1. 事件

是促使事情状态和关系改变的条件。

1. 事件抽取的任务可以分为2大类：
   1. 事件识别和抽取：input描述事件信息的文本，output时间、地点、人物、与之相关的动作/状态的改变。
   2. 事件检测和追踪：将文本分解为：事件、新（不可预见）的事件。  
      分为：从按照时间排序的新闻文档中发现以前没有识别的事件；从实时新闻流中实时发现新的事件。
2. 核心概念  
   Event Mention:描述事件的词组/句子/句群

Event Trigger:一般是动词/名词，Event Mention中最能代表事件发生的词汇

Event Argument:事件的重要信息/entity mention,主要由实体、属性值等表达完整语义的细粒度单位组成

Argument Role: Event Argument在事件中扮演的角色，Event Argument与Event的关系

Event Type

1. 事件抽取  
   从文本中找到信息，进行填表

EX:

Text: 苹果公司将于西部时间9月12日上午10点，北京时间9月13日凌晨1点举行新品发布会。这一次的发布会地点是全新建造的史蒂夫·乔布斯剧院。根据目前的消息，这次发布会上苹果发布iPhone8(命名不确定)，iPhone7, iPhone 7sPlus, Apple Watch以及全新的AppleTV.

Table:

|  |  |
| --- | --- |
| 事件类型 | 发布会 |
| 公司 | 苹果公司 |
| 地点 | 西部时间9月12日上午10点 |
| 地点 | 史蒂夫·乔布斯剧院 |
| 产品 | iPhone8, iPhone7s. iPhone7s Plus, Apple watch3, apple TV |

1. 基于模式匹配的方法

在特定领域中性能好

（召回率：所有准确的条目中，有多少被检索出来了）

1. 基于人工标注语料

从一个大的标注集开始创建模板。模板的产生完全基于人工标注语料。

1. AutoSlog(Riloff)

基本假设：①事件首次提及之处即可确定该元素与事件间的关系。②事件周围的语句中包含了事件元素在事件中的角色描述。

通过监督学习和人工审查来建立抽取规则。通过训练数据中已经填充好的槽，来解析槽附近的句法结构，从而自动生成抽取规则。这个过程中抽取产生的模板过于general,所以需要人工审核。

本质上形成的是一个字典。

1. PALKA

基本假设：特定领域中高频出现的语言表达方式是可数的。

用语义框架和短语模式结构来表示特定领域中的抽取模式。通过融入WordNet的语义信息，PALKA在特定领域可以取得接近纯人工抽取的效果。

1. 基于弱监督

人工标注耗时耗力，且存在一致性问题。

弱监督方法不需要对语料完全标注，只需要人工一定的预分类or指定种子模板即可。

机器将根据分类语料or种子模板自动进行模式学习。

1. AutoSlog-TS( Riloff and Shoen,1995)

不需要文本的标注，只需要一个预先分类好的语料库，对于每一个根据句法分析识别出来的名词短语都产生对应的抽取规则，然后整体过一遍语料库，产生每个规则的统计数据。

1. TIMES （Chai AND Biermann）

领域无关的概念知识库WordNet，提升模式学习的泛化能力。通过人工或规则进行词义消歧，使得最终的模式更加准确。

1. NEXUS （Piskorski et al,2001）

用聚类方法对语料进行预处理

1. GenPAM （ Jiang,2005）

利用模式间的相似性实现词义消歧

（词义消歧：在特定语境中，识别出某个歧义词的正确含义。

比如“小米”，既可以指谷物，又可以指小米科技公司）

从而在由特例生成泛化模式的学习过程中，减少人的工作量。

1. 基于统计的方法——传统机器学习
   1. Pipeline:将事件抽取转化为多阶段的分类问题。

需要顺序执行以下分类器：

Trigger Classifier:是事件触发词？事件类别是？

Argument Classifier:词组是事件元素？

Role Classifier:判断元素的角色类别

Attribute Classifier:判断事件属性

Reportable-Event Classifier:是否存在值得报告的事件实例

分类器：MaxEnt , SVM

重点：提取和集成有区分性的特征。包括句子级信息和篇章级信息

1. 句子级信息
2. 篇章级信息

跨文档利用全局信息

不仅要考虑当前的置信度，还要考虑与带抽取文本相关的文本对它的影响，以及全局信息。

* Ji and Grishman,2008
* Liao and Grishman,2010
* Hong et al,2011
* Liu et al,2016a

1. 评价

早期大部分研究都基于Pipeline方法。

存在的问题：误差传递；各个环节任务是独立的，缺少互动；无法处理全局的依赖关系

* 1. Joint model
     + 1. Joint Inference

将各个模型通过整体优化目标整合起来

* + - 1. Joint Modeling

把事件结构看作依存树，抽取任务转化为依存树结构预测问题

评价：对于词义消歧非常棒，但是在未见词上缺乏泛化

* *什么是依存树？*

*··ex:猴子喜欢吃香蕉*

*·· 简单的短语分词：（方法有：正向逆向最大匹配，n-gram,机器学习）*

*猴子 喜欢 吃 香蕉*

*·由分词转为词性标注*

*猴子/NN 喜欢/VV 吃/VV 香蕉/NN. /PU*

*·由词性标注生成短语句法树*

*（ROOT*

*(IP*

*(NP(NN 猴子))*

*（VP (VV喜欢)*

*（IP*

*(VP(VV吃*

*（NP(NN香蕉)））））*

*（PU .）))*

*·依存树：有依存关系的树形图*

*短语句法树转成依存树的工具：*

*Penn2malt*

*Stanford parser*

*LTH*

相关工作：

* Li et al,2013a. Li提出基于结构感知机的联合模型同时完成事件触发词识别和事件元素识别两个任务，并通过beam search缩小搜索解空间
* Li et al.2014 利用结构预测模型。将实体、关系和事件进行联合抽取。从而利用了更多的句子级信息

1. 基于统计的深度学习
   1. 深度学习相对于传统机器学习的优势：

· 对外部NLP工具的依赖很小

·使用词向量作为输入

·自动提取句子特征

* 1. DMCNN

传统CNN只保留句子中最重要的信息，会丢失部分；而DMCNN可以实现对一个句子中的不同部分的最大值获取，保留了更多有价值的信息。

* + - 1. DMCNN把事件抽取看作是两个阶段的多分类任务：触发词分类；论元分类。
  1. JRNN