

Assignment1. Chart Parsing文档报告

MG1733092 张则君

2017年11月7日

1 任务描述

本实验是实现一个基于简单英语语法的chart句法分析器。首先用户可以通过指定文法，对输入的任意英语句子，运用一种基于图的句法分析技术（ChartParsing）判断用户给定的句子是否符合句法。

2 技术路线

1. 实验创建4个txt，其中包含文法产生式，要分析的英语句子，英语句子的词性标注和要输出的数据（输出为句子是否满足文法、活动边和非活动边）。
2. 实验创建3个类，用于表示agenda，非活动边，活动边。
 - agenda:记录每个文法元素、起始位置和终止位置
 - 活动边：记录所用产生式、起始位置、终止位置和下一个要判断的文法元素的位置。
 - 非活动边：记录当前边的文法元素、起始位置和终止位置。
3. 实验通过graphviz包去创建图。节点表示每个词所在的位置（比实际多一个），虚线边表示活动边，实线边边说非活动边。
4. ChartParsing算法:

当agenda为空并且输入中没有下一个词时算法终止。

 - (1) 若agenda为空，则把句子中下一个词的各种词法符号（词性）和它们的位置加入进来。
 - (2) 从agenda中取一个元素（设为C，位置为：p1-p2）对下面形式的每个规则增加活动边：
 - $X \rightarrow CX_1 \dots X_n$ ，增加一条活动边： $X \rightarrow C \circ X_1 \dots X_n$ ，位置为：p1-p2；
 - $X \rightarrow C$ ，把X加入agenda，位置为：p1-p2。
 - (3) 将C作为非活动边加入到chart的位置p1-p2。
 - (4) 对每个形式为： $X \rightarrow X_1 \dots \circ C \dots X_n$ 的活动边：

- 若它在 p_0 - p_1 之间，则增加一条活动边： $X \rightarrow X_1 \dots C \circ \dots X_n$ ，位置： p_0 - p_2 ；
- 对每个形式为： $X \rightarrow X_1 \dots X_n \circ C$ 的活动边，若它在 p_0 - p_1 之间，则把 X 加入agenda，位置为： p_0 - p_2 。

3 实验结果

实验以文本和图片的方式输出，其中以“The cat caught a mouse”为例子，输出的文本以及图片结果展示如下：

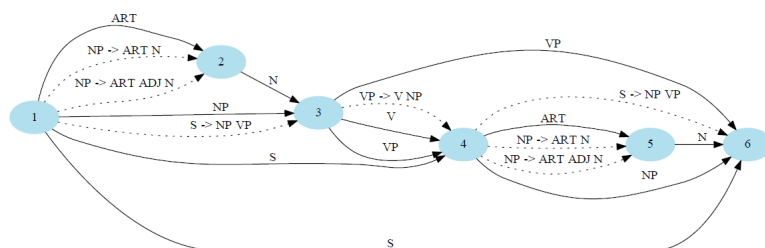


图 1: chartparsing图

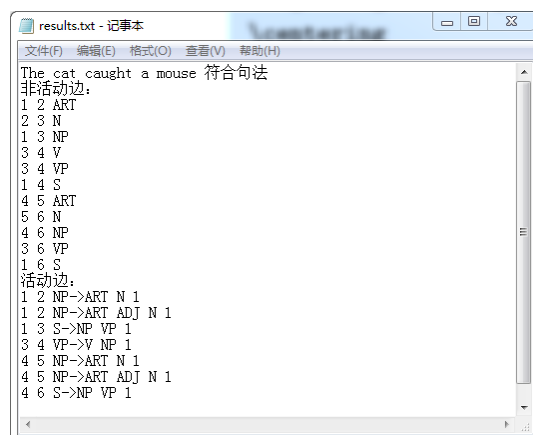


图 2: result信息

4 代码说明

实验通过Python实现，在windows7下使用python3.5版本下实现，需要用到numpy、graphviz包。

运行时，可直接在命令行下执行 `python parse.py` 运行，实验输出result.txt和chart-parsing.gv.pdf。

用户可以自己制定文法在rules.txt，英语句子写入到sentence.txt，对应词性标注在cixing.txt。