态以及栈顶符号发生改变。

在这里，我们的讨论只限于作为接受器的下推自动机。



非确定型下推接受器

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

定义 (非确定型下推接受器)

非确定型下推接受器 (*nondeterministic pushdown acceptor, NPDA*) 由一个七元组定义

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

其中,

- Q 是控制部件内部状态的有穷集;
- Σ 是输入字母表;



非确定型下推接受器(续)

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

定义 (非确定型下推接受器(续))

- Γ 是一个有穷符号集, 称为栈字母表(*stack alphabet*);
- $\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma \rightarrow Q \times \Gamma^*$ 的有穷子集, 称为转移函数;
- $q_0 \in Q$ 是控制部件的初始状态;
- $z \in \Gamma$ 是栈开始符(*stack state symbol*);
- $F \subseteq Q$ 是终止状态的集合。



关于 δ 函数

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

- ① δ 函数的第二个参数可以是 ε ，表明自动机的一次迁移可以不需要输入符号，我们称之为 ε 转移。
- ② δ 的执行总需要一个栈符号，如果栈为空，相应的转移将无法完成。
- ③ δ 函数的值域为一个有穷子集是必要的，因为 $Q \times \Gamma^*$ 是一个无穷集合。



例子

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

在一个非确定型下推自动机中，其转移规则集合包含

$$\delta(q_1, a, b) = \{(q_2, cd), (q_3, \varepsilon)\}。$$

我们假定向栈中插入一个符号串的操作是从符号串的右端开始逐个完成的。



例子

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

给定一个非确定型下推自动机中，其中 $Q = \{q_0, q_1, q_2, q_3\}$, $\Sigma = \{a, b\}$, $\Gamma = \{0, 1\}$, $z = 0$, $F = \{q_3\}$, 并且

$$\delta(q_0, a, 0) = \{(q_1, 10), (q_3, \varepsilon)\},$$

$$\delta(q_0, \varepsilon, 0) = \{(q_3, \varepsilon)\},$$

$$\delta(q_1, a, 1) = \{(q_1, 11)\},$$

$$\delta(q_1, b, 1) = \{(q_2, \varepsilon)\},$$

$$\delta(q_2, b, 1) = \{(q_2, \varepsilon)\},$$

$$\delta(q_2, \varepsilon, 0) = \{(q_3, \varepsilon)\}.$$



瞬时描述

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

在一个非确定型下推自动机中，三元组 (q, w, u) 称为下推自动机的瞬时描述(instantaneous description)，其中，

- q 是控制部件的状态，
- w 是输入符号串的未读入部分，
- u 是栈的当前内容(最左边的符号表示栈顶符号)。



迁移的表示

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

从一个瞬时描述到另一个瞬时描述的迁移采用符号 \vdash 来表示, 这样 $(q_1, aw, bx) \vdash (q_2, w, yx)$ 当且仅当 $(q_2, y) \in \delta(q_1, a, b)$ 才是可能的。

包含任意多个步骤的迁移可以用 \vdash^* 表示。在存在多个自动机时, 用 \vdash_M 表明该迁移是由特定的自动机 M 完成的。



下推自动机接受的语言

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

定义

设 $M = (Q, \Sigma, \Gamma, \delta, q_0, z, F)$ 是一个非确定型下推自动机，则 M 接受的语言是集合 $L(M) = \{w \in \Sigma^* : (q_0, w, z) \vdash_M^* (p, \varepsilon, u), p \in F, u \in \Gamma^*\}$ 。总之，由 M 接受的语言就是所有这些符号串的集合，它们都能够在结束接受符号串时将 M 置为终态，而栈的最终内容 u 与这种接受形式的定义是无关的。



例子

下推自动机

姚刚

目录

非确定型下推
自动机

下推自动机与
上下文无关语言

确定型下推自
动机和确定型
上下文无关语言

确定型上下文
无关语言的文法

针对语言 $L = \{w \in \{a, b\}^* : n_a(w) = n_b(w)\}$, 构造一个非确定型下推接受器接受该语言。

针对语言 $L = \{ww^R : w \in \{a, b\}^*\}$, 构造一个非确定型下推接受器接受该语言。



上下文无关语言与下推接受器

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

人上下文无关语言 L 与非确定型下推接受器 M 之间存在着普遍性的关系。

- 对于每个上下文无关语言 L ，都存在一个非确定型下推接受器 M 能够接受它；
- 任何一个非确定型下推接受器 M ，它接受的语言是一个上下文无关语言 L 。



上下文无关语言相应的下推接受器

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

构造一个NPDA能够以某种方式对于上下文无关语言中任何符号串产生一个最左推导。

假定上下文无关语言由格里巴克范式生成，构造NPDA采用如下方式表示推导：将句型的右部变量压入栈中，而使完全由终结符组成的左部作为读输入。

构造一个下推接受器能够接受具有如下产生式的文法生成的语言： $S \rightarrow aSbb|a$ 。



定理

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

定理

对于任何的上下文无关语言 L , 存在一个非确定型下推接受器 M , 使得

$$L = L(M).$$

令 $M = (\{q_0, q_1, q_f\}, T, V \cup \{z\}, \delta, q_0, z, \{q_f\})$ 。其中转移函数包含:

$$\delta(q_0, \varepsilon, z) = \{(q_1, Sz)\};$$

$(q_1, u) \in \delta(q_1, a, A)$, 如果存在 $A \rightarrow au$;

$$\delta(q_1, \varepsilon, z) = \{(q_f, z)\}.$$



例子

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

考虑文法

$$S \rightarrow aA, \quad A \rightarrow aABC|bB|a,$$

$$B \rightarrow b, \quad C \rightarrow c,$$

构造一个下推接受器能够接受该文法生成的语言。



构造等价的NPDA

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

对于一个NPDA，可以构造一个具有下面形式的NPDA与之等价：

- 它只有一个终态 q_f ，并且当且仅当栈为空的时候才进入终态；
- 所有转移函数的形式应该是 $\delta(q_i, aA) = \{c_1, c_2, \dots, c_n\}$ ，其中， $c_i = (q_j, \varepsilon)$ 或者 $c_i = (q_j, BC)$ 。



构造文法

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

构造一个文法，使得文法变量形如 $(q_i A q_j)$ ，而产生式为 $(q_i A q_j) \xRightarrow{*} v$ 当且仅当在读到 v 并且从状态 q_i 向 q_j 转换时，下推自动机擦去 A 。

对于下推接受器的两种形式的转移函数，分别构造产生式 $(q_i A q_j) \rightarrow a$ 和 $(q_i A q_k) \rightarrow a(q_j B q_l)(q_l C q_k)$ ，其中 q_l 和 q_k 取遍 Q 中所有可能的值。



定理

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

考虑NPDA，它的转移函数如下

$$\delta(q_0, a, z) = \{(q_0, Az)\},$$

$$\delta(q_0, a, A) = \{(q_0, A)\},$$

$$\delta(q_0, b, A) = \{(q_1, \varepsilon)\},$$

$$\delta(q_1, \varepsilon, z) = \{(q_2, \varepsilon)\},$$

给出其对应的上下文无关文法。

定理

对于任何非确定型下推接受器 M ，如果 $L = L(M)$ ，则 L 是一个上下文无关语言。



确定型下推自动机

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

定义 (确定型下推自动机)

一个下推自动机 $M = (Q, \Sigma, \Gamma, \delta, q_0, z, F)$ 称为确定型的, 如果它是满足下推自动机的定义, 并且满足如下限制条件: 对于任意 $q \in Q$, $a \in \Sigma \cup \{\varepsilon\}$, 以及 $b \in \Gamma$, 有

- ① $\delta(q, a, b)$ 最多包含一个元素;
- ② 如果 $\delta(q, \varepsilon, b)$ 非空, 则对于每个 $c \in \Sigma$, $\delta(q, c, b)$ 都必须为空。



确定型下推自动机

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

在确定型下推接受器(deterministic push-down acceptor, DPDA)的定义中，第一个条件要求对于任意给定的输入符号与栈顶符号，最多只能执行一种迁移；第二个条件说明，如果某一格局存在 ε 迁移，则不能有读入输入符号的迁移。



确定型上下文无关语言

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

定义 (确定型上下文无关语言)

一个语言 L 是确定型上下文无关语言 (*deterministic context-free language*) 当且仅当存在一个确定型下推自动机 M , 满足 $L = L(M)$ 。

语言 $L = \{a^n b^n : n \geq 0\}$ 是确定型上下文无关语言。

我们给出 $\{ww^R\}$ 的一种非确定上下文无关文法的表示, 但是我们不能就此认为它是非确定型上下文无关语言。



例子

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

事实上, $\{ww^R\}$ 是非确定型上下文无关语言。

注

非确定型下推自动机与确定型下推自动机是不等价的。

设 $L_1 = \{a^n b^n : n \geq 0\}$, $L_2 = \{a^n b^{2n} : n \geq 0\}$, 则 $L = L_1 \cup L_2$ 是非确定型上下文无关语言, 但是不是确定型上下文无关语言。



句型匹配

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

假定我们自顶向下地进行语法分析，并试图为特定的句子找到其最左推导。

a_1	a_2	a_3	$a_4 \dots a_n$	输入 w
a_1	a_2	a_3	$A \dots$	句型
匹配部分			尚未匹配部分	



简单文法匹配

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

对于简单文法，我们在分析的每一步骤中，都可以准确地知道使用的是哪一个产生式：假定 $w = w_1w_2$ ，并且到达了句型 w_1Ax ，为了得到与 w 中下一个符号匹配的句型中的下一个符号，我们只需要看 w_2 的最左符号，比如是 a 。如果文法中不存在 $A \rightarrow ay$ 的产生式，则 w 不属于该语言，如果存在这样的规则，分析可以继续。



LL文法(LL grammar)

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

简单文法要求太严格，我们引入LL文法：如果给定当前输入的符号以及“向前看” $k-1$ 个符号，我们就可以唯一地确定使用哪一个产生式，则该文法称为 $LL(k)$ 文法。

这里，第一个 L 表示采用的是从左到右的扫描方式，第二个 L 表示采用的是最左推导。



例子

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

文法 $S \rightarrow aSb|ab$ 是 $LL(2)$ 文法。

文法 $S \rightarrow SS|aSb|ab$ 不是 LL 文法。

设 L 是文法 $S \rightarrow SS|aSb|ab$ 对应的语言，上面的例子不能说明 L 就是非确定型的，或者不存在产生语言 L 的 LL 文法。

例如文法 $S_0 \rightarrow aSbS, S \rightarrow aSbS|\varepsilon$ ，这个 LL 文法可以生成语言 L 。



LL文法

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

定义 (LL文法)

设 $G = (V, T, S, P)$ 是一个上下文无关文法, 如果对于每个最左推导对

$$S \xRightarrow{*} w_1 A x_1 \Rightarrow w_1 y_1 x_1 \xRightarrow{*} w_1 w_2,$$

$$S \xRightarrow{*} w_1 A x_2 \Rightarrow w_1 y_2 x_2 \xRightarrow{*} w_1 w_3,$$

其中 $w_1, w_2, w_3 \in T^*$, w_2 和 w_3 最左的 k 个符号的等同性意味着 $y_1 = y_2$, 则称 G 为 $LL(k)$ 文法。(如果 w_2 或 w_3 的长度小于 k , 则取它们的最小长度代替 k 。)



LL文法

下推自动机

姚刚

目录

非确定型下推自动机

下推自动机与上下文无关语言

确定型下推自动机和确定型上下文无关语言

确定型上下文无关语言的文法

通过LL文法的定义，我们可以清楚地看到，在最左推导(w_1Ax)的任意阶段，如果我们能够知道输入的下 k 个符号，则推导的下一步是唯一确定的(如同 $y_1 = y_2$ 所表示的)。



下推自动机

姚刚

目录

非确定型下推
自动机

下推自动机与
上下文无关语
言

确定型下推自
动机和确定型
上下文无关语
言

确定型上下文
无关语言的文
法

谢谢！

主 讲 人： 姚 刚

电子邮箱： yaogang@iie.ac.cn