论文阅读笔记一

Named Entity Recognition as Dependency Parsing

补充知识：ElMo、cross-view training(CVT)

平面实体：

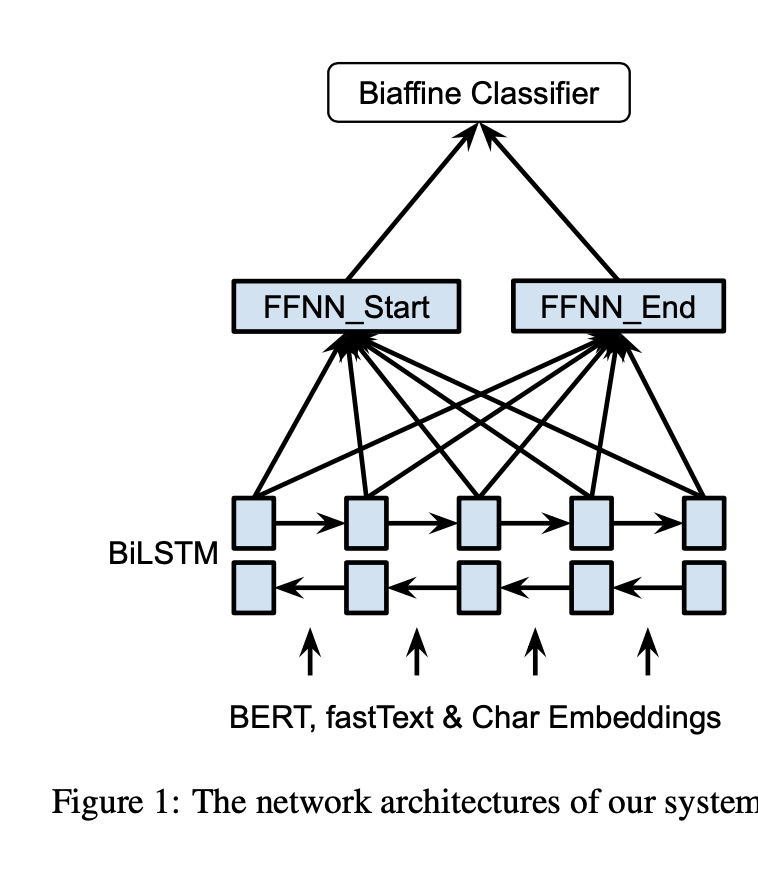
嵌套实体：Meizhi Ju 2018，LSTM-CRF，A Neural layered model for nested named entity recognition，迭代预测实体，直到再无实体产生。韩先培团队，Sequence-to-nuggets: Nested entity mention detection via anchor-region networks。先探测实体头，再推断实体的边界，给实体贴标签。

本文灵感来源：DEEP BIAFFINE ATTENTION FOR NEURAL DEPENDENCY PARSING

1. 贡献
2. 方法

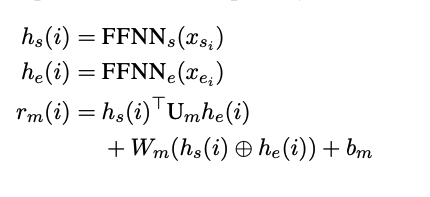
输入：词向量+字符向量

模型：BiLSTM——>Biaffine model



BERTlarge和fasttext作为词向量输入。使用BERT获取上下文依赖向量；用CNN获取字符级信息。单词和字腹肌的向量送到BiLSTM中进行建模，获取单词表示。

用两个独立的FFNN来获取一串文本的开始字和结束字，创建开始字和结束字的两种不同的表示形式hs和he。最后用双仿射模型(Biaffine Model)创建一个l×l×c评分张量，其中l为句子长度，c为实体种类数目+1(非实体的标签)。



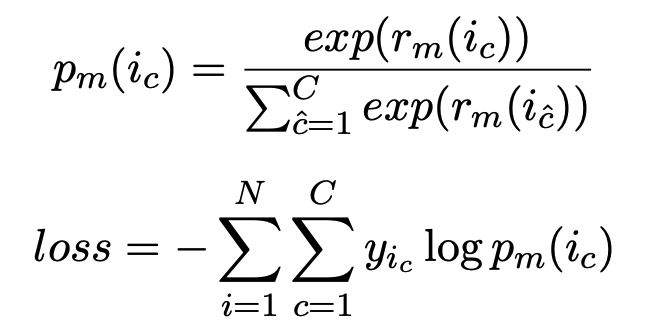
其中Um是d×c×d的张量，Wm是2d×c的矩阵，Wm是偏置项。最终得到的rm是所有可能的实体段的得分。对所有的得分进行最大化，获取每个字段最有可能的类别：



最后对除了非实体的所有字段的得分进行降序排列，然后对嵌套实体和平面实体用以下两种约束进行处理：

对于嵌套实体，只要它不与排名较高的实体的边界冲突即可；对于平面实体若实体已经被包含到嵌套实体中，则该实体不会被选中参与计算。

模型的学习目标是为每个字段分配一个正确的类别标签。使用softmax交叉熵模型进行模型优化：

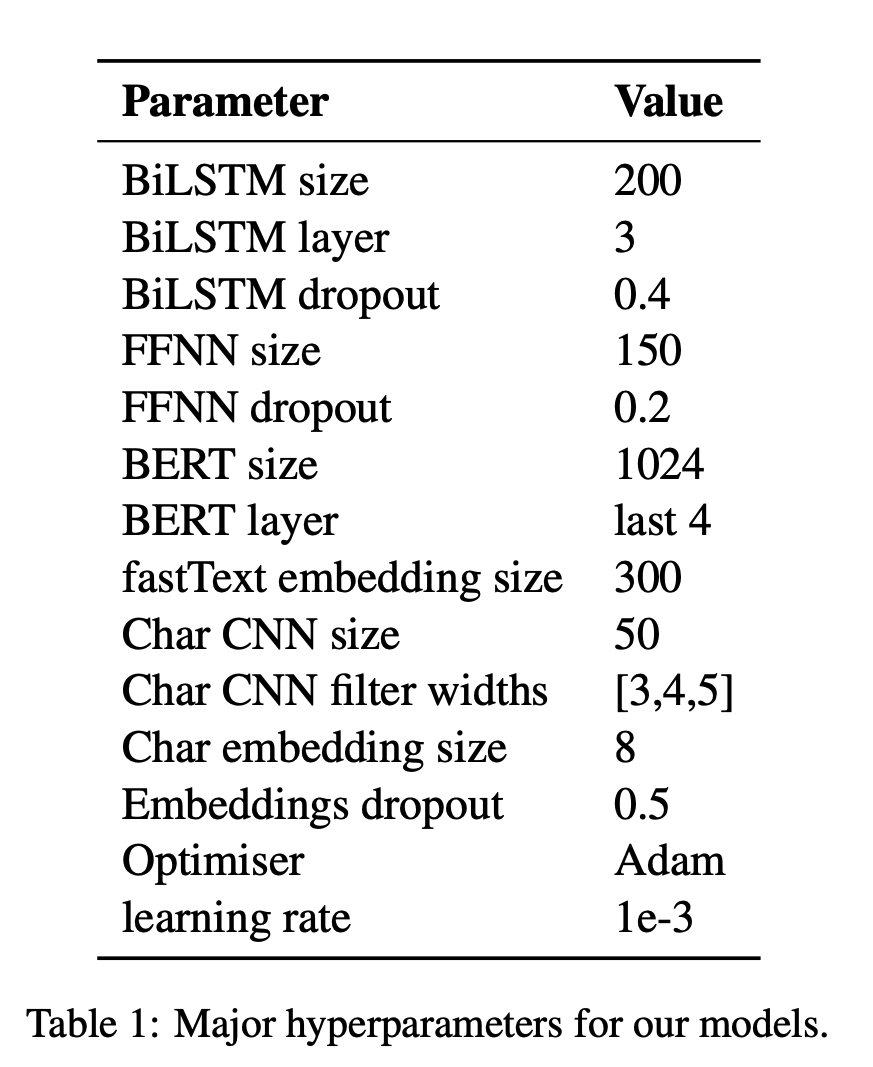


1. 实验结果

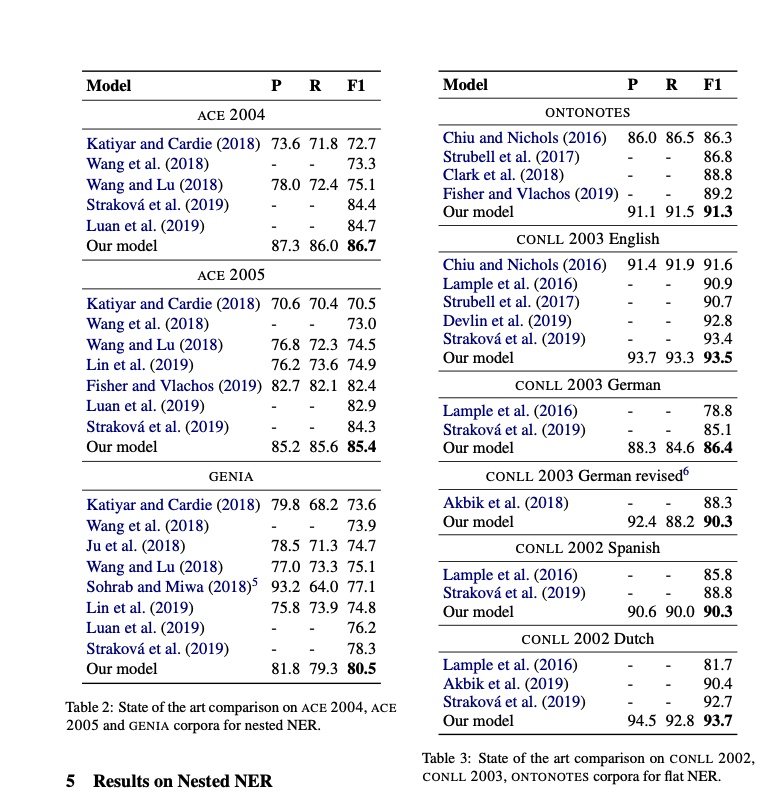
数据集：嵌套实体，ACE2004，ACE2005，GENIA；平面实体，CONLL2002、CONLL2003、ontonotes。

评估指标：recall、precision、F1

参数设置：



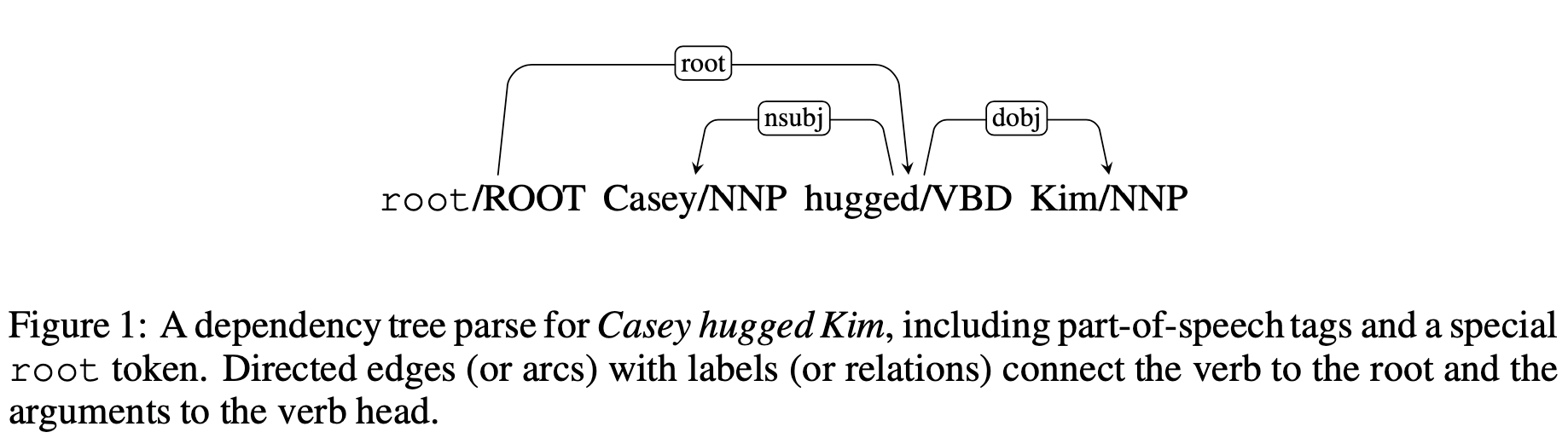
不同数据集上的表现：



论文阅读笔记二

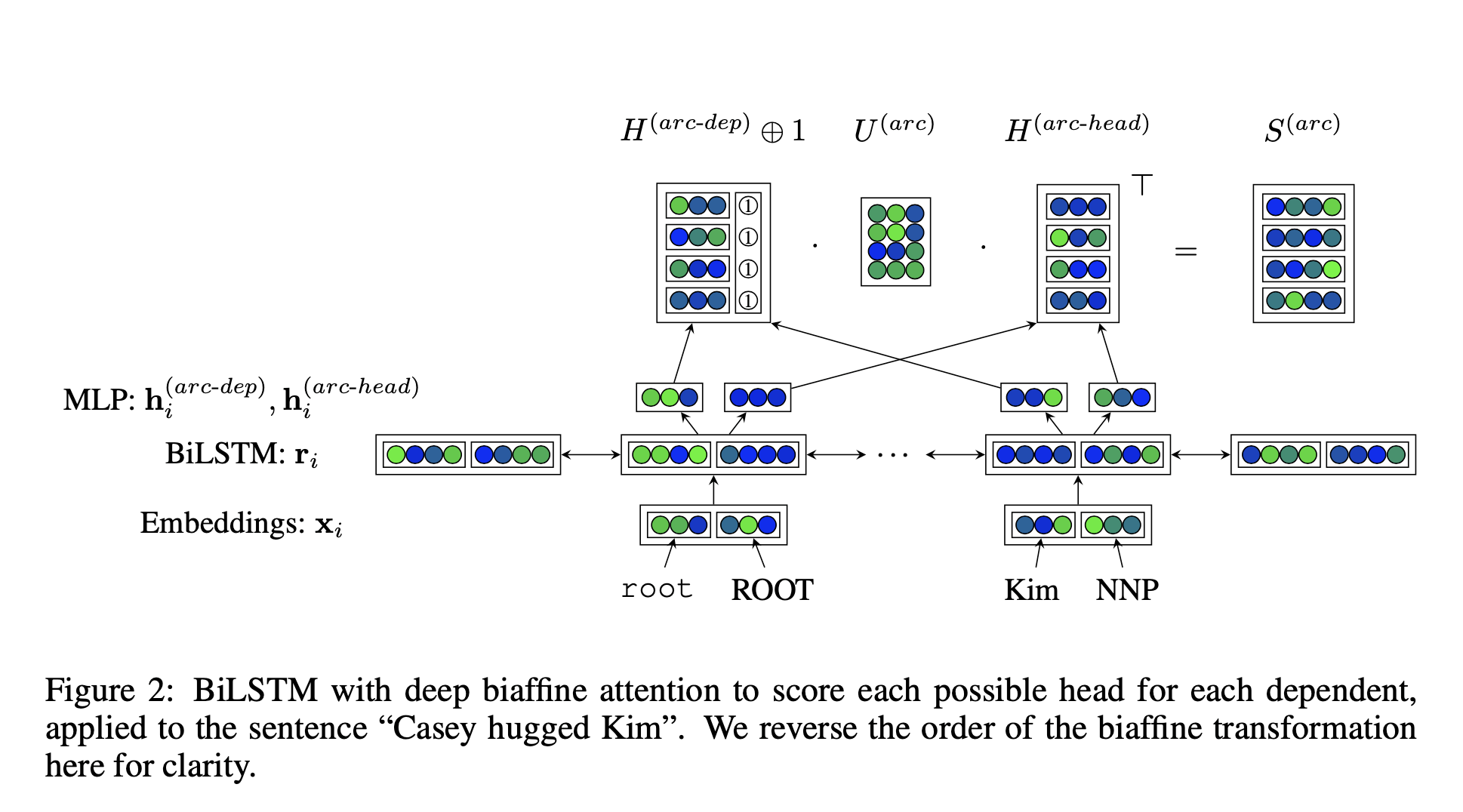
Deep Biaffine Attention for Dependency Parsing

1. 预备知识
2. 句法分析（syntactic parsing）又称短语结构分析（phrase structure parsing）、成分句法分析（constituent syntactic parsing），作用是分析输入的句子获取其句法结构。句法分析包括识别句子中短语结构及短语之间的层次语法分析（syntactic structure parsing），识别句子中词与词之前的相互依存关系的依存分析（dependency parsing），利用词汇化树邻接文法（Lexicalized Tree Adjoining Grammar, LATG）深层的句法及与语义分析.
3. 依存分析是指将句子分析成一棵依存句法树，描述各个词语之间的依存关系。处于支配地位的词成为支配者（governor，regent，head），而处于被支配地位的词被称为从属者（modifier， subordinate，dependency）。依存关系连接的两个词分别是核心词（head）和依存词（dependent）。依存关系由一个有向弧表示，由从属词指向支配词。如：

****

依存分析主要有transition-based和graph-based两种方法。

1. 基于图的依存语法分析



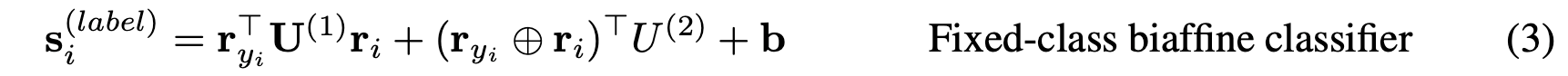
基于图的依存句法分析从左到右解析句子，找到核心词head和依存词dependent之间的关系以及关系类型，也就是依存弧及弧的标签。模型针对图的依存语法分析进行修改：

1. 使用Biaffine 依存标签分类器；
2. 在双仿射变换（Biaffine transformation）之前将降维MLP应用于每个循环输出。
3. Biaffine使用一个仿射变换在单个LSTM输出状态ri上进行得分预测。双仿射注意力机制Biaffine Attention可看作一个传统的仿射分类器（公式1），但是对于Stack LSTM的输出RU(1)进行d×d的线性变换，而不是用权重矩阵W，同时将偏置项b用一个d×1的线性变换RU(2)来替代。



**依存弧的标签**

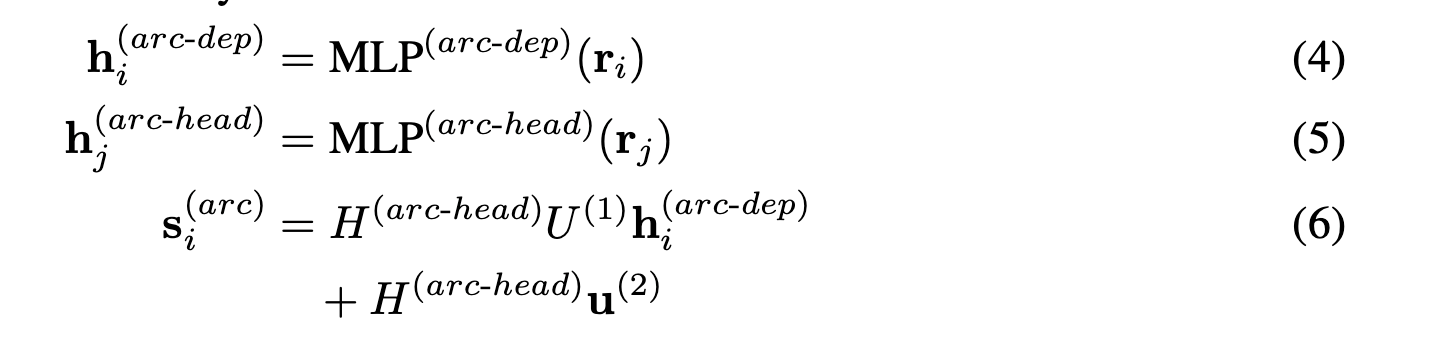
核心词和依存词之间的依存关系数目是确定的。采用公式3计算标签分数：



直接对单词j在第二项收到任何依存词的先验概率和j在第一项中收到特定依存项i之间进行建模。用双仿射分类器来预测给定的核心词head或预测对应的依存标签。假设总共m个标签，U(1)是m×d×d的高阶张量，ri是第i个词在BiLSTM的输出向量表示（大小为d×1），yi是第i个词head，ryi对应的是其BiLSTM的向量表示（大小为d×1）。

**依存弧的得分**

对于一个由N个词组成的句子，包含虚根root在内共d=N+1个单词，每个词都会有一个分数，故这是一个不定类别分类问题，一般MLP是个固定类别的分类器（如公式1），故本文采用两个MLP对BiLSTM隐层的输出向量进行重新编码（公式4和公式5），然后套用公式2得到公式6，得到分数，两个MLP分别针对核心词head和依存词dependent，MLP降维后的向量表示能够去除多余信息，比如除掉BiLSTM隐层中的预测依存弧标签的信息。



1. 实验结果评估

数据集：English Penn Treebank，PTB-SD，CoNLL09，Chinese Penn Treebank

