*(2017 ACL)Joint Learning for Event Coreference Resolution*

时间：2021/8/15

绝大多数事件共指解析器都是基于管道的，其中从触发检测组件到事件共指组件的错误传播是主要的性能限制因素。为此，提出了一个联合学习事件共指、触发检测和事件回指的模型。选择这三个任务的部分原因是它们之间的相互依赖性。

1 触发词检测性能对事件共指性能有巨大影响，事件共指也可以改进触发词检测。

2 Wiseman等人(2015)的工作表明联合训练一个名词短语回指模型（a noun phrase

anaphoricity model）和一个提及排序模型（a mention-ranking model）对于实体共指解析是有好处的。

3 事件回指和触发词检测也可以相互受益。回指是一个辅助任务：它的预期用途是改进触发词检测和事件共指，潜在地调解触发词检测和事件共指之间的交互。

**数据集**

**英语训练集**：LDC2015E29、E68、E73和E94。这些语料库由两种类型的文档组成，新闻专线文档和论坛文档。它们包含648个文档，18739个事件提及，分布在9955个事件共指链上。

**英语测试集**：由169篇新闻报道和论坛文档组成，4155个事件提及分布在3191个事件共指链上。

**中文训练集**：LDC2015E78、E105和E112。这些语料库仅由论坛文档组成。它们包含383个文档，4870个事件提及，分布在3614个事件共指链上。

**中文测试集**：由167个新闻和论坛文档组成，2518个事件提及分布在1912个事件共指链上。

**模型**

给定一个测试文档，首先从中提取(1)所有单个单词的名词和动词，以及(2)在训练数据中作为触发词至少出现过一次的所有单词和短语，将这些提取的单词和短语视为候选事件提及。

模型的目标是对文档中提到的候选事件进行联合预测。将对每个候选事件提及做出三个预测，对应模型中的三个任务:触发词子类型、回指性和先行词。

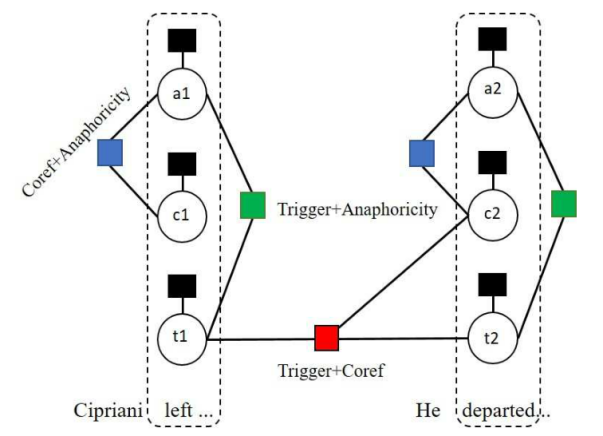
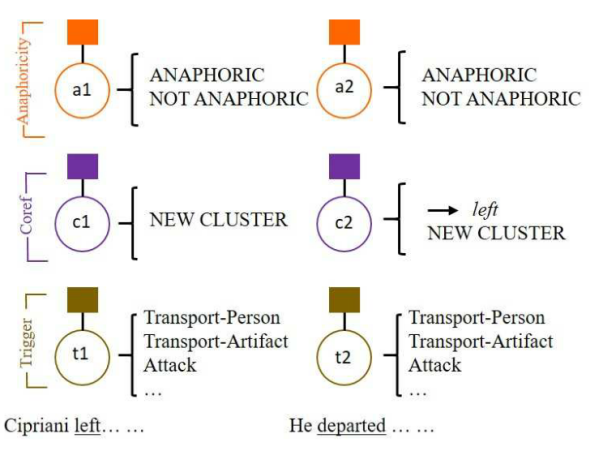
定义三种变量：事件子类型变量（t）、回指变量（a）、共指变量（c）。每个候选事件都与一个共指变量、一个事件子类型变量和一个回指变量相关联。模型在这些变量上归纳出以下对数线性概率分布:



其中θ是与特征函数fi关联的权重，x是输入文档。

**模型捕获两种特征：**

1. 特定任务的特征，特定任务的特征被编码成一元因子，每个因子都与相应的变量相关联(图1)；
2. 跨任务特征，跨任务特征用二元或三元因子编码，每个因子耦合两个任务的输出变量(图2)。



**特定任务的特征**

1. 触发检测模型

当单独应用时，触发词检测模型返回给定候选触发词的可能子类型的分布。

候选触发词t的表示：t的单词、t的引理、由来自t的窗口大小为3的单词二元模型以及通过将t的引理与以下每个特征配对而创建的特征连词：语法上最接近t的实体的头词，文本上最接近t的实体的头词，语法上最接近t的实体的实体类型，以及文本上最接近t的实体的实体类型。对于带有**动词触发词**的事件提及，使用头词及其主语和宾语的实体类型作为特征；对于带有**名词触发词**的事件提及，创建了与动词触发词相同的特征，用启发式提取的动因和受动者（agents and patients）替换主语和动词。

对于**中文触发检测**，另外从t中的每个字符创建两个特征，一个编码字符本身，另一个编码中文同义词词典中相应字符的条目号。

1. 事件共指模型

采用一个提及排序模型，从候选先行词集合中为要解决的提及选择最可能的先行词。单独应用时，该模型被训练为最大化条件可能性，在训练文本中将提及共同解析为其正确的先行词。

**代表空候选先行词的特征：**除了mj的词和mj的引理之外，使用特征连词——在mj的引理和mj前面的句子数之间创建一个连词，在mj的引理和文档中mj前面的提及数之间创建一个连词。

**代表非空候选先行词的特征**：mi：mi的词，mi的引理，mi和mj是否有相同的引理，特征连词包括:

1. mi的词与mj的词成对，
2. mi的引理与mj的引理成对，
3. mi和mj的句子距离与mi的引理和mj的引理成对，
4. mi和mj之间的提及距离与mi引理和mj引理配对，
5. 由mi和mj的主语及其引理组成的四元组，
6. 由mi和mj的宾语及其引理组成的四元组。
7. 回指确定模型

单独使用时，回指模型返回给定事件提及是回指的概率。

为了训练模型，使用以下特征来表示每个事件提及mj:

1. 每个候选先行词的头词与mj的词配对，
2. 是否至少有一个候选先行词具有与mj相同的引理，
3. mj在训练数据中是回指的概率(如果mj从未出现在训练数据中，则该概率被设置为0.5)。

**跨任务交互特征**

1. 触发检测和回指

结合每个候选事件提及的事件子类型、其触发器的引理及其回指的特征。

1. 触发检测和共指

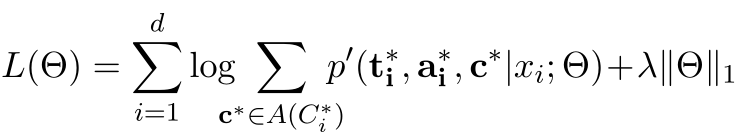
定义联合共指和触发检测因子，使得子类型变量ti和tj上定义的特征仅在提及mj与前面提及的mi共指时触发。这些特征是:

1. mi和mj的子类型对，
2. mj的子类型和mi的词对，
3. mi的子类型和mj的词对。
4. 共指和回指
5. 事件提及mj的回指性，
6. 是否为mj选择了非空先行词。

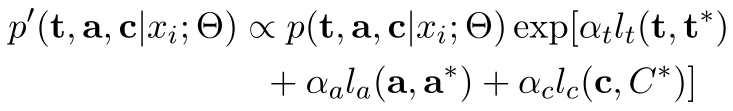
**训练**

学习前准备：首先，需要从C\*i（gold event coreference partition）中导出黄金回指标签a\*i。这很简单:首先提到的每个共指链都不是回指，而其余的都是回指；第二，使用黄金事件提及进行模型训练，此外，还根据对应于非触发词的候选事件提及来训练模型；第三，由于模型以先行向量的形式产生事件共指输出(每个事件提及一个先行)，它需要在先行向量上进行训练。

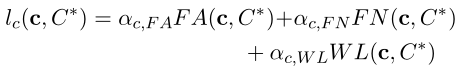
接下来，学习模型参数，以利用L1正则化最大化训练数据的以下条件似然性:



p’是通过用特定任务的参数化损失函数扩充分布p获得的:



事件共指的损失函数*l*c是三种错误类型计数的加权和:



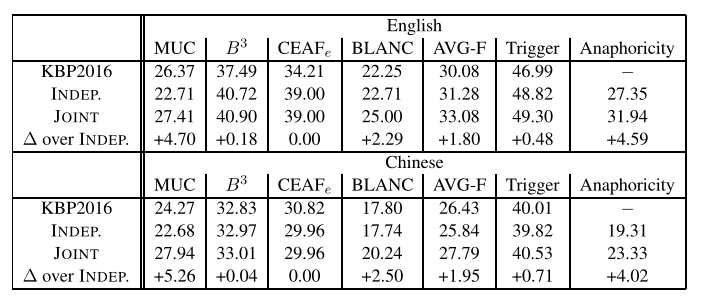
其中FA(c，C\*)是非回指提及被误分类为回指的数量，FN(c，C\*)是回指提及被误分类为非回指的数量，WL(c，C\*)是被错误解析的回指提及的数量。

**推理**

推理是在训练和解码过程中进行的，涉及到计算一个变量或一组与一个因子相关联的变量的边际值。为了提高效率，使用信念传播而不是精确推理来执行近似推理。作者使用最小贝叶斯风险，计算每个变量的边际，并导出每个变量的最大可能分配。

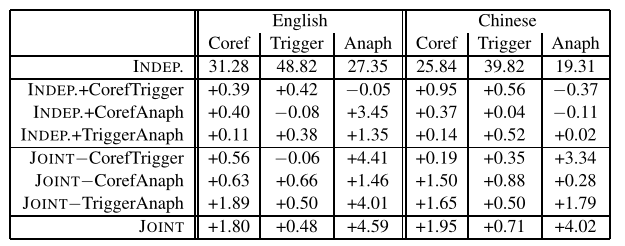
**评估**

指标：MUC、B3、CEAFe、 BLANC、AVG-F



第1行显示了参加KBP2016的最佳共指系统的表现(Lu和ng，2016)。第2行显示了独立模型的性能，每个模型都是独立于其他模型进行训练的。联合模型的结果显示在第3行。第4行显示了联合模型和独立模型之间的绝对性能差异。

**模型消融**



**错误分析**

1. **两种主要类型的精度误差**
   1. 错误的触发词和错误类型的触发词
   2. 无法提取参数
2. **召回错误的三种主要类型**
   1. 缺少触发词
   2. 缺少实体共指信息
   3. 缺乏上下文理解

**总结**

作者提出了一个事件共指解析、触发检测和事件回指确定的联合模型。该模型在任务选择和跨任务交互方面进行了创新。