【之前做原子事件抽取都是采用的词法分析层的，于是想尝试着用依存句法分析去做原子事件边界检测任务，然后上网搜了一下相关资料，就找到了这篇文章。】

【题目】基于依存分析的事件识别
【作者】付剑锋、刘宗田、付雪峰、周文、仲兆满
【单位】上海大学计算机工程与科学学院
【期刊】计算机科学

【方法描述】用依存分析发掘触发词与其他词之间的句法关系，以此为特征在 SVM 分类器上对事件进行分类，最终实现事件识别。
【评测会议】MUC 消息理解会议；TDT 话题识别与跟踪；ACE 自动内容抽取
【相关工作】

文献【5】针对生物医学领域，根据该领域文献语法特征构造了一系列规则，封装成抽取器，抽取其中的事件；

文献【6】对在线新闻中的气象事件构造规则，抽取关于气象方面的事件。

文献【7】用 MegaM 作为二元分类其和 TiMBL 作为多元分类器两种机器学习方法实现了事件抽取，在 ACE 英文语料上取得了不错的效果。

文献【8】利用手工确定的句型模版构造了抽取规则，用于从处理后的文本中抽取事件信息填充句型模版中的槽。

文献【9】通过语句聚类的方法获得事件的信息结构（事件模版），以抽取相应事件。

文献【10】采用及其学习的方法改进了文献【7】中训练集的正反例不平衡以及数据稀疏的不足，在 ACE 中文语料上取得了较好的效果。

[5] Yakushiji A, Tateisi Y, Miyao Y, et al. Event extraction from biomedical papers using a full parser. 2001: 408-419

[6] Lee C S, Chen Y J, Jian Z W. Ontology-based fuzzy event extraction agent for Chinese e-news summarization[J]. Expert Systems with Applications, 2003, 25(3): 431-447

[7] Ahn D. The stages of event extraction[C]// Proceedings of the COLING-ACL 2006 Workshop on Annotating and Reasoning About Time and Events. 2006: 1-8

[8] 吴平博，陈群秀，马亮. 基于事件框架的事件相关文档的智能检索研究[J]. 中文信息学报，2003，17(06)：25-30

[9] 杨尔弘. 突发事件信息提取研究[D]. 北京：北京语言大学，2005

[10] 赵妍妍，秦兵，车万翔，等. 中文事件抽取技术研究[J]. 中文信息学报，2008，22(1)：3-8

采用机器学习的方法识别事件，就是借鉴文本分类的思想，将事件的识别转化成为分类问题。
【依存分析】

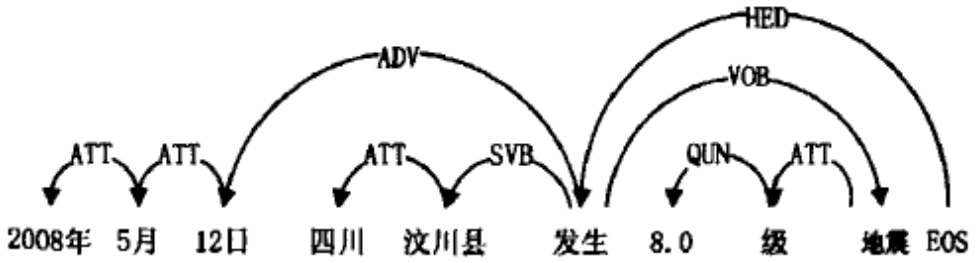
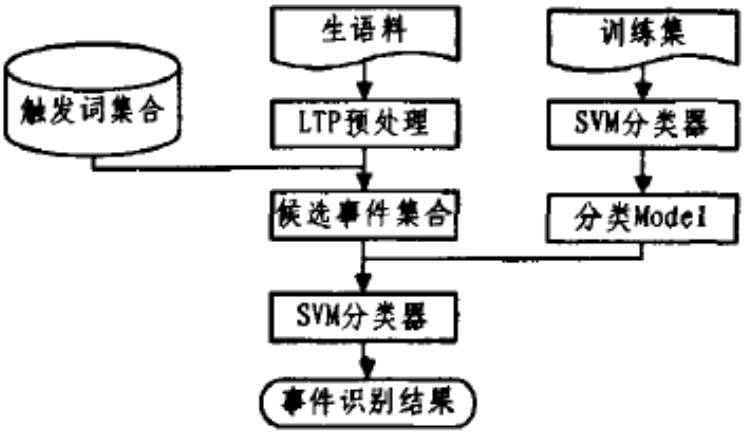


图 1　依存分析示例

HED 表示句子的核心词；
ATT 表示定中关系；
SVB 表示主谓关系；
VOB 表示动宾关系；
QUN 表示数量关系；
本文采用了哈工大的 LTP 平台实现依存分析，LTP 对句子处理后的输出是一个以词为单位的三元组链表，每个元组都可以表示为（POS,DR,PID），其中 POS 表示词性，DR 表示依存关系，PID 表示父节点编号。
【事件识别】
定义 1 事件：在某个特定的时间和地点下发生、由若干角色参与、表现出若干动作特征的一件事情。其中时间、地点、事件参与的对象称为事件要素。
定义 2 事件触发词：可以用来清晰的表示所发生的事情的词。一般情况下，触发词是句子中的主要动词（也可能是名词），触发词直接描述了事件。
定义 3 事件识别：从包含触发词的句子（文本）中找出现实世界发生的事件。
【事件识别过程】



本文选取的与触发词相关的特征如下：
1）触发词以及触发词的词性；
2）触发词左侧 8 个词及其词性；
3）触发词右侧 9 个词及其词性；
4）上述词之间的依存关系。
【实验结果分析】
SVM 分类器采用 libSVM
实验语料来自于 Web 上收集的各类事件的报道，包括地震、台风和海啸等 8 类事件的 416

篇文档。其中 300 篇作为训练语料、116 篇作为测试语料，并建立了相关事件的触发词库。

表 2　事件识别实验结果

Features	precision	Recall	F-measure
Word	58.4%	56.2%	57.3%
Word+POS	65.6%	60.3%	62.8%
Word+DR	67.7%	63.5%	65.5%
Word+POS+DR	71.6%	67.2%	69.3%