**论文阅读笔记一**

Deep Exhaustive Model for Nested Named Entity Recognition

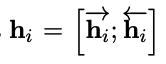
1. Motivation
2. 识别语料中的嵌套实体
3. Contribution
4. 枚举文本中的所有可能为实体的跨度
5. 用深度神经网络对实体进行分类
6. 为减少计算复杂度以及获取跨度的上下文信息，用LSTM的输出作为共享的跨度表示。
7. 实现neural Exhaustive Model的Method

模型用单个神经网络来穷举句子中所有可能的实体区域，模型使用一个共享的LSTM层来进行编码，然后用LSTM的输出来表示实体区域，从这些区域中检测可能的实体，区域的长度取决于预定义的长度L。

1. 单词表示

将单词和字符级的向量表示进行拼接，然后将由字符组成的单词的词向量送到LSTM中得到前向和后向的隐藏层向量，将前向和后向向量进行拼接得到单词的向量表示。

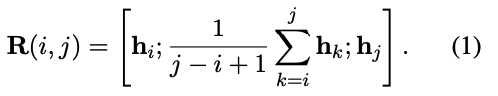
1. 用LSTM进行穷举组合

给定句子X= {x1, x2, ...xn}，i表示第i个单词，n表示句子长度，离散的单词用LSTM获取上下文信息，得到；

通过LSTM的输出hi，穷举模型通过穷举组合共享所有可能区域的表示。生成最大长度为L的所有可能的实体区域，用region（i,j）表示从第i个词到第j个词组成的区域，其中1≤i＜j≤L，L为最大实体长度。

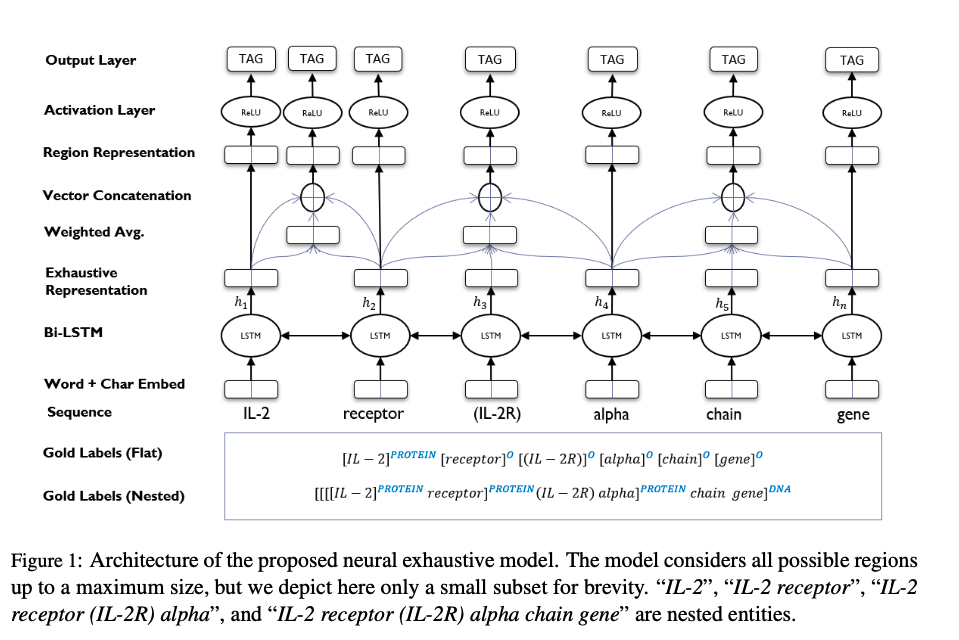
1. 跨度表示和分类

将区域的表示分为边界表示和内部表示，边界表示仅取决于LSTM的输出，内部表示则取LSTM输出的均值。区域region(i,j)的表示R(i,j)如下：



将每个区域表示作为激活函数的输入送到ReLU中，将激活函数的输出传到softmax输出层，将区域分类为实体或非实体。

该模型对所有可能的区域进行穷举和分类，每个句子分类众数为O(lmn)，L为句子中的单词总数，m是最大的实体长度，n是可能的实体类型总数。



该模型对每个区域进行分类，可以直接合并短语级词典信息，并且能像CRF那样为每种类型调整偏差。

1. 优点
2. 有效识别嵌套实体
3. 枚举的方式，正确率高
4. Idea
5. LSTM编码
6. Hypergraph+LSTM编码训练
7. CRF解码