

形式语言与自动机中关于 ϵ 的一些问题

陈文字 王晓斌 程小鸥 孙世新

(电子科技大学计算机科学与工程学院 成都 610054)

摘要 讨论了形式语言与自动机理论中关于空串 ϵ 的一些问题。分析了 ϵ 产生式对文法和语言分类的影响;从文法和有限状态自动机的角度讨论了开始符号 S 和开始状态 q_0 的作用;提出了语言增加或减少 ϵ 句子的简单方法;研究了 ϵ -NFA 的 ϵ 状态转换函数的本质;提出了 ϵ -NFA 转换为 NFA 的新方法,即先将 ϵ -NFA 转换为文法形式,消除 ϵ 产生式和单产生式后得到正则文法,再将正则文法转换为 NFA。并用实际例子进行了验证。

关键词 ϵ 句子, ϵ 产生式, ϵ 状态转换函数,带 ϵ 动作的有限状态自动机

中图分类号 TP301.2 文献标识码 A

Issues Regarding ϵ in Formal Language and Automata Theory

CHEN Wen-yu WANG Xiao-bin CHENG Xiao-ou SUN Shi-xin

(School of Computer Science and Engineering, University of Electronic Science & Technology of China, Chengdu 610054, China)

Abstract The paper discussed some issues regarding blank string ϵ in the formal language and automata theory. After analysis of the influence of production ϵ on grammar and language classification, the paper discussed the effect of starting symbol S and the starting state q_0 from the perspective of grammar and infinite state and proposed a simple method to increase language or decrease sentence ϵ . The paper also proposed a new method to transit ϵ -NFA to NFA after studying the essence of ϵ state transition function of ϵ -NFA. The method is: first transit ϵ -NFA to formal grammar and eliminate production ϵ and single production. After that, regular grammar was obtained. Then transited regular grammar to NFA. Examples were given to support the discussion.

Keywords ϵ -sentence, ϵ -producer, ϵ -state transform function, ϵ -NFA

文法 G 的开始符号可以作为一般的非终结符使用,有限状态自动机(FA)的开始状态也可以作为一般状态使用, G 产生的语言和 FA 接收的语言要增加或减少空句子是困难的。本文通过限制开始符号和开始状态的作用,提出了语言增加或减少空句子的简单方法。通过研究 ϵ -NFA 的 ϵ 状态转换函数的本质,提出了 ϵ -NFA 转换正则文法和 NFA 的简便方法。

1 ϵ 产生式对文法和语言分类的影响

ϵ 不属于任何字母表,但可以属于语言。文法要产生 ϵ 句子,最简单的方法是提供特殊的 ϵ 产生式:

$$S \rightarrow \epsilon$$

文法 G 产生的语言如果不包括 ϵ 句子,则文法 G 可以没有任何的 ϵ 产生式^[1,2]。

如果文法的开始符号 S 不出现在文法的任意产生式的右边,那么 S 不能够作为一般非终结符使用, S 仅仅负责推导过程的开始^[1-3]。在句子的推导过程中,所有出现的句型都不包括开始符号 S ,即

$$S \Rightarrow^+ \alpha A \beta$$

其中, $\alpha, \beta \in \Sigma, A \neq S$ 。

增加 ϵ 产生式 $S \rightarrow \epsilon$,则该产生式只使用一次,只能产生 ϵ 句子。

如果文法的开始符号 S 出现在文法的某个产生式的右边,增加的 ϵ 产生式可以产生多余的句子。

对于任意文法 $G = (\Sigma, V, S, P)$,如果 S 不出现在文法 G 的任何产生式的右边,且 G 是 1、2 或 3 型文法,则

$$G' = (\Sigma, V, S, P \cup \{S \rightarrow \epsilon\})$$

$$G'' = (\Sigma, V, S, P - \{S \rightarrow \epsilon\})$$

仍然为 1、2 或 3 型文法,对应产生的语言仍然为 1、2 或 3 型语言。

2 FA 接收 ϵ 句子

FA 接收 ϵ 句子,最简单的方法是设置开始状态为接收状态^[4,5]。

设 FA 接收语言 L ,设置语言 $L' = L \cup \{\epsilon\}$ 或 $L'' = L - \{\epsilon\}$,可以方便地构造新的 FA 接收 L' 和 L'' 。

2.1 DFA 接收 ϵ 句子

设 DFA = $(Q, \Sigma, \delta, q_0, F)$, $L = L(DFA)$,直接将 DFA 的

到稿日期:2009-02-20 返修日期:2009-05-10 本文受 863 计划(2006AA01Z174)资助。

陈文字(1968-),男,副教授,CCF 高级会员,研究方向为模式识别、编译技术、面向对象技术,E-mail: cwy@uestc.edu.cn;王晓斌(1964-),男,副教授,研究方向为编译技术、人工智能;程小鸥(1986-),女,硕士研究生,研究方向为编译技术;孙世新(1936-),男,教授,博士生导师,研究方向为分布式算法、模式识别、计算数学。

开始状态也设置为接收状态,则可以接收 ϵ 句子,如果 DFA 存在 $\delta(q, a) = q_0$ 形式的产生式,就可以接收不属于 L 的句子,例如图 1 所示的 DFA_1 。

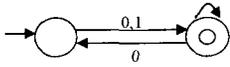


图 1 DFA_1 的状态图

开始状态 q_0 不属于接收状态,且其他状态可以回到开始状态 q_0 。如果直接将 DFA 的开始状态也设置为接收状态,则得到如图 2 所示的 DFA_2 ,可以接收 ϵ 句子,还可以接收 00 等多余句子。

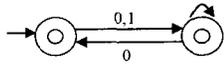


图 2 DFA_1 的状态图

对 DFA 进行改进,增加开始状态 q_s ,使得开始状态只负责接收整个输入串的第一个符号(对比文法,类型与文法的开始符号 S 不出现在文法的任何产生式的右边),即开始状态的入度为 0,不允许任何状态再改变为开始状态,那么开始状态就不能够作为一般状态使用。

任意 $a \in \Sigma$,对于 DFA 的 $\delta(q_0, a) = q'$,增加 $\delta(q_s, a) = q'$ 。改造 DFA_1 为图 3 所示的 DFA_3 。

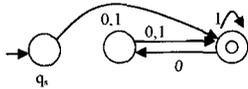


图 3 DFA_3 的状态图

若直接将开始状态设置为接收状态,则仅仅多接受 ϵ 句子,即改造 DFA_1 为图 4 所示的 DFA_4 。

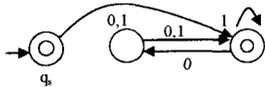


图 4 DFA_4 的状态图

2.2 NFA 接收 ϵ 句子

NFA 的情况类似 DFA,只是在改造 NFA 时,只针对 NFA 存在的 $\delta(q_0, a) = Q'$ 增加 $\delta(q_s, a) = Q'$ 即可。

DFA 和 NFA 接收的语言需要除去 ϵ 句子,则只需要将开始状态设置为非接收状态即可。

3 ϵ -NFA 转换为 NFA

ϵ -NFA 可以转换为 NFA' 的传统方法是扩充 NFA 的状态转换函数,使得

$$\delta'(q, a) = \delta * (q, a)$$

可以将 ϵ -NFA 先转换为扩展的正则文法,再转换为 NFA。

对于 ϵ -NFA $= (Q, \Sigma, \delta, q_0, F)$,构造等价的文法

$$G = (\Sigma, Q, q_0, P)$$

其中, $P = \{q \rightarrow aq' \mid q' \in \delta(q, a)\} \cup \{q \rightarrow q' \mid q' \in \delta(q, \epsilon)\} \cup \{q \rightarrow a \mid \delta(q, a) \cap F \neq \Phi\} \cup \{q \rightarrow \epsilon \mid \delta(q, \epsilon) \cap F \neq \Phi\}$ 。该文法不属于正则文法,主要是增加了单产生式和 ϵ 产生式^[9]。

对于单产生式的推导作用,推导过程 $S \Rightarrow^* aA \Rightarrow^* aB \Rightarrow^* \alpha w_1 \Rightarrow^* \dots$ 可以简化为 $S \Rightarrow^* aA \Rightarrow^* \alpha w_1 \beta \Rightarrow^* \dots$,即省略掉 $A \Rightarrow B$ 的推导。

因此,需要将 $A \rightarrow B$ 的产生式使用下面的产生式进行替换:

$$A \rightarrow w_1 \mid w_2 \mid w_3 \mid \dots \mid w_n$$

而 $B \rightarrow w_1 \mid w_2 \mid w_3 \mid \dots \mid w_n$ 是文法 G 中关于 B 的所有产生式。

实际上,就是在产生式中体现出推导过程。

需要保留关于 B 的所有产生式,因为可能存在 $S \Rightarrow^+ aB \Rightarrow^+ \alpha w C \Rightarrow^+ \dots$ 的推导,即该推导过程没有使用产生式 $A \rightarrow B$ 。

对于 $A \rightarrow \epsilon$ 产生式的推导作用,推导过程可能是 $S \Rightarrow^* aA \Rightarrow^* \alpha w B \Rightarrow^* \dots$,即不使用 ϵ 产生式;或 $S \Rightarrow^* aA \Rightarrow^* a \Rightarrow^* \dots$,即使用 ϵ 产生式。对于 $C \in V$,考虑产生式 $C \rightarrow wA$,增加产生式 $C \rightarrow w$,则能够体现出 A 的 ϵ 产生式的作用,可以删除 A 的 ϵ 产生式。

如果 $\epsilon \in L(G)$,则要增加新的开始符号 S' 和产生式 $S' \rightarrow \epsilon$ 。对于 $S \rightarrow r$,增加 $S' \rightarrow r$,使得新文法是正则文法。

正则文法 $G = (\Sigma, V, S, P)$ 将文法的非终结符当作 NFA 的状态,并增加一个接收状态 q (若文法 G 中有 $S \rightarrow \epsilon$,即 $\epsilon \in L$,则开始状态 S 也是接收状态),使得

$$NFA = (Q, \Sigma, \delta, Q_0, F)$$

其中, $Q = V \cup \{q\}$, $Q_0 = \{S\}$, $F = \{q\}$, $\delta(A, x) = \{B \mid B \in V, \text{且 } A \rightarrow xB \text{ 在 } P \text{ 中}\} \cup \{q \mid A \rightarrow x \text{ 在 } P \text{ 中}\}$,则 $L(G) = L(NFA)$ 。

例如图 5 所示的 ϵ -NFA,接收语言 $0^* 1^* 2^*$ 转换为文法 $S \rightarrow 0S \mid A, A \rightarrow 1A \mid B, B \rightarrow 2B \mid 2 \mid \epsilon$ 。消除单产生式后,得

$$S \rightarrow 0S \mid 1A \mid 2B \mid 2 \mid \epsilon$$

$$A \rightarrow 1A \mid 2B \mid 2 \mid \epsilon$$

$$B \rightarrow 2B \mid 2 \mid \epsilon$$

消除 ϵ 产生式后,得

$$S \rightarrow 0S \mid 0 \mid 1A \mid 1 \mid 1 \mid 2B \mid 2 \mid \epsilon$$

$$A \rightarrow 1A \mid 1 \mid 1 \mid 2B \mid 2$$

$$B \rightarrow 2B \mid 2$$

增加 S' 和相应产生式后得正则文法

$$S' \rightarrow \epsilon \mid 0S \mid 0 \mid 1A \mid 1 \mid 1 \mid 2B \mid 2$$

$$S \rightarrow 0S \mid 0 \mid 1A \mid 1 \mid 1 \mid 2B \mid 2$$

$$A \rightarrow 1A \mid 1 \mid 1 \mid 2B \mid 2$$

$$B \rightarrow 2B \mid 2$$

转换为如图 6 所示的 NFA。

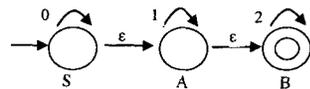


图 5 接收 $0^* 1^* 2^*$ 的 ϵ -NFA

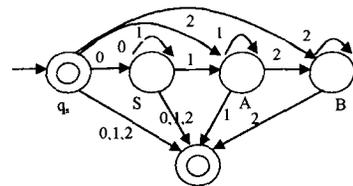


图 6 接收 $0^* 1^* 2^*$ 的 NFA

结束语 形式语言与自动机理论中,文法的基本分类原则是只允许 0 型文法中出现 ϵ 产生式,不允许 1,2 和 3 型文法中出现 ϵ 产生式。本文通过限制开始符号 S 的作用,使得 1,2 和 3 型文法中可以出现关于 S 的 ϵ 产生式,同时提出了

(下转第 264 页)

不能直接应用。因此,为了更为清晰地阐释不同层级间基本模型的交互映射关系,可以在高层级的基本模型和其低层级的基本模型之间增加一个逻辑中间层级,而这个逻辑中间层级又恰好将该层级基本模型的输入及其低层级基本模型间的输出有机地衔接起来,从而保证了层级映射能够在这两个层级之间进行。这种逻辑中间层之所以被称为“软构层”,是因为它不是 C2 组织的物理结构组成部分,因而并不构成一个层级;但是它却又具有在 C2 组织探索性分析模型的层级映射中存在的必要性,因为通过它能更好地解释探索性分析模型的各个分辨率层级间的层级映射关系。

如图 6 所示,假设某基本模型所在层级为 n ,那么增加软构层后,软构层模型所在层级应该为 $n+1$,而其低层级基本模型的层次为 $n+2$ 。由图可知,基本模型和组成该基本模型的低层级基本模型间由于没有直接的输入输出关联,因而不能建立直接的层级映射关系。但是,增加软构层之后,基本模型和组成该基本模型的低层级基本模型之间的层级映射关系可以通过软构层的两次层级映射来完成。

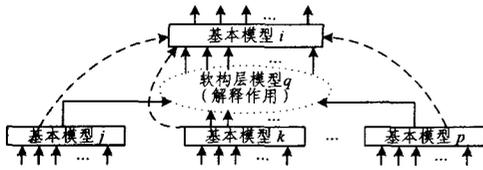


图 6 结构层级概念模型中的软构层映射

由于基本模型 i 分析处理的输入直接来自软构层模型 q 分析处理的输出,所以基本模型和相应的软构层模型间可以进行层级映射,即

$$X_{n,i} \subseteq Y_{n+1,q} \Rightarrow Y_{n,i} = F_{n,i}(X_{n,i})$$

$$F_{n,i}(X_{n,i}) = F_{n,i}(Y_{n+1,q}) = F_{n,i}(F_{n+1,q}^i(X_{n+1,q})) = G_{n,i}(X_{n+1,q}^i)$$

其中,

$$Y_{n+1,q}^i \subseteq Y_{n+1,q}$$

$$X_{n+1,q}^i \subseteq X_{n+1,q}$$

$$F_{n+1,q}^i \subseteq F_{n+1,q}$$

同理可得软构层模型和相应的低层级系统层节点间的层级映射,即

$$X_{n+1,q} \subseteq (Y_{n+2,j} \cup Y_{n+2,k} \cup \dots \cup Y_{n+2,p}) \Rightarrow Y_{n+1,q} = F_{n+1,q}$$

$$(X_{n+1,q}) = F_{n+1,q}(Y_{n+2}^q) = F_{n+1,q}(F_{n+2}^q(X_{n+2}^q))$$

(上接第 244 页)

文法产生 ϵ 句子的简单方法。相应地,限制了 FA 中开始状态的作用,使得开始状态只接收输入串的第 1 个字母。提出了 ϵ -NFA 转换为正则文法和 NFA 的方法,并通过实例进行了验证。

参考文献

- [1] 蒋宗礼. 形式语言与自动机理论(第 2 版)[M]. 北京:清华大学出版社,2008
- [2] 陈有祺. 形式语言与自动机[M]. 南京:南开大学出版社,1999
- [3] Sipser M. Introduction to the theory of computation[M]. Boston, USA: PWS Publishing Company, 1997

$$= G_{n+1,q}(X_{n+2}^q)$$

其中,

$$Y_{n+2}^q \subseteq (Y_{n+2,j} \cup Y_{n+2,k} \cup \dots \cup Y_{n+2,p})$$

$$X_{n+2}^q \subseteq (X_{n+2,j} \cup X_{n+2,k} \cup \dots \cup X_{n+2,p})$$

$$F_{n+2}^q \subseteq (F_{n+2,j} \cup F_{n+2,k} \cup \dots \cup F_{n+2,p})$$

结合以上两种层级映射,基本模型和组成该基本模型的低层级基本模型之间的层级映射关系可表示为

$$Y_{n,i} = F_{n,i}(F_{n+1}^i(X_{n+1}^i)) = F_{n,i}(F_{n+1}^i(Y_{n+2}^q))$$

$$= F_{n,i}(F_{n+1}^i(F_{n+2}^q(X_{n+2}^q))) = \Psi_{n,i}(X_{n+2}^q)$$

其中,

$$Y_{n+2}^q \subseteq (Y_{n+2,j} \cup Y_{n+2,k} \cup \dots \cup Y_{n+2,p})$$

$$X_{n+2}^q \subseteq (X_{n+2,j} \cup X_{n+2,k} \cup \dots \cup X_{n+2,p})$$

$$F_{n+2}^q \subseteq (F_{n+2,j} \cup F_{n+2,k} \cup \dots \cup F_{n+2,p})$$

综上所述,C2 组织的每个高层级的基本模型都可以相应地添加一个软构层模型,与低层级的基本模型间建立关联映射,软构层模型的数量代表了整个 C2 组织中所包含的各层级基本模型的数量。软构层的逻辑作用使得层级映射可以在 C2 组织的不同分辨率模型的各个层级之间依次进行下去,从而确保了整个 C2 组织探索性分析模型的层级映射的完备性。

结束语 构建 C2 组织探索性分析模型是实现 C2 组织优化设计理论探索性分析验证的基础和关键。本文围绕 C2 组织探索性分析模型的构建问题,阐述了分析模型的特征,明确了分析模型的描述内容,提出了分析模型的构建原则与基本方法,并对分析模型的层级映射关系进行了分析。在此基础上设计与构建 C2 组织探索性分析模型,进一步实现了探索性分析环境,对 C2 组织优化设计理论和方法进行了实验验证和分析。

参考文献

- [1] 周少平,李群,王维平. 探索性分析建模[J]. 计算机仿真,2006(4):50-53
- [2] Davis P K, Bigelow J H. Experiment in Multi-resolution Modeling[R]. MR-1004- DARPA. 1998
- [3] 耿振余,毕义明. 基于探索性分析法仿真的建模方法研究[J]. 情报指挥控制系统与仿真技术,2005(4):31-34
- [4] 周启海. NFA \rightarrow FA \rightarrow GFA 的自动转换算法[J]. 电子科技大学学报,2005,34(3):234-236
- [5] 周启海. 有穷自动机的规范化[J]. 计算机科学,2004(增刊):156-157
- [6] Hopcroft J E, Ullman J D. Introduction to Automata Theory, Languages and Computation[M]. Addison-Wesley, 1979: 112-134
- [7] 张选芳. 形式语言中的上下文无关和上下文相关[J]. 电子科技大学学报,1997,26(4):435-439
- [8] 蒋龙龙,陈文字. 利用等价类构造有限状态自动机[J]. 计算机科学,2006,33(11):272-275