第六章 上下文无关文法的化简与范式

姚 刚

中国科学院信息工程研究所

目录

限制形式

在上下文无关文法的定义中，它没有对产生式的右部做任何限制，然而这种完全自由的形式并不总是必要的，在某些情况下，甚至会对文法产生不良影响。从前面的定理中可以看出，限制形式在应用中是方便的，而且通过消除$A \to \varepsilon$和$A \to B$这样的产生式也可以简化文法。因而，上下文无关文法的限制形式是必要的。

空串现象

就很多文法与语语言而言，空串现象都是一个麻烦的问题。在很多定理及其证明中，空串往往表现为一种特殊的形式，需要特别考虑。

为了化简，我们可以选择消除空串，使得待考察的语言中不包含ε。

消除空串的方法

消除空串不会使文法失去一般性。设L是一个上下文无关语言，而$G=(V,T,S,P)$是表示$L-\{\varepsilon\}$的上下文无关文法。

通过在V中增加S_0，并使其作为初始变量，同时在P中增加$S_0 \rightarrow S|\varepsilon$，这样得到的新文法产生的语言即为$L$。因此，$L-\{\varepsilon\}$总是可以转化为$L$。

对于任意的上下文无关文法G，存在一种方法可以得到$\hat{G}$，并使得$L(\hat{G}) = L(G)-\{\varepsilon\}$。

消除空串

在实际应用中，我们认为包含ε的上下文无关文法与不包含ε的上下文无关文法是没有区别的。

在这一章中，除非特别指出，我们讨论的都是不包含ε 的上下文无关语言。

代入方法

通过代入方法，很多规则能够生成等价文法。下面介绍其中的一种，它可以用来化简文法。

在这里，我们认为化简(simplification)就是对某些不必要的产生式的消除，但是，在实际的化简过程中，它并不必然导致产生式数目的减少。

代入规则

定理

设$G=(V,T,S,P)$是一个上下文无关文法，P中包含一个形如$A \to x_1Bx_2$的产生式，其中A和B是不同的变量，并且$B \to y_1|y_2|\cdots|y_n$是P中所有以B为左部的产生式的集合。而对于上下文无关文法$\hat{G}=(V,T,S,\hat{P})$，其中$\hat{P}$是通过删除P中的产生式$A \to x_1Bx_2$并增加产生式$A \to x_1y_1x_2|x_1y_2x_2|\cdots|x_1y_nx_2$而得到的。则$L(G)=L(\hat{G})$。

例子

考虑文法$G=(\{A,B\},\{a,b,c\},A,P)$，其中产生式为

$$A \to a|aaA|abBc, \quad B \to abbA|b,$$

对变量B运用代入规则。

注

这里一个必要条件是A和B是不同的变量。

删除无用表达式

我们总是希望将在任何推导中不起作用的产生式从文法中删除。例如产生式集合中有$S \to aSb|\varepsilon|A$，$A \to aA$，这里$S \to A$是无用的。

删除不起作用的产生式，对语言没有任何影响，并且，从定义上来看，这也是对文法的化简。

变量是有用的和无用的

定义

设$G = (V, T, S, P)$是一个上下文无关文法，如果变量$A \in V$是有用的(useful)，则当且仅当至少存在一个$w \in L(G)$，使得$S \stackrel{*}{\Rightarrow} xAy \stackrel{*}{\Rightarrow} w$，其中$x, y \in (V \cup T)^*$。而对于不出现在任何句子推导中的变量，我们称之为无用的(useless)。如果某个产生式包含无用变量，则该产生式是无用的。

例子

注

总而言之，某个变量是有用的当且仅当它至少在一个句子的推导中出现。

考虑某个文法中，开始符号为S，产生式包括

$$S \to A,\quad A \to aA|\varepsilon,\quad B \to bA,$$

这里，变量B是无用的。

变量无用的原因

一个变量无用的两个原因：

- 或者无法从开始符号达到它；
- 或者无法从它推导出一个终结符串。

去除无用变量与产生式的过程就是在认识到这两种情况的基础上进行的。

例子

消除文法$G=(V,T,S,P)$中的无用符号与产生式，其中$V=\{S,A,B,C\}$，$T=\{a,b,c\}$，产生式包括

$$S\to aS|A|c,\qquad A\to a,$$

$$B\to aa,\qquad C\to aCb。$$

依赖图

定义

依赖图(*denpendency graph*)作为对复杂关系的一种可视化表示而被广泛应用。对于上下文无关文法的依赖图，其中顶点表示变量，而从顶点C到顶点D之间存在边当且仅当存在如下形式的表达式$C \to xDy$。

如果一个变量是有用的，仅当在依赖图中存在一条路径，它从以S为标记的顶点出发能够到达以该变量标记的顶点。

删除无用产生式

定理

设$G=(V,T,S,P)$是一个上下文无关文法，则存在一个与之等价的文法$\hat{G}=(\hat{V},\hat{T},S,\hat{P})$，它不包含任何的无用符号或者产生式。

第一步：标记有用记号

1. 置V_1为空；
2. 重复下面的步骤直到再也没有变量能够加到V_1中为止：对于任意$A \in V$，如果P中存在形如$A \to x_1 x_2 \cdots x_n$的产生式，其中所有x_i都在$V_1 \cup T$中，则将A加到V_1中；
3. 对于P中的每一个产生式，如果它的符号都在$(V_1 \cup T)$中，则将其加入到P_1中。

第二步：构造新文法

1. 基于G_1的变量依赖图，可以得到所有不能从S到达的变量。
2. 删除这些变量以及与其相关的产生式，同时也删除不出现在有用产生式中的终结符。

ε产生式

定义

上下文无关文法中，任意形如$A \to \varepsilon$的产生式称为ε产生式(ε-*production*)。任何存在推导

$$A \overset{*}{\Rightarrow} \varepsilon$$

的变量A称为是可空的(*nullable*)。

考虑文法$S \to aS_1b$，$S_1 \to aS_1b|\varepsilon$，该文法生成的语言$\{a^nb^n : n \geq 1\}$不包含$\varepsilon$。

删除ε产生式

定理

设$G = (V, T, S, P)$是任意的一个上下文无关文法，并且ε不在$L(G)$中，则存在一个等价的不包含ε产生式的文法$\hat{G}$。

第一步：标记可空变量

1. 对于所有形如$A \to \varepsilon$的产生式，将A加入V_N中。
2. 重复下述步骤直到再也没有新的变量可以加入到V_N中。对于所有形如

$$B \to A_1 A_2 \cdots A_n$$

的产生式，如果A_1，A_2，$\cdots$，A_n都属于V_N，则将B加入到V_N中。

第二步：构造新文法

考察P中所有形如$A \rightarrow x_1x_2\cdots x_m$的产生式，其中$m \geq 1$，$x_i \in V \cup T$。
对于P中每个上述形式的产生式，我们将该产生式以及使用ε替换其可空变量所有可能组合而得到的产生式加入$\hat{P}$。

例子

考虑文法

$$S \to ABaC, \qquad A \to BC,$$

$$B \to b|\varepsilon, \quad C \to D|\varepsilon, \quad D \to d,$$

构造一个等价的上下文无关文法，使得它不包含ε产生式。

单位产生式

定义

上下文无关文法中，任意形如$A \to B$的产生式，其中$A, B \in V$，称为单位产生式。

定理

设$G = (V, T, S, P)$是任意的一个上下文无关文法，并且不包含ε产生式，则存在一个上下文无关文法$\hat{G} = (\hat{V}, \hat{T}, S, \hat{P})$，它与$G$等价且不存在单位产生式。

构造过程

对每个变量A，找到所有的变量B，使得$A \stackrel{*}{\Rightarrow} B$。
先将P中所有的非单位产生式加入$\hat{P}$，然后针对所有满足$A \stackrel{*}{\Rightarrow} B$的$A$与$B$，在$\hat{P}$加入产生式$A \rightarrow y_1|y_2|\cdots|y_n$，其中$B \rightarrow y_1|y_2|\cdots|y_n$是$\hat{P}$中所有$B$为左部的产生式集合。

例子

消除文法

$$S \to Aa|B, \quad B \to A|bb, \quad A \to a|bc|B$$

中的单位产生式。

定理

设L是不包含ε的上下文无关语言，则存在一个产生L的上下文无关文法，该文法不包含无用产生式、ε产生式和单位产生式。

范式

范式(normal form)指的是一种受限制但又具备足够能力的文法形式，任何文法都可以表示成相应的等价范式。

针对上下文无关文法G，我们将讨论一些相应的范式：乔姆斯基范式(Chomsky normal form)和格里巴克范式(Greibach normal form)。

乔姆斯基范式

定义

如果一个上下文无关文法的所有产生式都形如$A \to BC$，或者$A \to a$，其中，$A, B, C \in V$，$a \in T$，则该文法属于乔姆斯基范式。

文法$S \to AS|a$，$A \to SA|b$是乔姆斯基范式。

文法$S \to AS|AAS$，$A \to SA|aa$不是乔姆斯基范式。

乔姆斯基范式

定理

对于任意一个上下文无关文法$G = (V, T, S, P)$，其中$\varepsilon \notin L(G)$，都存在一个等价的上下文无关文法$\hat{G} = (\hat{V}, \hat{T}, S, \hat{P})$属于乔姆斯基范式。

证明

对于$n \geq 2$，将产生式$A \to x_1 x_2 \cdots x_n$改写为$A \to C_1 C_2 \cdots C_n$，（其中$C_i = x_i$，如果$x_i \in T$；$x_i = B_a$，如果$x_i = a$。）并在新文法中加上产生式$B_a \to a$。

在新文法中加上产生式$A \to C_1 D_1$，$D_1 \to C_2 D_2$，$\cdots$，$D_{n-2} \to C_{n-1} C_n$。

例子

将具有产生式

$$S \to ABa,\ \ A \to aab,\ \ B \to Ac$$

的文法转换为乔姆斯基范式。

格里巴克范式

定义

对于一个上下文无关文法，如果所有的产生式都具有如下形式：$A \to ax$，其中，$a \in T$，$x \in V^*$，则该文法属于格里巴克范式。

将具有产生式$S \to AB$，$A \to aA|bB|b$，$B \to b$ 的文法转换为格里巴克范式。

将具有产生式$S \to abSb|aa$的文法转换为格里巴克范式。

格里巴克范式

定理

对于任意一个上下文无关文法$G = (V, T, S, P)$，其中$\varepsilon \notin L(G)$，都存在一个等价的上下文无关文法$\hat{G} = (\hat{V}, \hat{T}, S, \hat{P})$属于格里巴克范式。

成员资格算法

给定上下文无关文法$G = (V, T, S, P)$，它满足乔姆斯基范式，对于符号串$w = a_1a_2\cdots a_n$，定义子串$w_{ij} = a_i\cdots a_j$，以及V的子集$V_{ij} = \{A \in V : A \overset{*}{\Rightarrow} w_{ij}\}$。

显然，$w \in L(G)$当且仅当$S \in V_{1n}$。

成员资格算法(续)

注意到$A \in V_{ii}$当且仅当G中存在产生式$A \to a_i$。

先计算V_{11}，V_{22}，$\cdots$，V_{nn}，

然后依次计算V_{12}，V_{23}，$\cdots$，$V_{n-1,n}$，

再依次计算V_{13}，V_{24}，$\cdots$，$V_{n-2,n}$，

一直下去得到V_{1n}。

这里$V_{ij} =$

$$\bigcup_{k \in \{i, \cdots, j-1\}} \{A : A \to BC, B \in V_{ik}, C \in V_{k+1,j}\}$$

例子

确定符号串$w = aabbb$是否属于由文法

$$S \to AB, \quad A \to BB|a, \quad B \to AB|b$$

生成的语言。

谢谢！

主 讲 人： 姚 刚

电子邮箱： yaogang@iie.ac.cn