Earley, J. (1970) An Efficient Context-Free-Parsing Algorithm, Communication of ACM, Vol. 6, No. 8, pp94-102.

Earley 分析算法

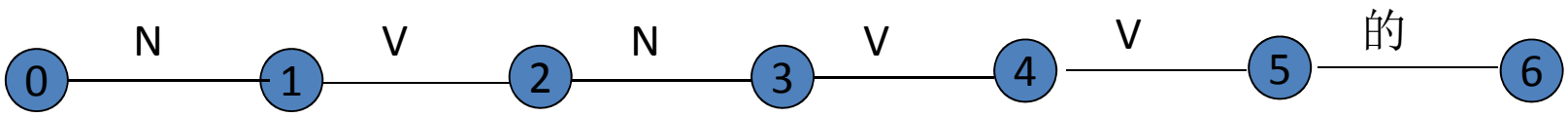
詹卫东
http://ccl.pku.edu.cn/doubtfire

Earley算法

Top-down（为主）与 bottom-up（为辅）相结合

预测能力 ＋ 数据驱动

张三是县长派来的　一　N　V　N　V　V　的



基本概念：状态(state)

一个状态由3部分组成：

（1）上下文无关文法规则

（2）圆点・ （圆点左边的部分是已分析的，右边是待分析的）

（3）状态的起止位置：

整数 i : 状态起点（已分析子串的起点）

整数 j : 状态终点（已分析子串的终点） $i \le j$

比如： <S → NP・VP [0, 4]>

点规则

点在最右端，为 [完成状态]

否则，为 [未完成状态]

基本操作/算子（operator）

1. 预测（Predicator）：如果圆点右方是一个非终结符，那么以该非终结符为左部的规则都有匹配的希望，也就是说分析器可以预测这些规则都可以建立相应的项目。

2. 扫描（Scanner）：如果圆点右方是一个终结符，就将圆点向右方扫描一个字符间隔，把匹配完的字符“让”到左方。

3. 归约（Completer）：如果圆点右方没有符号（即圆点已经在状态的结束位置），那么表示当前状态所做的预测已经实现，因而可以将当前状态（S_i）与已有的包含当前状态的状态(S_j)进行归约（合并），从而扩大S_j覆盖的子串范围

算子的形式定义

Predicator: 对于状态 $Z \rightarrow \alpha \cdot X\ \beta\ [j, k]$ 其中X是非终结符
对于语法中每条形如 $X \rightarrow \gamma$ 的规则，都可以形成一个新状态： $X \rightarrow \cdot \gamma\ \ [k, k]$

Scanner: 对于状态$Z \rightarrow \alpha \cdot X\ \beta\ [j, k]$ 其中X是终结符
如果X与输入字符串中第k个字符匹配，就形成一个新状态：$Z \rightarrow \alpha X \cdot \beta\ [j, k+1]$

Completer : 对于一个已经“完成”的状态 $Z \rightarrow \gamma \cdot\ [j, k]$
如果已有状态集合中有形如 $X \rightarrow \alpha \cdot Z\ \beta\ [i, j]$这样的状态，就形成一个新状态：$X \rightarrow \alpha Z \cdot \beta\ [i, k]$

说明：以上α，β，γ是终结符或非终结符串，其中α，β均可为空字符 $i \le j \le k$

Earley算法：算法描述

设输入字符串长度为n, 字符间隔可记做0,1,2,…,n

（1）将文法规则中形如 S→ α 的规则形成为状态：

<S→ · α [0, 0]> 加入到状态集合中（种子状态/seed state）

（2）对当前分析句子的每个词，依次进行循环：

对状态集中的每个状态，依次进行循环：

i）如果当前状态是[未完成状态]，且点后不是终结符，则

执行Predicator；

ii）如果当前状态是[未完成状态]，且点后是终结符，则

执行Scanner；

iii）如果当前状态是[完成状态]，则

执行Completer；

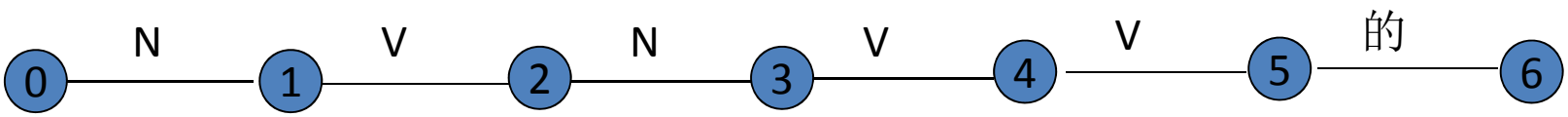
（3）如果最后得到形如<S→ α · [0, n]>这样的状态，那么

输入字符串被接受为合法的句子，否则分析失败

Earley算法过程示例

张三是县长派来的
老虎是瞎子打死的
主意是董永想出来的
……
N V N V V 的

(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V



Earley算法过程示例-1

(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V

当前间隔 0

预测

<S→ · NP VP [0,0]>
<NP→ · N [0,0]>
<NP→· CS 的 [0,0]>
<CS→· NP V' [0,0]>

当前间隔 1

<NP→ N · [0,1]> → 扫描
<S→ NP · VP [0,1]> → 归约
<CS→NP · V' [0,1]> → 归约
<VP→· V NP [1,1]> → 预测
<V' → · V V [1,1]> → 预测

预测 → 扫描 → 归约 → 预测 → 扫描 → 归约 → 预测 .....

Earley算法过程示例-2

(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V

当前间隔 1

<NP→ N · [0,1]>
<S→ NP · VP [0,1]>
<CS→NP · V' [0,1]>
<VP→· V NP [1,1]>
<V' → · V V [1,1]>

当前间隔 2

<VP→ V · NP [1,2]>
<V' → V · V [1,2]>
<NP→ · N [2,2]>
<NP→· CS 的 [2,2]>
<CS → · NP V' [2,2]>

扫描

预测

<S→ NP · VP [0,1]>
<CS→NP · V' [0,1]>

保留状态

Earley算法过程示例-3

(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V

当前间隔 2

<VP→ V · NP [1,2]>
<V' → V · V [1,2]>
<NP→ · N [2,2]>
<NP→· CS 的 [2,2]>
<CS → · NP V' [2,2]>

<S→ NP · VP [0,1]>
<CS→NP · V' [0,1]>

当前间隔 3

<NP→ N · [2,3]> → 扫描
<VP → V NP · [1,3]>
<S→ NP VP · [0,3]> → 归约
<CS→NP · V' [2,3]>
<V' → · V V [3,3]> → 预测

<S→ NP · VP [0,1]>
<CS→NP · V' [0,1]>
<NP→· CS 的 [2,2]>
<VP→ V · NP [1,2]>

Earley算法过程示例-4

(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V

当前间隔 3

<NP→ N · [2,3]>
<VP → V NP · [1,3]>
<S→ NP VP · [0,3]>
<CS→NP · V' [2,3]>
<V' → · V V [3,3]>

<S→ NP · VP [0,1]>
<CS→NP · V' [0,1]>
<NP→· CS 的 [2,2]>
<VP→ V · NP [1,2]>

当前间隔 4

<V' → V · V [3,4]> → 扫描

<S→ NP · VP [0,1]>
<CS→NP · V' [0,1]>
<NP→· CS 的 [2,2]>
<VP→ V · NP [1,2]>
<CS→NP · V' [2,3]>
<S→ NP VP · [0,3]>

Earley算法过程示例-5

(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V

当前间隔 4

<V' → V · V [3,4]>

<S→ NP · VP [0,1]>
<CS→NP · V' [0,1]>
<NP→· CS 的 [2,2]>
<VP→ V · NP [1,2]>
<CS→NP · V' [2,3]>
<S→ NP VP · [0,3]>

当前间隔 5

<V' →V V · [3,5]> → 扫描
<CS→NP V' · [2,5]> → 归约
<NP→CS · 的 [2,5]> → 归约

<S→ NP · VP [0,1]>
<CS→NP · V' [0,1]>
<VP→ V · NP [1,2]>
<S→ NP VP · [0,3]>

Earley算法过程示例-6

(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V

当前间隔 5

<V' → V V · [3,5]>
<CS→NP V' · [2,5]>
<NP→CS · 的 [2,5]>

<S→ NP · VP [0,1]>
<CS→NP · V' [0,1]>
<VP→V · NP [1,2]>
<S→ NP VP · [0,3]>

当前间隔 6

<NP →CS 的 · [2,6]> → 扫描
<VP→V NP · [1,6]> → 归约
<S→NP VP · [0,6]> → 归约

<CS→NP · V' [0,1]>
<S→ NP VP · [0,3]>

6							
5							
4							
3							
2							
1							
0							
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6							
5							
4							
3							
2							
1							
0	S → · NP VP						种子
	0	1	2	3	4	5	6

N 张三　V 是　N 县长　V 派　V 来　的 的

6							
5							
4							
3							
2							
1							
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6							
5							
4							
3							
2							
1	NP → N ·						扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6							
5							
4							
3							
2							
1	S → NP · VP CS → NP · V ' NP → N ·						归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6							
5							
4							
3							
2							
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6							
5							
4							
3							
2		VP → V · NP V ' → V · V					扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6							
5							
4							
3							
2		VP → V · NP V ' → V · V	NP → · N NP → · CS 的				预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6							
5							
4							
3							
2		VP → V · NP V ' → V · V	CS → · NP V ' NP → · N NP → · CS 的				预测 预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6

N 张三　V 是　N 县长　V 派　V 来　的 的

6							
5							
4							
3			NP → N ·				扫描
2		VP → V · NP V ' → V · V	CS → · NP V ' NP → · N NP → · CS 的				预测 预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6							
5							
4							
3			CS → NP · V ' NP → N ·				归约 扫描
2		VP → V · NP V ' → V · V	CS → · NP V ' NP → · N NP → · CS 的				预测 预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6							
5							
4							
3		VP → V NP ·	CS → NP · V ' NP → N ·				归约 扫描
2		VP → V · NP V ' → V · V	CS → · NP V ' NP → · N NP → · CS 的				预测 预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6							
5							
4							
3	S → NP VP ·	VP → V NP ·	CS → NP · V ' NP → N ·				归约 扫描
2		VP → V · NP V ' → V · V	CS → · NP V ' NP → · N NP → · CS 的				预测 预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6							
5							
4							
3	S → NP VP ·	VP → V NP ·	CS → NP · V ' NP → N ·	V ' → · V V			预测 归约 扫描
2		VP → V · NP V ' → V · V	CS → · NP V ' NP → · N NP → · CS 的				预测 预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6							
5							
4				V ' → V · V			扫描
3	S → NP VP ·	VP → V NP ·	CS → NP · V ' NP → N ·	V ' → · V V			预测 归约 扫描
2		VP → V · NP V ' → V · V	CS → · NP V ' NP → · N NP → · CS 的				预测 预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6							
5				V ' → V V ·			扫描
4				V ' → V · V			扫描
3	S → NP VP ·	VP → V NP ·	CS → NP · V ' NP → N ·	V ' → · V V			预测 归约 扫描
2		VP → V · NP V ' → V · V	CS → · NP V ' NP → · N NP → · CS 的				预测 预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6

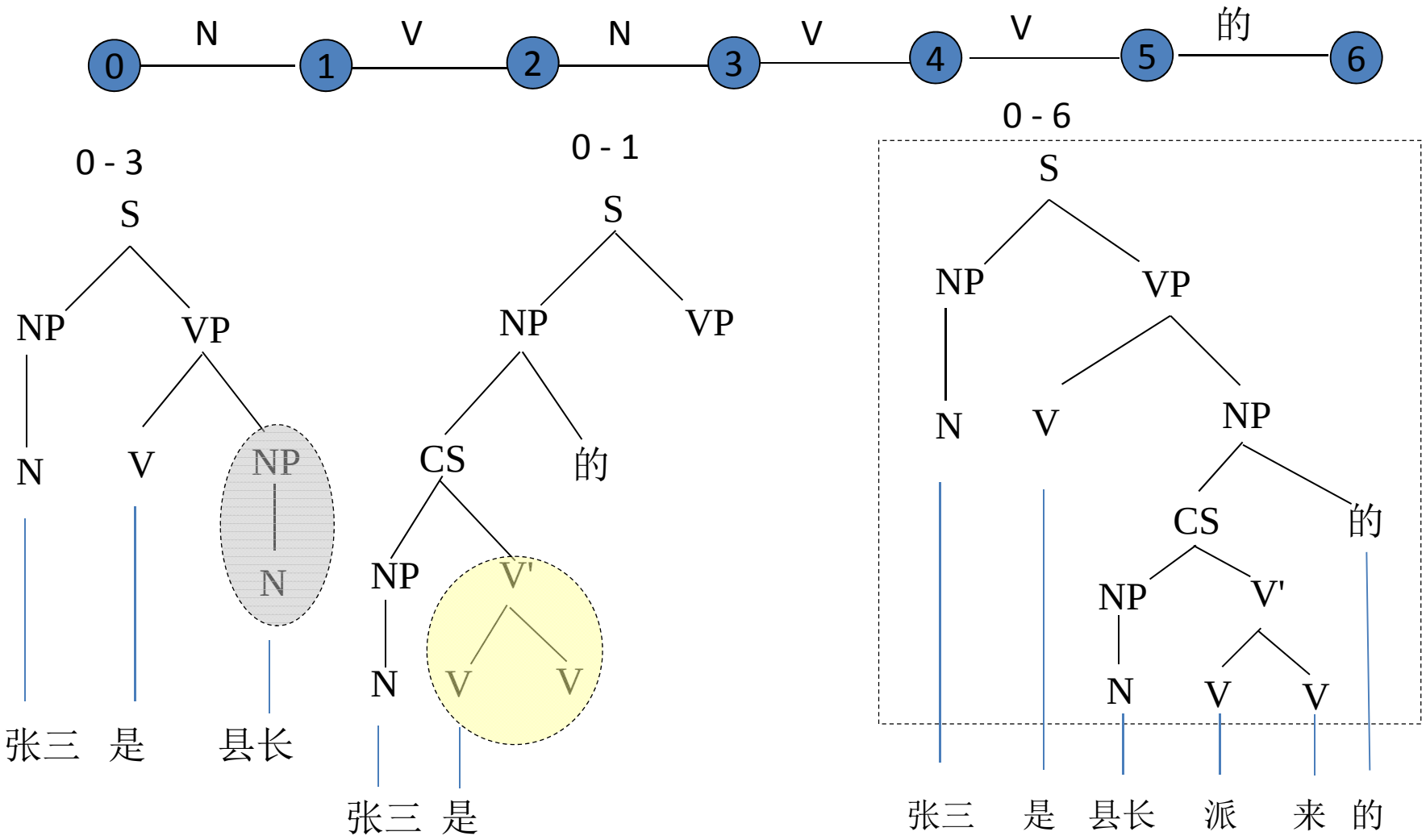
N 张三　V 是　N 县长　V 派　V 来　的 的

6							
5			NP → CS · 的 CS → NP V ' ·	V ' → V V ·			归约 扫描
4				V ' → V · V			扫描
3	S → NP VP ·	VP → V NP ·	CS → NP · V ' NP → N ·	V ' → · V V			预测 归约 扫描
2		VP → V · NP V ' → V · V	CS → · NP V ' NP → · N NP → · CS 的				预测 预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6			NP → CS 的 ·				扫描
5			NP → CS · 的 CS → NP V ' ·	V ' → V V ·			归约 扫描
4				V ' → V · V			扫描
3	S → NP VP ·	VP → V NP ·	CS → NP · V ' NP → N ·	V ' → · V V			预测 归约 扫描
2		VP → V · NP V ' → V · V	CS → · NP V ' NP → · N NP → · CS 的				预测 预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6		VP → V NP ·	NP → CS 的 ·				归约 扫描
5			NP → CS · 的 CS → NP V ' ·	V ' → V V ·			归约 扫描
4				V ' → V · V			扫描
3	S → NP VP ·	VP → V NP ·	CS → NP · V ' NP → N ·	V ' → · V V			预测 归约 扫描
2		VP → V · NP V ' → V · V	CS → · NP V ' NP → · N NP → · CS 的				预测 预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6	S → NP VP ·	VP → V NP ·	NP → CS 的 ·				归约 扫描
5			NP → CS · 的 CS → NP V ' ·	V ' → V V ·			归约 扫描
4				V ' → V · V			扫描
3	S → NP VP ·	VP → V NP ·	CS → NP · V ' NP → N ·	V ' → · V V			预测 归约 扫描
2		VP → V · NP V ' → V · V	CS → · NP V ' NP → · N NP → · CS 的				预测 预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

6	S → NP VP ·	VP → V NP ·	NP → CS 的 ·				归约 扫描
5			NP → CS · 的 CS → NP V ' ·	V ' → V V ·			归约 扫描
4				V ' → V · V			扫描
3	S → NP VP ·	VP → V NP ·	CS → NP · V ' NP → N ·	V ' → · V V			预测 归约 扫描
2		VP → V · NP V ' → V · V	CS → · NP V ' NP → · N NP → · CS 的				预测 预测 扫描
1	S → NP · VP CS → NP · V ' NP → N ·	VP → · V NP V ' → · V V					预测 归约 扫描
0	CS → · NP V ' NP → · N NP → · CS 的 S → · NP VP						预测 预测 预测 种子
	0	1	2	3	4	5	6
	N 张三	V 是	N 县长	V 派	V 来	的 的	

Earley算法构造分析树示意图