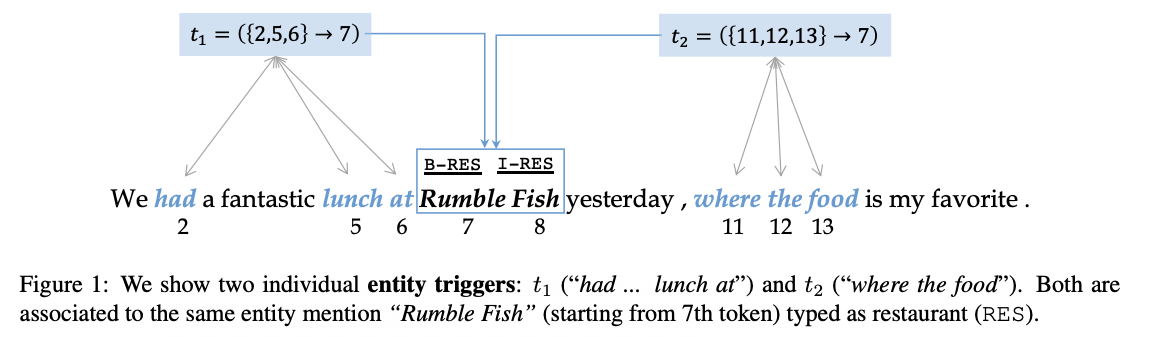
**阅读笔记一**

TriggerNER: Learning with Entity Triggers as Explanations for Named Entity Recognition

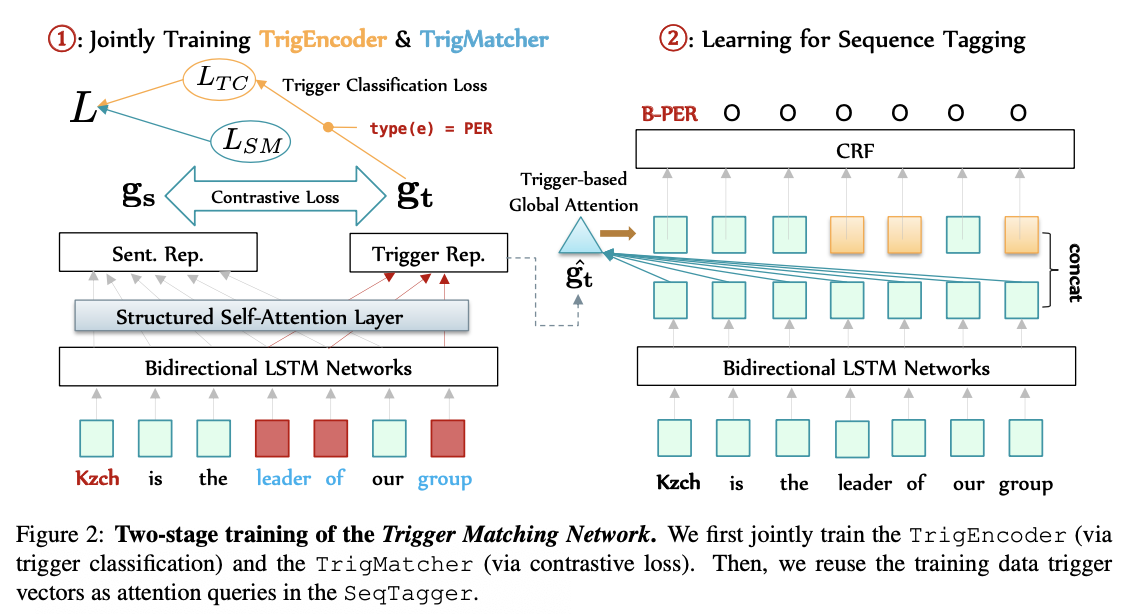
1. 贡献
2. 引入了实体触发器(entity trigger)的概念，在CoNLL和BC5CDR这两个数据集上众包并公开发布了14k带实体触发器标注的数据集。
3. 提出Trigger Matcing Network框架，用于编码实体触发器和在无标注的句子基础上进行软连接，提高实体标记器的效率。
4. 方法
5. 问题构想



如Figure 1，给定句子只有一个实体，也就是餐厅“Rumble Fish”，用B-RES和I-RES标注，两个实体触发器和就是描述餐厅的单词集合。，带标注的语料，，未标注的语料； T(x,y)，带触发器标注的语料，每个触发器中，有一个实体索引e和多个单词索引w,即

添加了实体触发器的新数据集格式：

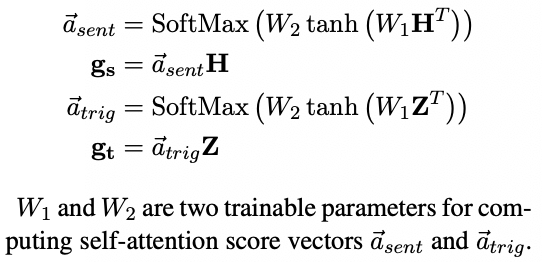
1. Trigger Matching Networks
2. TrigEncoder & TrigMatcher



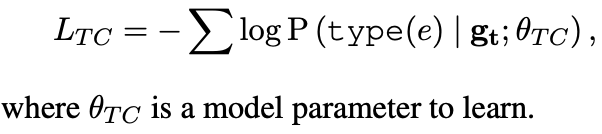
给定含有多个实体的句子输入x，假设每个实体都有一系列触发器，为了便于批处理，将数据集重新格式化为一个实体仅有一个触发器的新序列。然后将一个实体匹配一个触发器，创建新的训练实例，格式为)。

首先，对格式化的训练实例(x,e,t)，用LSTM获取x和t的上下文词向量，分别用H和Z表示；

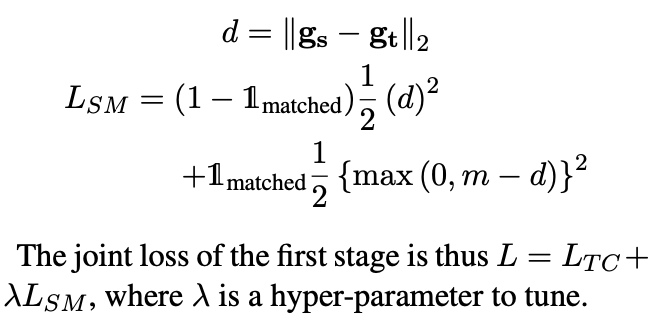
然后，学习触发器和句子基于注意力的表示，得到代表输入x的加权向量表示和触发器加权向量表示：



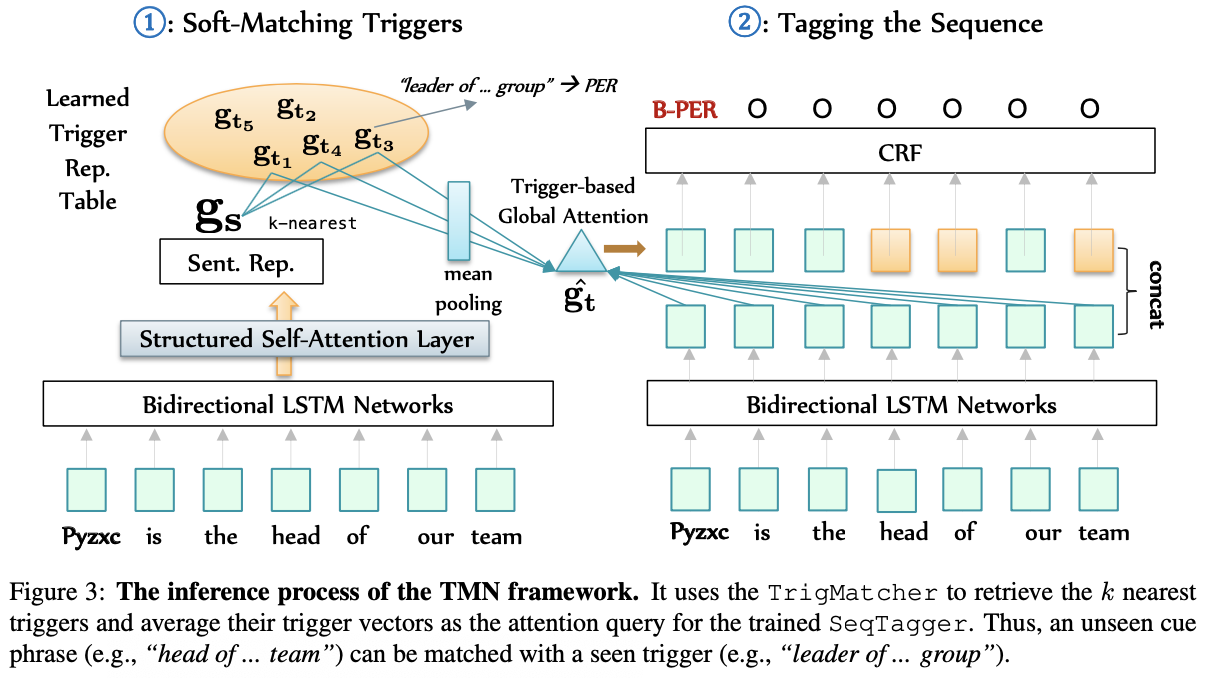
要使用关联实体作为指导触发器表示的监督，因此将送到多分类器中，预测在触发器和参数条件下，关联实体e的类型，触发器分类的损失函数如下：



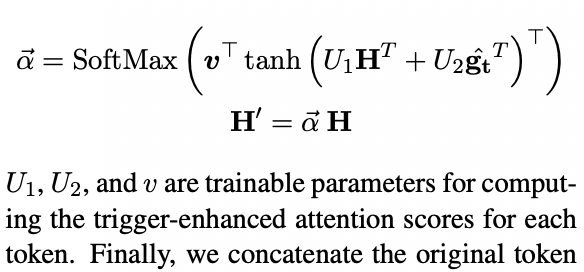
为了学习基于注意的向量表示来匹配触发词和句子，随机引入负样本进行训练，最小化触发器t和句子x距离，其中，λ是待调整的超参数：



1. SeqTagger



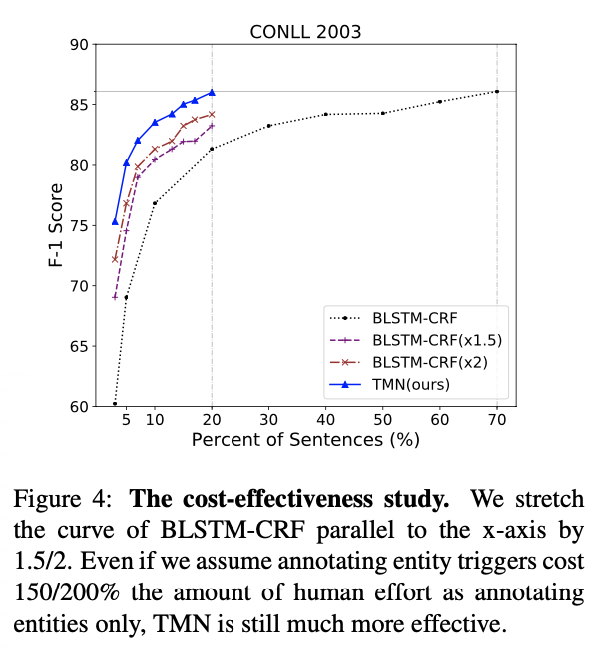
与TrigEncoder和TrigMatcher模块一样，使用同一个BiLSTM来获取所有输入x的隐藏层向量H，用前文训练得到的TrigMatcher来计算所有触发器向量的平均值，用来创建基于注意力的向量表示：



将原始的输入向量表示和触发器增强的向量表示拼接起来作为CRF的输入，进行序列标注。进行解码的时候并不知道句子的触发词，故使用TrigMatcher来计算句子和触发词向量之间的相似度，将最合适的触发词作为额外输入送到SeqTagger中。

对每个待标注的句子x来说，首先计算其基于字关注的向量，用L2范数距离来计算对比损失，在共享空间中有效检索与句子相似度最高的触发向量，然后计算出发向量平均值，将其SeqTagger的关注查询，进行序列预测。

1. 实验

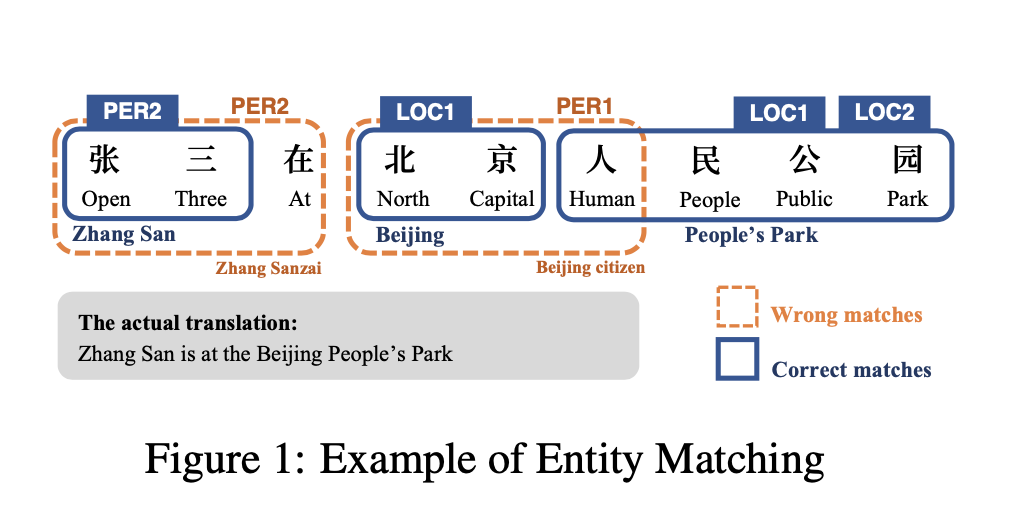


有Fig4可以看出，TMN在仅用20%带触发词标注的数据集的情况下，相比于用了50-70%原始训练集的模型，模型的表现有了显著的提升。

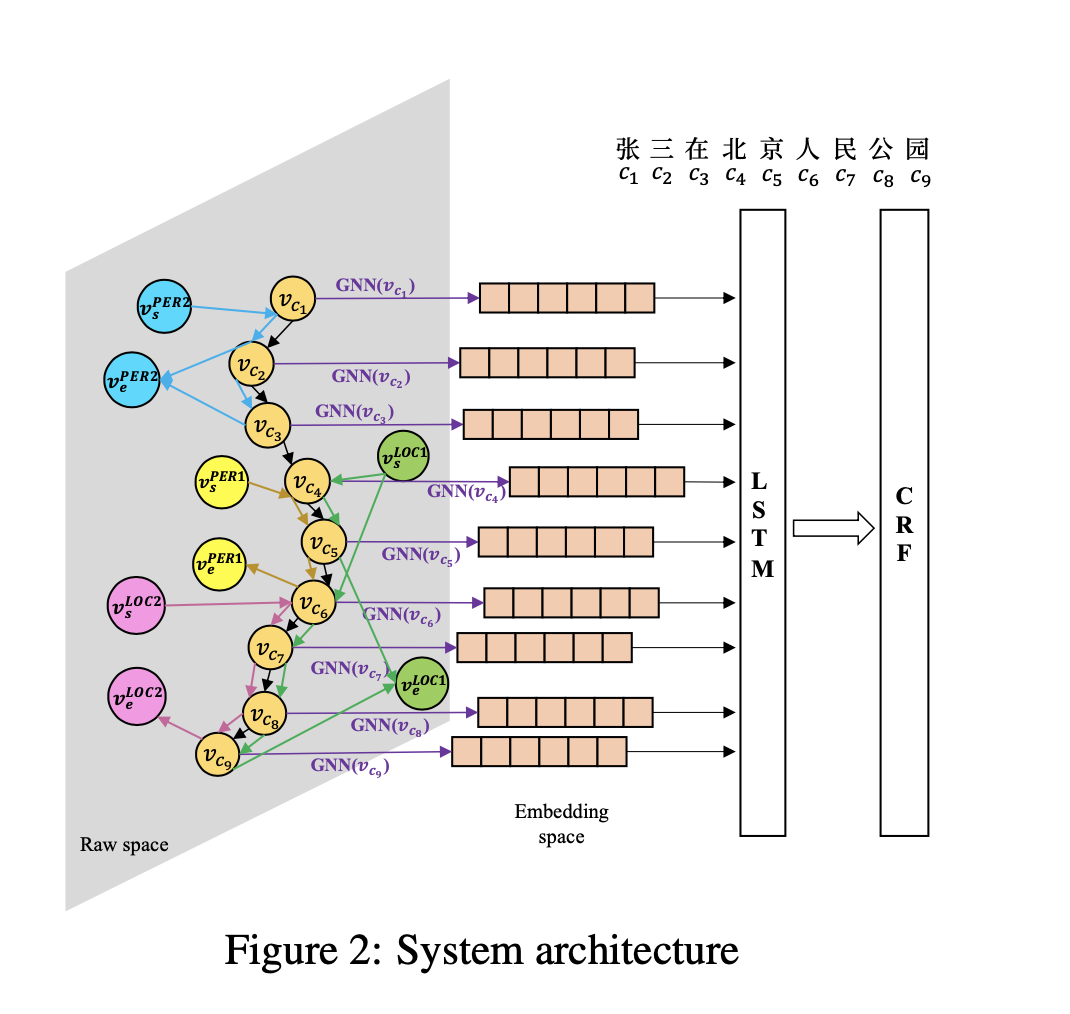
**阅读笔记二**

A Neural Multi-digraph Model for Chinese NER with Gazetteers

1. 贡献
2. 本文提出一个多图模型，该模型学习如何整合gazetter中的信息，并解决上下文学习中的冲突匹配；
3. 发布了一个电商领域的新的数据集
4. 方法

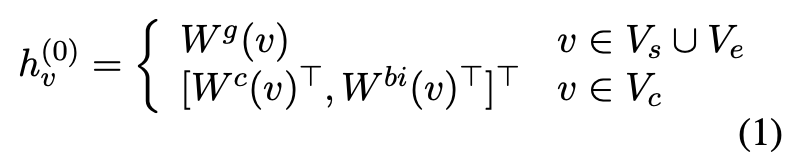


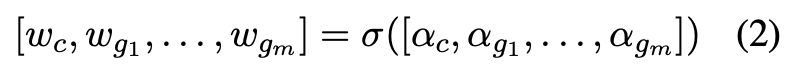
对于“张三在北京人民公园”这句话，有9个中文字符，4种不同类型的实体（北京人、张三在、北京、人民公园）



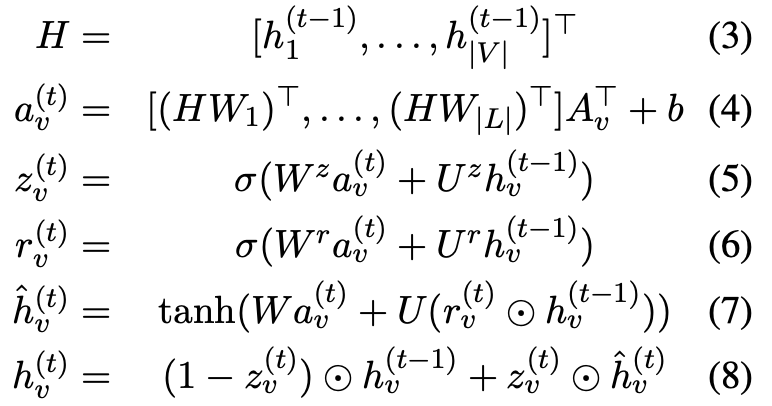
图中分为三个部分，一是中文字符的表示，二是GGNN中的字符特征学习，三是用LSTM+CRF解码。

1. Text Graph：节点为字符，边为文本词的顺序；其中节点包括实体的起始字符和终止字符。边表示词之间的关系，图中黑色的边表示词的顺序，彩色的边和箭头表示gazetter中包括起始节点和终止节点的实体（蓝、黄、绿、粉）。
2. Adapted GGNN：一种改进的GGNN，用于学习节点的向量表示。图存储在邻接矩阵A中，A包括不同标签的边，学习适用于我们任务的gazetter索引信息的加权组合。首先扩展邻接矩阵A使其涵盖不同标签的边；然后为每条边定义参数，,…,，这些参数用于描述每种信息对任务的贡献。

，其中Wg是实体起始词和终止词，Wc 是字符向量表示，W bi 是双字（多个字组成的词）嵌入。

每条边的参数α即为邻接矩阵A中边的权值：

同一个标签的边具有相同的权重，然后用GRU更新权重，神经网络的训练过程如下：

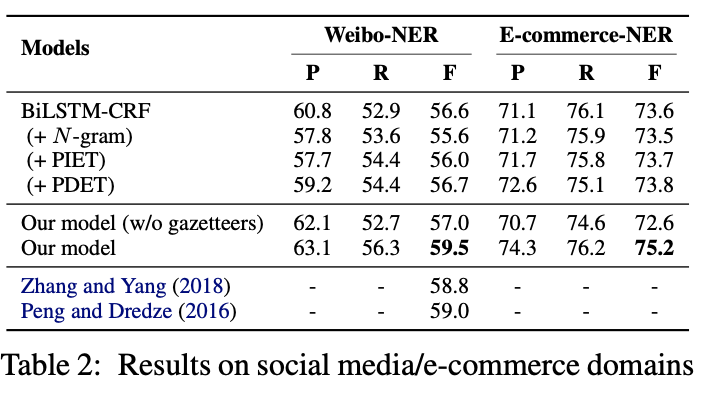
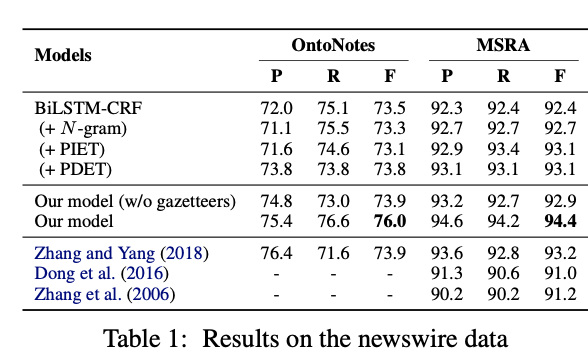


其中是节点v在t时刻的隐藏层向量，是邻接矩阵A中节点v对应的行向量，W(字符向量的权重)和U(隐层向量的权重)是待调节的参数。等式(3)创建隐层状态矩阵H，等式(4)表示要通过相邻节点传播的信息。公式(5)、(6)、(7)、(8)表示整合来自相邻节点的信息和节点当前的隐层状态，用以计算t时刻的隐藏层状态，最终得到节点的最终状态。

1. LSTM+CRF：前面GRU学习到字符特征表示后，BiLSTM-CRF先用LSTM获取长距离上下文特征，CRF进行似然概率统计输出最佳序列。
2. 实验

数据：新闻数据：OntoNotes、MSRA；社交媒体数据Weibo-NER；作者构建的电商数据：E-commerce-NER，四个数据集都有一定的有匹配冲突的实体，其比例分别是36.10%, 44.75%, 36.10%, 46.05%。

基准模型是LSTM-CRF，作者也探索了三种添加了gazetter特征的方法：n-gram、位置无关实体类型(PIET)、位置相关实体类型(PDET)。



如上所示，table 1是在新闻数据及上做实验得到的结果，table 2是在社交媒体数据集和作者发布的数据集上做实验得到的结果。

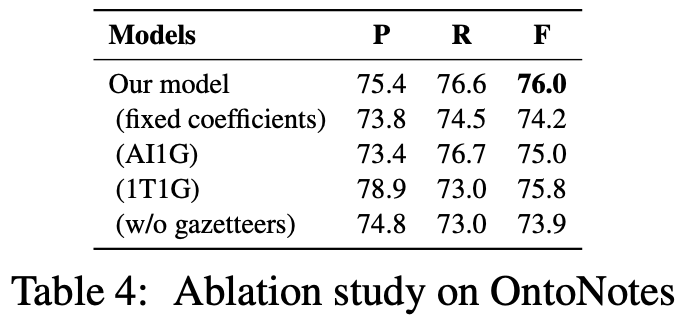
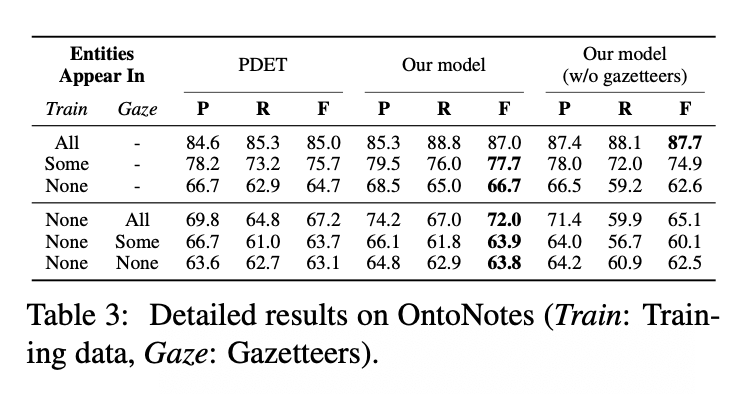
 

Table 3 是对比了PDET方法和作者提出来的方法得到的实验结果，最后，作者做了OntoNotes数据集上做了实体冲突的消解实验，探索gazetter的加权组合对模型的贡献。

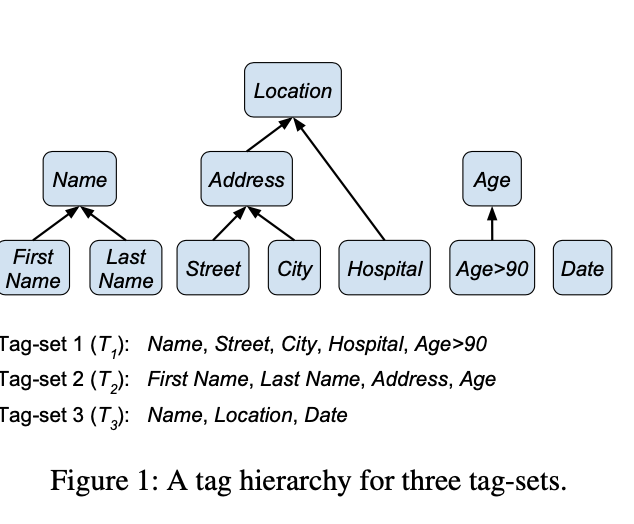
**阅读笔记三**

A Joint Named-Entity Recognizer for Heterogeneous Tag-sets Using a Tag Hierarchy

1. 贡献

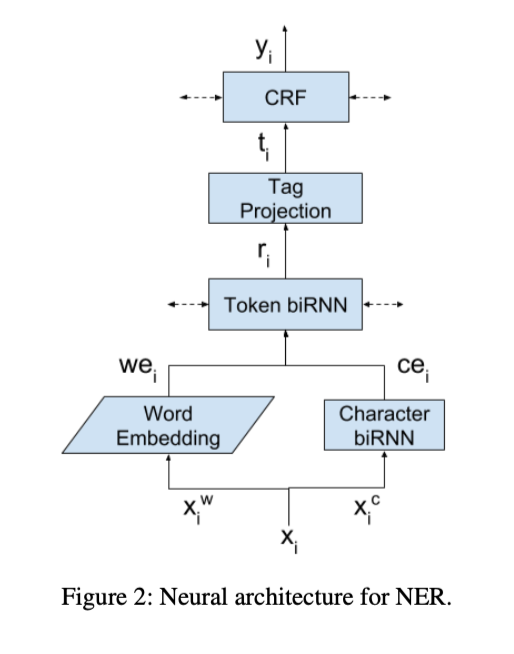
针对NER任务中异构标签集合提出了标签层次结构模型，在预处理阶段组合标签，应用了标签层次 ，学习和预测阶段共享模型所有参数，故最后的结果不需要其他处理，创新点在于利用语义相关性构建层次标签。

1. 方法介绍

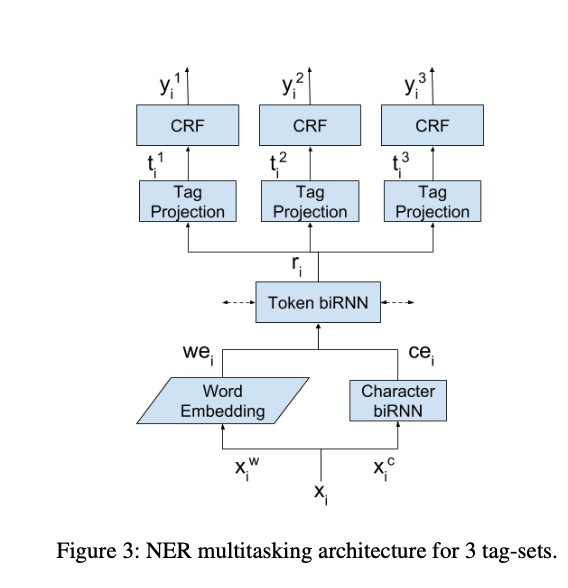


对三种不同的数据集，有两种训练方法：

**Independent Model**：每次对单一数据集进行训练：对图一中不同训练集构建CRF模型训练，训练结束后将各个模型的预测结果合并。



**Multitasking for heterogeneous：**不同的label共享语料的特征，但每个数据集使用独立的tagging layer。最后分别使用CRF进行似然概率统计。统计得到的结果有三种合并的方式（随机选择、CRF条件概率最高的序列、CRF边缘概率最高的一项）。

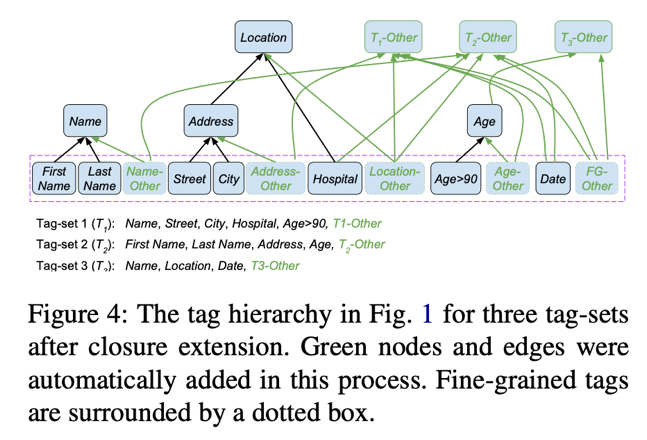


1. 给定输入，每个输入首先用词向量矩阵和字符向量矩阵嵌入到向量空间中，得到和；
2. 将这两个向量进行拼接得到；
3. 然后再将送到BiRNN中再次编码得到文本的向量表示；

(4) 最后，将每一个重新编码的投影到目标标记空间T中，得到；将序列作为CRF层的输入进行解码。

**本文创新点：在标签上构建一种标签层次**

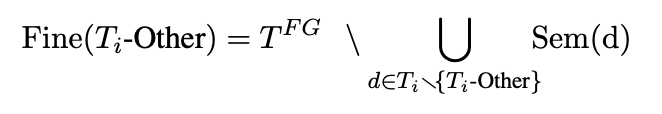
1. **预处理阶段将不同数据集的标签统一格式**
2. **训练时只有训练一个CRF模型，即训练时不仅共享语料特征，也共享tag layer，即手工构造的标签层次，内容如下图：**



在这个有向无环图中，节点表示句子中的单词，一条直接相连的边c->d表示c是d的下义词，c是d的子集，如‘LastName’->‘Name’,‘Street’->‘Location’

用Sem(d)={d}U{c|c->d}表示捕获d语义的某些子集的所有标记的集合，其中->表示存在c到d的定向路径。如Sem(Name)={Name,LastName,FirstName}；如果一个词没得子集，Sem(d)=(d)它表示某细粒度标记，将这些标记用表示，用Fine(d)=∩Sem(d)表示所有与d同义的标签，如Fine(Name)={LastName, FirstName}

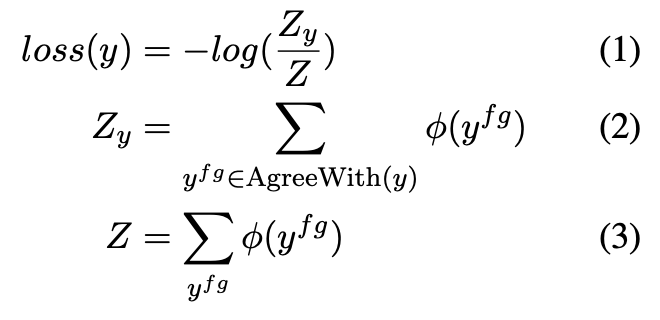
Hierarchy model仅从中预测标签序列，然后将预测结果映射到一个目标标签集上。  
为了解决语义完备问题，用“Other”标签来表示其他未被标注出来的语义。如d∉，则‘d-Other’->d会被自动添加到图中，‘d-Other’∈Fine(d)；同时，未被捕获到的标签集合会被表示为‘-Other’并自动添加到标签集合中，如图4所示。表示所有未被获取到的语义信息：



在训练过程中被用于远程监督，这个步骤中所有扩展都是作为模型算法的一部分自动完成的。

模型训练：由于只有一个标签层，故所有训练实例都共享模型参数。

模型损失函数如下：其中φ(y) 表示序列y的得分，Z为规范化因子，Zy为预测值得分



1. 实验

实验在医疗数据集(I2B2-2006, I2B2-2014)和新闻数据集CoNLL-2003的英语部分上完成，  
