

【读后感】这篇文章主要工作是从已标注的语料中根据层次聚类算法得出事件模式，也就是说对事件语义标签进行了一个聚类。

从实验结果来看：

- (1) 初始事件模式，列出了 15 个事件模式的数量分布；
- (2) 合并事件模式，列出了 20 个事件模式的数量分布；
- (3) 聚类之后事件模式，列出了 20 个事件模式的数量分布；

不是很清楚最终聚类之后包含多少类事件模式，而且这些事件模式有什么用途也没有说清楚。我感觉这个的用途就是：把一个句子中的事件语义都标注好，然后根据聚类结果对这个句子的事件语义结构进行一个分类。

不过，这个方法中也有值得借鉴的地方。

通过那天与您讨论，感觉可以做以下工作（不是很完善一个初步设想）：

(1) 可以考虑引入词性标注、命名实体识别、依存句法分析等技术，通过一些方法，找到词法/句法标签和事件模式之间的关系，这样只要对生语料进行 POS、NER、Parser 等处理，就可以大体确定这个句子中含有哪种模式的事件（这个可以包含事件识别和事件分类两个部分）；

(2) 还可以通过词性标注、命名实体识别、依存句法分析等技术，通过一些方法，找到词法/句法标签和事件语义角色之间的关系，然后利用事件模式、标签等特征，进行事件语义角色的识别；

(3) 但是上述过程中具体采用什么方法，还是没想到，等我看李妍师姐的硕士论文，看一下是不是可以尝试用 SVM，或者您给一些建议，那种模型/算法比较合适/可行。。。

---

【题目】基于聚类分析的事件语义模式获取

【作者】季陶美，刘茂福，张璐，杨晓

【单位】武汉科技大学，计算机科学与技术学院

【期刊】微型机与应用

【时间】2013 年 2 期

---

【笔记】

【摘要】

选取新闻类型的 web 页面，遵照事件语义标注规范对选取的未标注 web 页面进行事件语义标注。

对标注了事件语义的语料实例进行抽象得到事件语义结构模式；

利用层次聚类算法，将所得的事件语义结构模式进行聚类分析，得到不同类别的事件语义模式。

【事件语义结构】是语法和语义层面的结合。它充分考虑的事件的时间结构特性和内部构成关系对谓词句法表现的影响，有效地克服了以动词为核心的投射在句法解释方面的理论缺陷。

【流程】

事件拆分

事件标注

根据标注结果中的事件语义标签进行事件语义模式聚类

层次聚类算法（离差平方和函数作为目标函数）

- 1、设定预期模式数目  $k$ ，计算不同模式间的距离，得到邻近度矩阵；
- 2、合并最小距离的两个模式；
- 3、计算新模式与其他模式的距离；

批注 [ZHe1]: 我一直有一个疑问，相比于事件语义结构，“主谓宾”这种句法层次的解释有哪些缺陷？

4、重复 2，3 直至矩阵中模式数目小于等于  $k$ ;