可交换上下文无关文法

张继军[1,2] 吴哲辉[2]
（山东农业大学信息学院 泰安 271018）[1] （山东科技大学信息学院 青岛 266510）[2]

摘 要 本文提出了可交换上下文无关文法及其该文法产生的语言——可交换上下文无关语言，证明了正规语言类是可交换上下文无关语言类的一个子集，而可交换上下文无关语言类是上下文无关语言类的一个子集；讨论了可交换上下文无关语言的结构特点，并给出了可交换上下文无关语言的 Pumping 引理。

关键词 可交换上下文无关文法，可交换上下文无关语言，Pumping 引理

Commutative Context-Free Grammar

ZHANG Ji-Jun[1,2] WU Zhe-Hui[2]
(Shandong Agricultural University, Tai'an 271018)[1] (Shandong University of Science and Technology, Qing'dao 266510)[2]

Abstract The concept of commutative context-free grammar is presented in this paper, L is said to be a commutative context-free language if there is a commutative context-free grammar G such that L=L(G); It is proofed that the class language is a proper subclass of context-free language; The structural feature of commutative context-free language is analyzed; The Pumping lemma of commutative context-free language is given.

Keywords Commutative context-free grammar, Commutative context-free language, Pumping lemma

1 引言

自 1956 年 Chomsky 建立文法的数学模型[1]，并以此定义了四类性质不同的文法和语言后，人们对形式语言理论和应用及与形式语言密切关系的自动机理论的研究，都非常活跃且富有成果。所取得的研究成果不仅对编译原理，而且对诸如信息工程、人工智能以及数理语言等领域均产生了深远的影响。随着应用领域的拓广和实际应用的需要，人们对文法增加限制，提出了一些新的文法及相应的语言，并对其性质进行研究，例如：有界上下文无关语言[2]，LL(k)文法、LR(k)文法、算符文法，优先文法[7]等等及其对应的语言，这些文法或语言，都具有不同的结构特点和性质，因而，对这些文法或语言进行语法分析时，根据不同的结构特点，采用不同的分析技术，方便了语法分析。

文法是一种语言的生成规则，而推导是一个句子的形成过程，在本文中，通过分析上下文无关文法的性质和特点，依据推导过程的推导变化特点，提出了可交换上下文无关文法及其可交换上下文无关语言，从传统经典语言中又划分出一个子集，证明了正规语言类是可交换上下文无关语言类的一个子集，而可交换上下文无关语言类是上下文无关语言类的一个子集；分析讨论了可交换上下文无关语言的结构特点，并给出了可交换上下文无关语言的 Pumping 引理。正规语言的 Pumping 引理是可交换上下文无关语言的 Pumping 引理的特例，而可交换上下文无关语言的 Pumping 引理是上下文无关语言的 Pumping 引理的特例。

2 上下文无关文法的性质和特点

定义 2.1 设 $G=(V,\Sigma,P,S)$为一个上下文无关文法，$A\in V$，文法 G 中的一个推导：

$$A\overset{*}{\Rightarrow}w_1Aw_2 \quad w_1、w_2\in\Sigma^* \tag{1}$$

称为变量 A 的一个递归推导。特别地，

当 $w_1\neq\varepsilon, w_2=\varepsilon$ 时，(1)式称为变量 A 的一个右递归推导；

当 $w_2\neq\varepsilon, w_1=\varepsilon$ 时，(1)式称为变量 A 的一个左递归推导；

如果 $w_1\neq\varepsilon$ 且 $w_2\neq\varepsilon$ 时，(1)式称为变量 A 的一个嵌套递归推导。

具有嵌套递归的变量，称为具有自嵌套特性，至少含有一个自嵌套特性的非终结符的文法，称该文法具有自嵌套特性。

显然，当 G 是一个左线性文法时，G 中的任意变量 A 若存在递归推导，那么必然都是左递归的；同样，当 G 是一个右线性文法时，G 中每一个递归推导都是右递归的。

命题 2.1 任何上下无关文法 G，如果不存在某非终结符的自嵌套特性，那么，就等价于一个正规文法；或者说：它只能产生正规语言。自嵌套特性区别了正规语言和上下文无关语言。

定义 2.2 设 $G=(V,\Sigma,P,S)$为一个上下文无关文法，$A\in V$，

$$A\overset{*}{\Rightarrow}w_1Aw_2 \tag{2}$$

为变量 A 的一个递归推导。如果存在一组（不同于(2)式的）A 的递归推导：

$$A\overset{*}{\Rightarrow}w_{i1}Aw_{i2} \quad i=1,2,\cdots,k; \tag{3}$$

反复使用(3)式中的某些递归推导式可以推出(2)式，则称(2)式的递归推导为派生的（或复合的）。非派生的递归推导称为基本递归推导。

定义 2.3 设 $G=(V,\Sigma,P,S)$为一个上下文无关文法，A

张继军 教授，主要研究方向：自动机理论及应用，Petri 网理论及应用；吴哲辉 教授，博导，主要研究方向：Petri 网理论及应用、自动机理论及应用、算法设计与分析等。

$\in V, A\overset{*}{\Rightarrow}w_1Aw_2$ 为变量 A 的一个递归推导：

若 $w_1\neq\varepsilon$ 且 $w_2\neq\varepsilon$，则称 w_1 为相对于变量 A 的增重复序列，而 w_2 为相对于变量 A 的减重复序列，其三者的关系，用一个三元组表示：(A,w_1,w_2)。

若 $w_1\neq\varepsilon\wedge w_2=\varepsilon$ 或 $w_2\neq\varepsilon\wedge w_1=\varepsilon$，则称 w_1 或 w_2 为相对于变量 A 的左(右)不变重复序列。

特别地，当 A 的递归推导是基本递归推导时，则分别称为：基本增重复序列，基本依赖重复序列，基本不变可重复序列。为了方便，统称它们为"重复序列"或"基本重复序列"。

3 可交换上下文无关文法

定义 3.1 设 $G=(V,\Sigma,P,S)$ 为一个上下文无关文法，$A\in V$，

$$A\overset{*}{\Rightarrow}w_{11}Aw_{12}$$

$$A\overset{*}{\Rightarrow}w_{21}Aw_{22}$$

为变量 A 的任意两个递归推导。如果

$$A\overset{*}{\Rightarrow}w_{11}w_{21}Aw_{12}w_{22}$$

和 $A\overset{*}{\Rightarrow}w_{21}w_{11}Aw_{22}w_{12}$

也是 A 的递归推导，则称变量 A 的递归推导是可交换的。

定义 3.2 设 G 为一个上下文无关文法。若 G 的每一个变量的递归推导都可交换，则称 G 为可交换上下文无关文法。

定理 3.1 正规文法都是可交换上下文无关文法。

证明：因为正规文法或者为右线性文法，或者为左线性文法，不妨假设 G 为一个为右线性文法，那么对 G 中的任一变量 A，其递归推导都有 $A\overset{*}{\Rightarrow}w_iA$ 的形式。设

$$A\overset{*}{\Rightarrow}w_1A \text{ 和 } A\overset{*}{\Rightarrow}w_2A$$

为 A 的任意两个递归推导，根据文法的推导规则，可得到

$$A\overset{*}{\Rightarrow}w_1A\overset{*}{\Rightarrow}w_1w_2A$$

$$A\overset{*}{\Rightarrow}w_2A\overset{*}{\Rightarrow}w_2w_1A$$

即 $A\overset{*}{\Rightarrow}w_1w_2A$ 和 $A\overset{*}{\Rightarrow}w_2w_1A$ 也是变量 A 的递归推导。这表明 G 中的任一个变量的递归推导都是可交换的。因此 G 是可交换上下文无关文法。

定理 3.2 若 G 是一个上下文无关文法，且 G 中的每个变量都最多有一个基本的嵌套递归推导，则 G 是可交换上下文无关文法。

证明：设 A 为 G 中的任一个变量，它可能存在的全部基本递归推导为：

$$A\overset{*}{\Rightarrow}w_{11}Aw_{12}\quad w_{11}\neq\varepsilon \text{ 且 } w_{12}\neq\varepsilon \tag{4}$$

$$A\overset{*}{\Rightarrow}w_{i1}A\quad i=1,2,\cdots,p \tag{5}$$

$$A\overset{*}{\Rightarrow}Aw_{j2}\quad j=1,2,\cdots,q \tag{6}$$

即 A 只有(4)式一个是基本的嵌套递归推导，显然从(5)的任一个右递归推导和(6)式的任一个左递归推导都可以得到一个嵌套递归推导，即可得：

$$A\overset{*}{\Rightarrow}w_{i1}Aw_{j2}\qquad i=1,2,\cdots,p;j=1,2,\cdots,q \tag{7}$$

共$(p-1)(q-1)$个嵌套的递归推导，但它们都是派生的(不是基本的)。

式(5)式中任意两个右递归推导的可交换性和式(6)式中任意两个左递归推导的可交换性已在定理 3.1 中得到证明，仿此也容易证明(5)式的任一个右递归推导同(6)式中的任一个左递归推导的可交换性，以及(4)式的嵌套推导同式(5)或(6)的右(左)递归推导的可交换性。下面我们证明(4)式的基本嵌套递归推导同式(7)式中的任一个(派生的)嵌套递归推导的可交换性：

$$A\overset{*}{\Rightarrow}Aw_{j2}\overset{*}{\Rightarrow}w_{11}Aw_{12}w_{j2}\overset{*}{\Rightarrow}w_{11}w_{i1}Aw_{12}w_{j2}\quad i=1,2,\cdots,p;\ j=1,2,\cdots,q$$

$$A\overset{*}{\Rightarrow}w_{i1}A\overset{*}{\Rightarrow}w_{i1}w_{11}Aw_{12}\overset{*}{\Rightarrow}w_{i1}w_{11}Aw_{j2}w_{12}\quad i=1,2,\cdots,p;\ j=1,2,\cdots,q$$

同理也可以证明(7)式中的任意两个嵌套递归推导也是可交换的。

在此基础上，用数学归纳法可以证明变量 A 的任意两个递归推导都是可交换的。由 A 的任意性知 G 是可交换的文法。

并不是所有的上下文无关文法都是可交换的文法。例如：字母表 $\Gamma=\{a,b\}$ 上的回文 $L=\{w\in\{a,b\}^*\mid w=w^R\}$，我们知道产生 L 的文法 G 的产生式集为：

$$P: S\to aSa\mid bSb\mid a\mid b\mid\varepsilon$$

这里，变量 S 存在两个基本嵌套递归推导

$$S\overset{*}{\Rightarrow}aSa\quad \text{和}\quad S\overset{*}{\Rightarrow}bSb$$

然而，从 P 中的产生式集，运用推导规则不可以得到下面的递归推导：

$$S\overset{*}{\Rightarrow}abSab\quad \text{和}\quad S\overset{*}{\Rightarrow}baSba$$

这表明回文文法不是可交换的文法。

定义 3.3 L 称为一个可交换上下文无关语言当且仅当存在一个可交换上下文无关文法 $G=(V,\Sigma,P,S)$，使得 $L=L(G)$。

例如：语言 $L_1=\{a^nb^n\mid n\geqslant0\}$、$L_2=\{(a+b)^+a^nb^n\mid n>1\}$、$L_3=\{a^nb^mc^md^n\mid n,m\geqslant0\}$、$L_4=\{a^nb^nc^md^m\mid n,m\geqslant0\}$、$L_5=\{w=(a+b)^{n1+n2}(c+d)^{n1+n2}\mid n_1,n_2\geqslant0\}$、$L_6=\{(a+b)^n(c+d)^n\mid n\geqslant0\}$都是可交换上下文无关语言。产生它们的文法都是可交换上下文无关文法，且依次是：

$G_1=(\{S\},\{a,b\},P_1,S)$，其中 $P_1: S\to aSb\mid\varepsilon$

$G_2=(\{S,A,B\},\{a,b\},P_2,S)$，其中 $P_2: S\to Aa\mid Ab, A\to Aa\mid Ab\mid aBb, B\to aBb\mid ab$

$G_3=(\{S,A\},\{a,b,c,d\},P_3,S)$，其中 $P_3: S\to aSd\mid A, A\to bAc\mid\varepsilon$

$G_4=(\{S,A,B\},\{a,b,c,d\},P_4,S)$，其中 $P_4: S\to AB, A\to aAb\mid\varepsilon, B\to cBd\mid\varepsilon$

$G_5=(\{S\},\{a,b,c,d\},P_5,S)$，其中 $P_5: S\to aSc\mid bSd\mid abScd\mid baSdc\mid\varepsilon$

$G_6=(\{S\},\{a,b,c,d\},P_6,S)$，其中 $P_6: S\to aSc\mid aSd\mid bSc\mid bSd\mid\varepsilon$

定理 3.3 正规语言类是可交换上下文无关语言类的一个子集；而可交换上下文无关语言类是上下文无关语言类的一个真子集。

证明：由以上讨论可得该结论。

4 可交换上下文无关语言的性质

文[1~5]给出了语言的运算，并证明了上下文无关语言在运算：连接(·)、选择(+)、星闭包(*)等操作下，对上下文无关语言都是封闭的。显然，由于可交换上下文无关语言类是上下文无关语言类的一个真子集，所以，这些运算对可交换上下文无关语言类也是封闭的，但可交换上下文无关语言有其独特的性质。

定义 4.1 设 Σ 是一个有限字符表，$\alpha_1,\alpha_2\in\Sigma^+$，则：$(\alpha_1$

$+\alpha_2)^{\alpha}=\{w|w\in(\alpha_1+\alpha_2)^*$ 且 α_2 出现次数等于 α_1 出现次数并且 w 的任意前缀语言中 α_2 出现次数小于等于 α_1 出现次数$\}$，将该运算称为 α 闭包运算。

定理 4.1 如果 L_1 和 L_2 是 Σ 上的可交换上下文无关语言，则 $(L_1+L_2)^{\alpha}$ 也是可交换上下文无关语言。

证明：设生成语言 L_1 的可交换上下文无关文法为 $G_1=(V_1,\Sigma_1,P_1,S_1)$，生成语言 L_2 的可交换上下文无关文法为 $G_2=(V_2,\Sigma_2,P_2,S_2)$，由文法 G_1 和 G_2 构造一个文法 $G=(V,\Sigma,P,S)$，使之产生语言 $(L_1+L_2)^{\alpha}$。

构造过程为：令 $\Sigma=\Sigma_1\cup\Sigma_2$；$V=V_1\cup V_2\cup\{S,T\}$；$P=P_1\cup P_2$；

再在 P 中添加产生式：$S\to S_1T|\varepsilon$

$T\to S_1TS_2|S_2TS_1|S_1S_2T|S_2S_1T|S_2$

从而，形成文法 $G=(V,\Sigma,P,S)$，所产生的语言为 $(L_1+L_2)^{\alpha}$。

显然，文法 G 是可交换上下文无关文法，所以，可交换上下文无关语言对运算 α 闭包是封闭的。

定义 4.2 设 Γ 是一个有限字符表，α_1,α_2 是字母表 Γ 上的有限长度的字符串，i、j 为任意非负整数，则：$(\alpha_1+\alpha_2)^{i+j}=\{w|w\in(\alpha_1+\alpha_2)^*$ 且 w 中 α_1 的出现次数为 i，α_2 的出现次数为 j$\}$。

对于上下文无关语言，文[1]给出了其 Pumping 引理，由于可交换上下文无关语言是上下文无关语言的子集，所以，上下文无关语言的 Pumping 引理也适用于可交换上下文无关语言，然而可交换上下文无关语言也有自身的 Pumping 引理。

定理 4.2(可交换上下文无关语言的 Pumping 引理) 设 L 为一可交换上下文无关语言，对 $\forall z\in L$，当 $|z|$ 充分大时，则 z 可写成 $z=uv_1v_2wx_1x_2y$ 的形式，且满足：

(1) $|v_1v_2x_1x_2|\geqslant 1$；

(2)对任意非负整数 i、j 都有：$u(v_1+v_2)^{i+j}w(x_1+x_2)^{i+j}y\in L$

证明：因 $z\in L$，并且 z 可写成 $z=uv_1v_2wx_1x_2y$ 的形式，设产生语言 L 的可交换上下文无关文法为 $G=(V,\Sigma,P,S)$，则 G 中的每一个变量的递归推导都可交换，则对于 $z=uv_1v_2wx_1x_2y$，存在推导：

$$S\overset{*}{\Rightarrow}uAy \tag{8}$$

$$uAy\overset{*}{\Rightarrow}uv_1v_2Ax_1x_2y \tag{9}$$

$$uv_1v_2Ax_1x_2y\overset{*}{\Rightarrow}uv_1v_2wx_1x_2y \tag{10}$$

而变量 A 必是递归可交换推导，且至少存在推导：

$$A\overset{*}{\Rightarrow}v_1v_2Ax_1x_2 \tag{11}$$

$$A\overset{*}{\Rightarrow}v_2v_1Ax_2x_1 \tag{12}$$

$$A\overset{*}{\Rightarrow}v_1Ax_1 \tag{13}$$

$$A\overset{*}{\Rightarrow}v_2Ax_2 \tag{14}$$

$$A\overset{*}{\Rightarrow}w \tag{15}$$

由定义 2.3 知，可分为两种情况来讨论：

第一种情况：若 v_1、v_2 和 x_1、x_2 为不变重复序列，则必有 $|v_1v_2|=0$ 并且 $|x_1x_2|\geqslant 1$，或者，$|v_1v_2|\geqslant 1$ 并且 $|x_1x_2|=0$；显然，对变量 A 只存在：$A\overset{*}{\Rightarrow}Ax_1$、$A\overset{*}{\Rightarrow}Ax_2$、$A\overset{*}{\Rightarrow}w$ 或者 $A\overset{*}{\Rightarrow}v_1A$、$A\overset{*}{\Rightarrow}v_2A$、$A\overset{*}{\Rightarrow}w$，在推导过程中，任意使用 i 次 $A\overset{*}{\Rightarrow}Ax_1$(或 $A\overset{*}{\Rightarrow}v_1A$)及 j 次 $A\overset{*}{\Rightarrow}Ax_2$(或 $A\overset{*}{\Rightarrow}v_2A$)，再使用 1 次 $A\overset{*}{\Rightarrow}w$，可得到：$uw(x_1+x_2)^{i+j}y\in L$(或 $u(v_1+v_2)^{i+j}wy\in L$)；

第二种情况：若 v_1、v_2 和 x_1、x_2 为重复序列，并且 x_1 依赖于 v_1，x_2 依赖于 v_2，此时，必有 $|v_1v_2|\geqslant 1$ 并且 $|x_1x_2|\geqslant 1$。

①若由(8)、(15)两式推导，可得：$S\overset{*}{\Rightarrow}uAy\overset{*}{\Rightarrow}uwy\in L$，而对于表达式：$u(v_1+v_2)^{i+j}w(x_1+x_2)^{i+j}y$，当 i、j 都为 0 时，其值为：uwy，所以，$u(v_1+v_2)^{i+j}w(x_1+x_2)^{i+j}y\in L$；

②若由(8)式，并且重复使用(11)、(12)、(13)、(14)式推导：可得 $S\overset{*}{\Rightarrow}uAy\overset{*}{\Rightarrow}u(v_1+v_2)^{i+j}A(x_1+x_2)^{i+j}y$(其中 i、j 都大于等于 1)，最后再使用(15)式，可推导出：$S\overset{*}{\Rightarrow}u(v_1+v_2)^{i+j}w(x_1+x_2)^{i+j}y\in L$。

由以上讨论，结论得证。

例 1 证明字母表 $\Gamma=\{a,b\}$ 上的回文：$L=\{w\in\{a,b\}^*|w=w^R\}$ 不是可交换上下文无关语言。

证明：在前面已证明，产生该语言的文法不是可交换上下文无关文法，所以回文语言不是可交换上下文无关语言。下面用 Pumping 引理来证明该语言不是可交换上下文无关语言。

假设该语言是可交换上下文无关语言，设 k_1、k_1 是较大的正整数，并让 $z=a^{k1}b^{k2}b^{k2}a^{k1}$，显然，子串 $a^{k1}b^{k2}$ 中的 a、b 是增重复序列，而子串 $b^{k2}a^{k1}$ 中的 a、b 是依赖于子串 $a^{k1}b^{k2}$ 中的 a、b 的减重复序列，由可交换上下文无关语言的 Pumping 引理，将 z 分解为 $a^{k1-1}abb^{k2-1}b^{k2-1}baa^{k1-1}$，并有：$a^{k1-1}(a+b)^{n1+n2}b^{k2-1}b^{k2-1}(b+a)^{n1+n2}a^{k1-1}\in L(n_1,n_2\geqslant 1)$，但是，由于 $(a+b)^{n1+n2}$ 和 $(b+a)^{n1+n2}$ 中的 a、b 的出现次序是随机的，无法保证 $(a+b)^{n1+n2}$ 和 $(b+a)^{n1+n2}$ 的对称性。所以 $a^{k1-1}(a+b)^{n1+n2}b^{k2-1}b^{k2-1}(b+a)^{n1+n2}a^{k1-1}\notin L$，这与假设相矛盾，所以回文语言不是可交换上下文无关语言。

结束语 通过分析上下文无关文法的性质和特点，依据推导过程的变化特点，提出可交换上下文无关文法与可交换上下文无关语言，从传统经典语言中又划分出一个子集，该子集是上下文无关语言的真子集，给出了该类语言的 Pumping 引理。

进一步研究的课题有：

1)该类文法的语法分析的方法；

2)产生该类语言的自动机。

参 考 文 献

1 Hopcroft J E, Ullman J D. Introduction to Automata Theory, Languages and Computation [M]. Addison-Wesleg, 1979

2 Lisovik L P, Koval D A. Language Recognition by Two-Way Determinstic Pushdown Automata [J]. Cybernetics and Systems Analysis, 2004, 40(6): 939～942

3 张继军，吴哲辉. 广义有界上下文无关语言与 Petri 网语言[J]. 系统仿真学报，2005，17(s)：26～29

4 张继军，吴哲辉. 下推自动机的状态转换图与下推自动机的化简[J]. 计算机科学，2006，33(3)：271～274

5 张立昂，等译. 计算理论导引[M]. 机械工业出版社，2000

6 吕映芝. 上下文无关文法与无限状态自动机[J]. 电子学报，1996，24(8)：23～27

7 陈火旺，等著. 程序设计语言编译原理(第三版)[M]. 国防工业出版社，2000

8 蒋立源，康慕宁. 编译原理[M]. 西北工业大学出版社，1999

9 石青云译. 形式语言及其句法分析[M]. 科学出版社，1987

10 肖军模. 程序设计语言编译方法[M]. 大连理工大学出版社，1988

11 杜淑敏，王永宁. 编译程序设计原理[M]. 北京大学出版社，1990