论文阅读笔记一

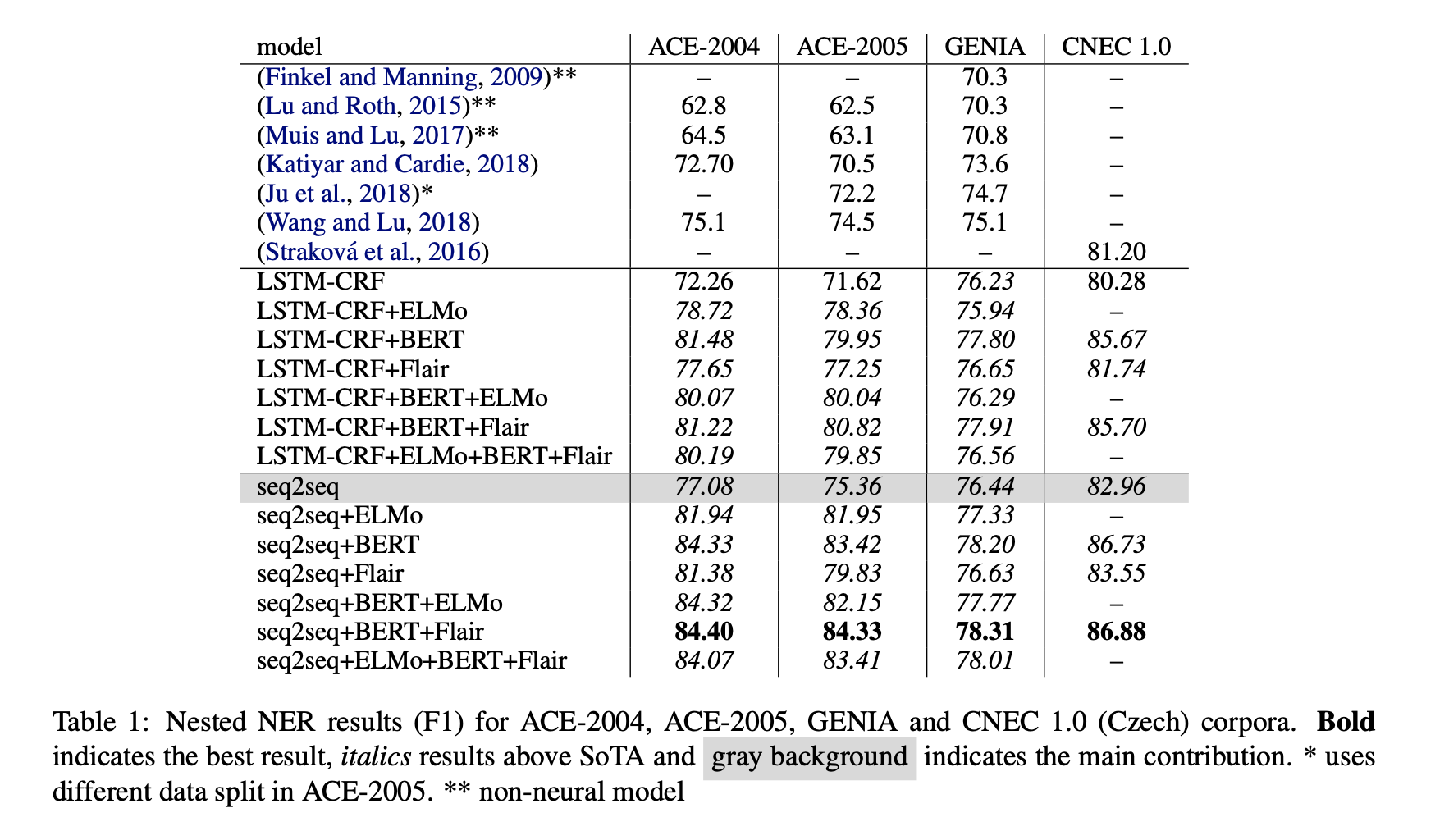
Neural Architectures for Nested NER through Linerization

1. 主要工作
2. 第一个模型将嵌套实体的多个标签连接为一个标签，使用BiLSTM-CRF进行预测。
3. 第二个模型将嵌套实体按序列进行编码，然后将任务视为seq2seq，输入为token形式，输出标签，解码器为每个token预测标签，直接预测到特殊标签<eow>才转移到下一个token。采用一个微启发式算法按照 顺序匹配连续单词的标签。如果I-或是L-标签出现在相邻的多标签中的相同位置并且具有相同的标签类型，则他们会与之前的B-或I-合并。
4. 方法

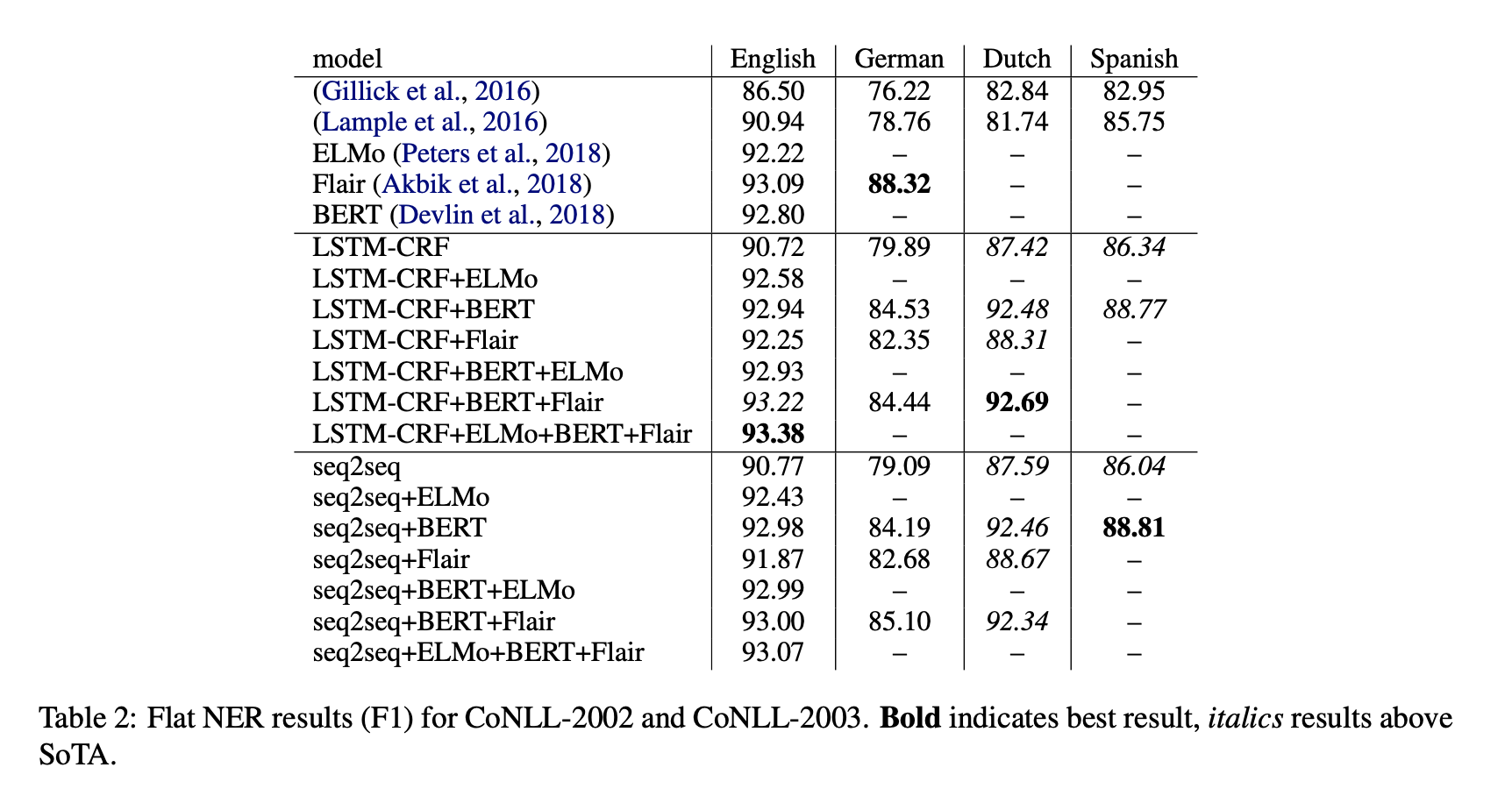
BILOU标注：B- (beginning), I- (inside), U- (unit-length entity), L- (last) or O (outside) labels

从token到多标签：1)较早开始的实体优先于较晚开始的实体；2)起始词相同的实体，较长的实体优先于较短的实体。

1. LSTM-CRF：编码器BiLSTM -解码器CRF，建模多标签。
2. Seq2seq：编码器BiLSTM-解码器单向LSTM。输入Token序列，一次输出标签，直到遇到<eow>然后转到下一个token，特别关注那些带预测的单词，并使用长度从高到低和起始词优先的策略进行预测。
3. 使用Adam 优化器的惰性变体训练网络，仅更新当前批次中出现的变量累加器。参数β1= 0.9和β2= 0.98。使用大小为8的微批处理，用0.5的比率进行dropout，并且用未知token替换20%的单词。
4. Baseline中的词嵌入：300维的Word2Vec，128维的Bidirectional GRU，上下文词嵌入ElMo、BERT、Flair
5. 实验结果和分析



Seq2seq的结果明显优于已知的方法，seq2seq模型会更复杂，也适合复杂的语料库，在ACE04和ACE05语料库中能明显看到收益。



实验结果显示，在输入中加入经过预训练的上下文词嵌入时F1分数会大大提高。

论文阅读笔记二

Chinese NER Using Lattice LSTM

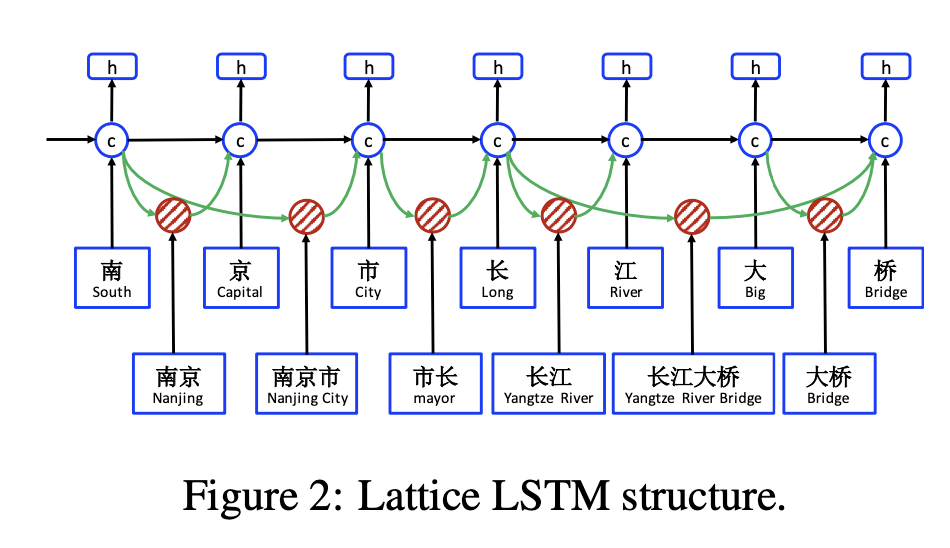
1. 贡献

引入中文分词信息避免分词错误

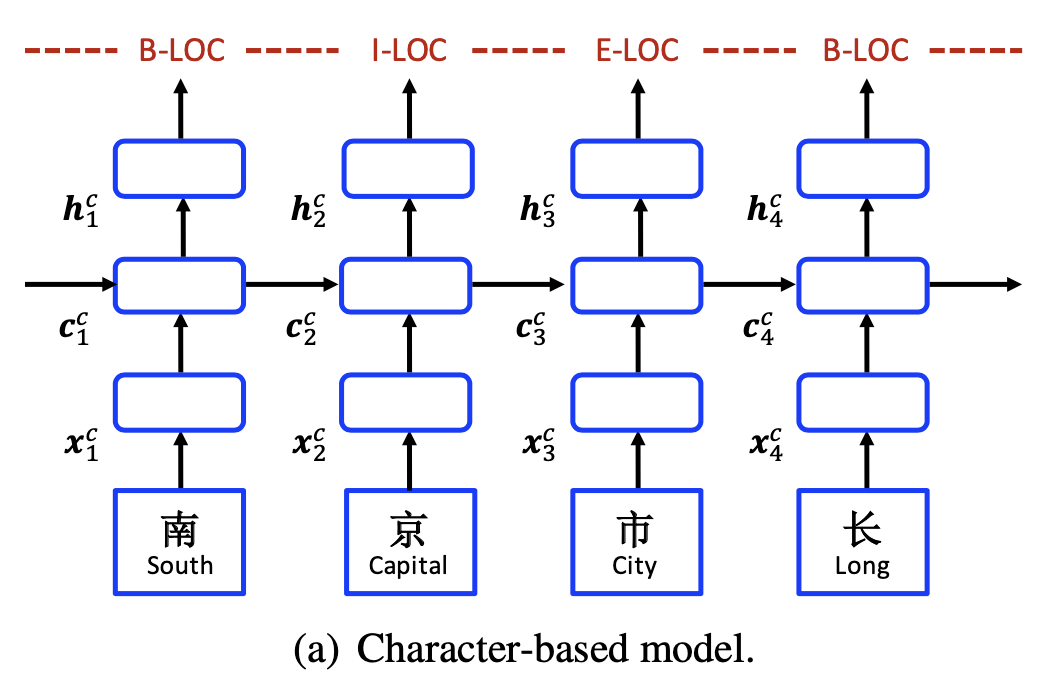
若当前字符与之前字符构成词汇，则从词汇中提取信息，联合更新记忆状态。

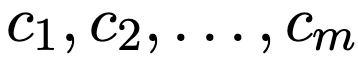
1. 方法

模型引入了分词信息对单个中文字符和中文词组进行建模。



1. 基于字符的模型

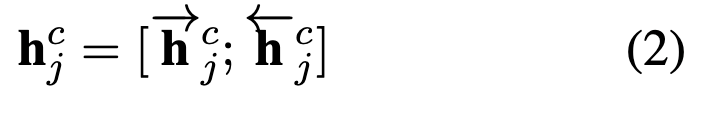


1. 对字符序列，每个字符同通过LSTM-CRF表示为向量：

首先从词嵌入表中查询到字符的词嵌入，其中ec表示字符的词嵌入表：



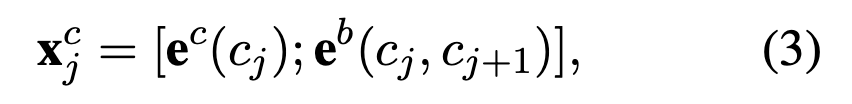
然后用BiLSTM进行训练：



CRF用于序列标注。

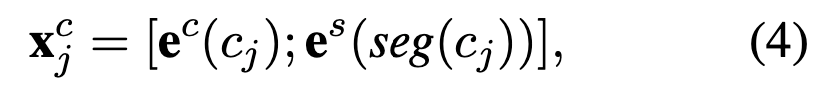
1. 字符+双字符组

将双字符（词语）与字符嵌入拼接起来扩充基于字符的模型，其中eb表示双字符组：

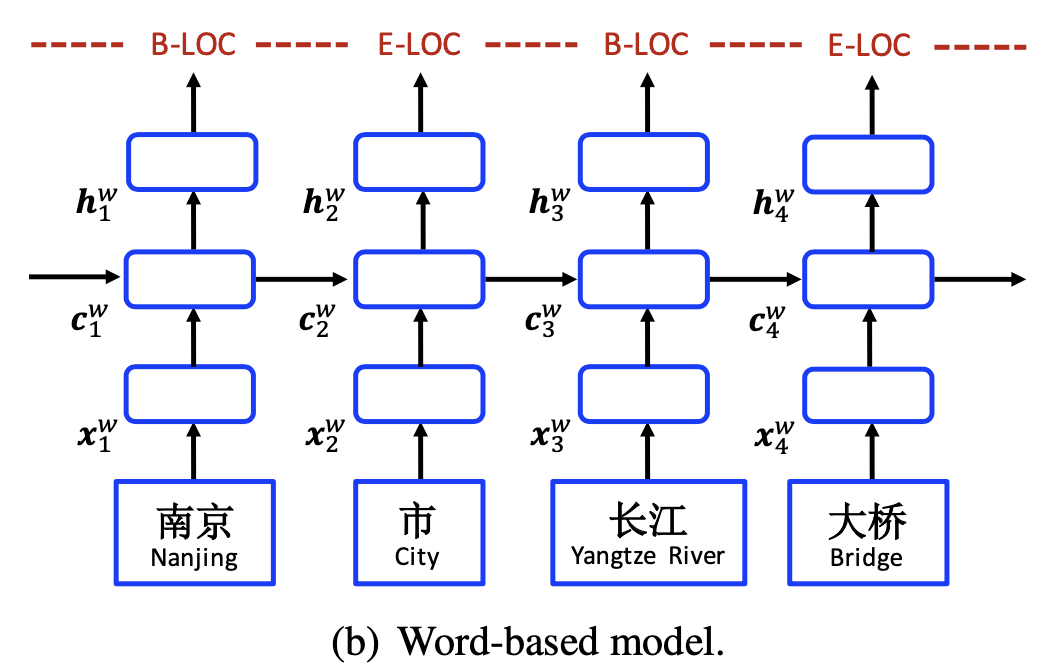


1. 字符+softword

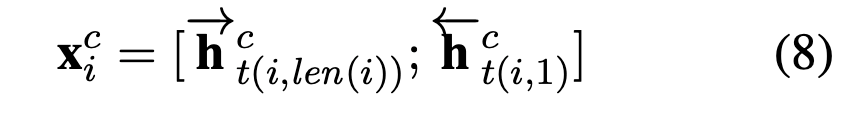
将分段信息作为字符模型的软特征，将分段标签嵌入和字符嵌入连接，用分段信息来增强字符表示，其中es表示段标签编码词嵌入表：



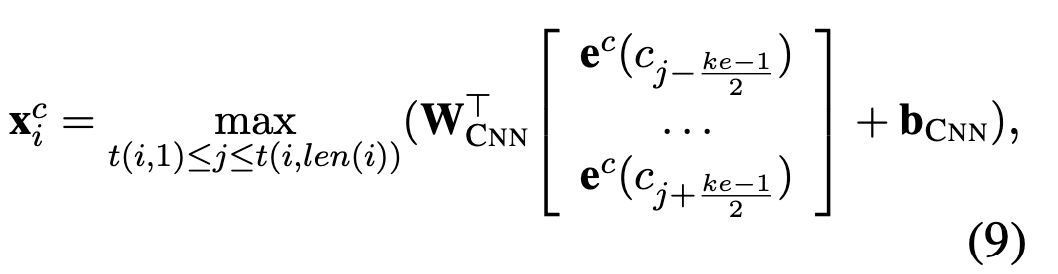
1. 基于单词的模型



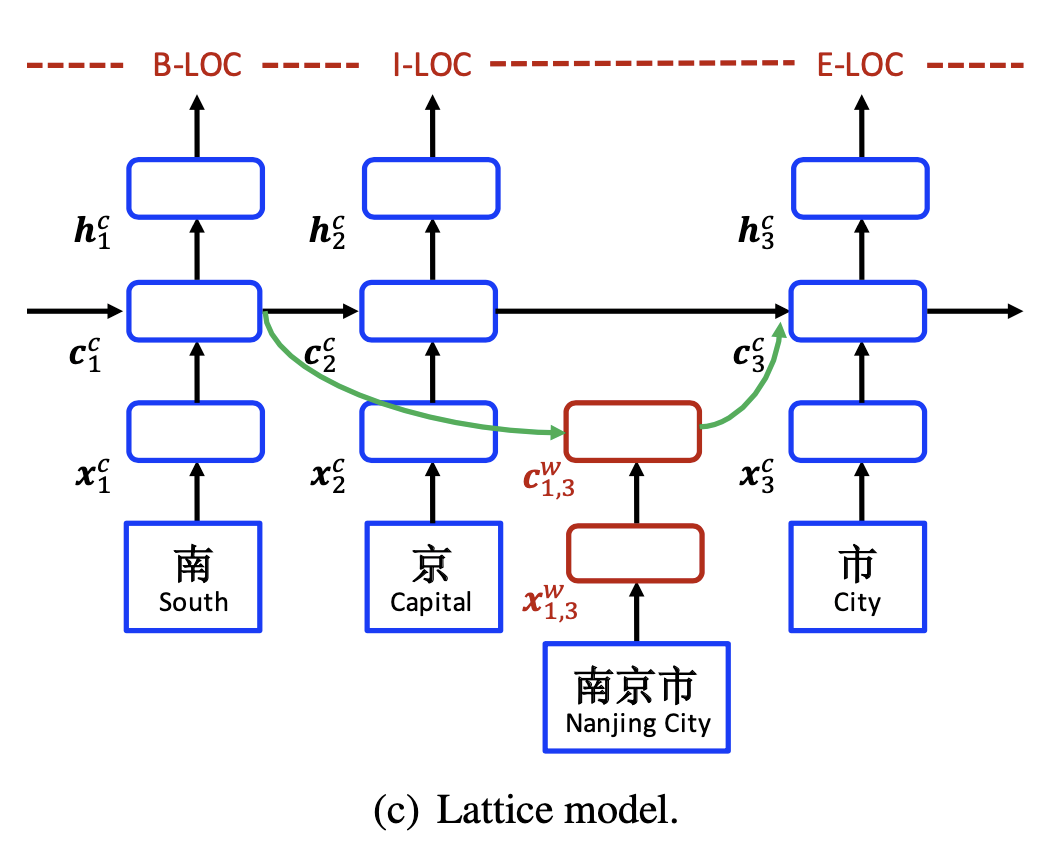
1. 词语和LSTM字符词嵌入，其中下标len(i)表示第i个词的字符的个数：



1. 词语和CNN字符词嵌入，其中WCNN和bCNN是参数，ke=3是内核的大小，max是最大池的大小。



1. Lattice模型

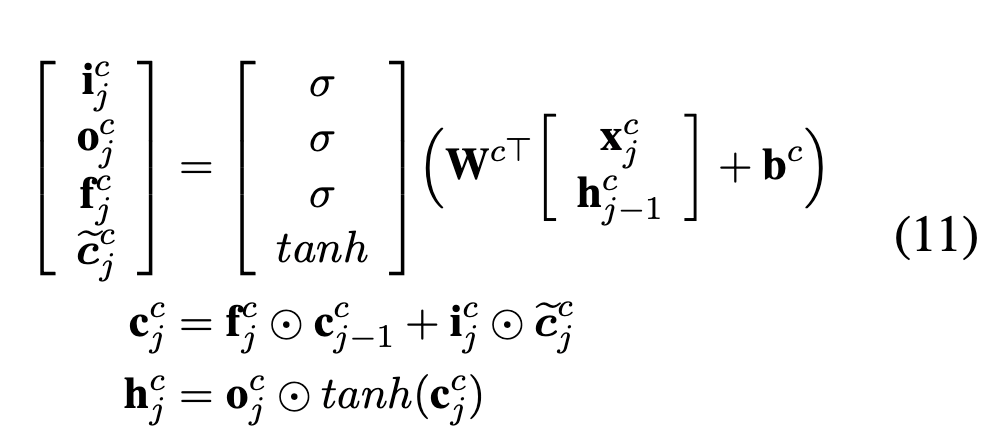


模型的输入由四个部分组成，分别是输入向量、输出隐藏层向量、细胞向量和门向量。

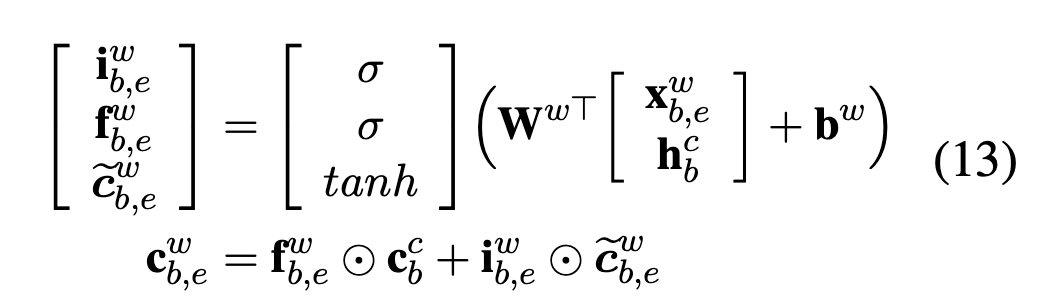
输入向量为：



LSTM提取的字符特征：

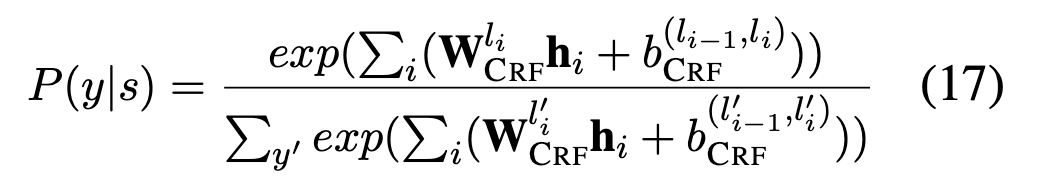


CNN提取的字符特征：

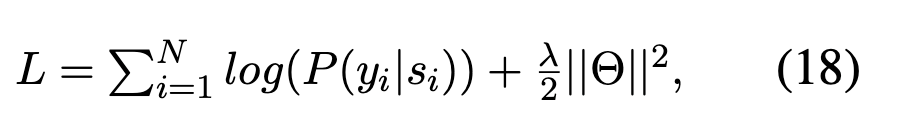


1. 解码和训练

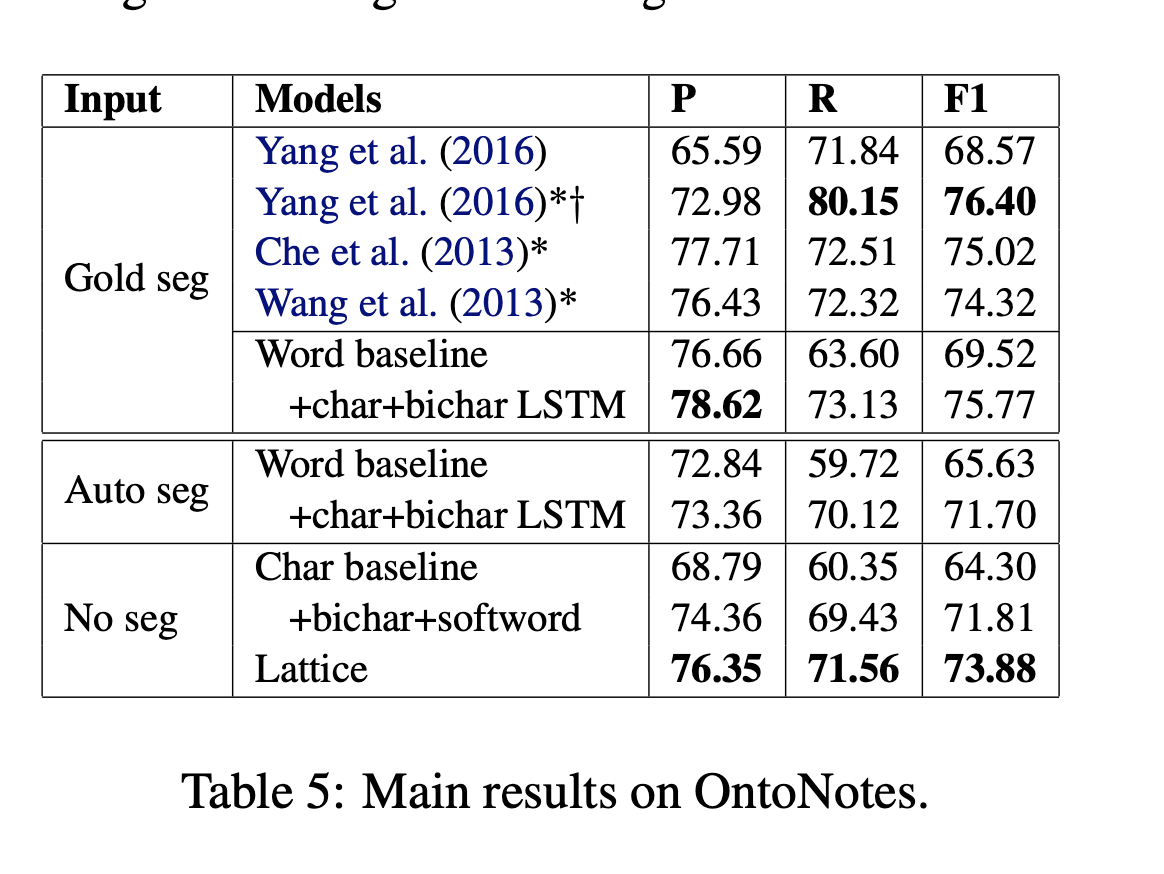
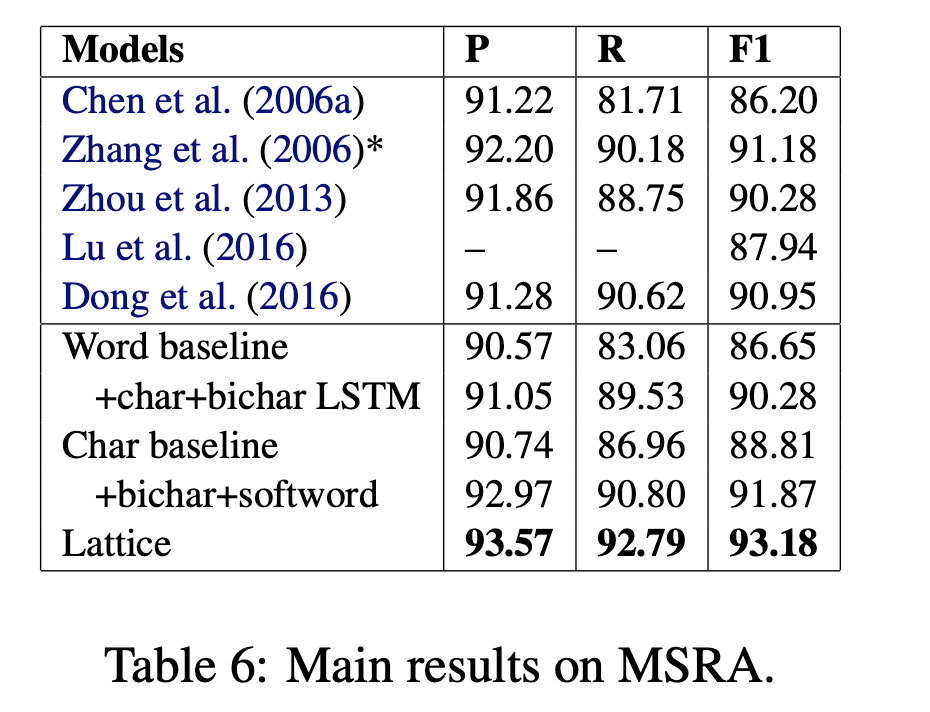
解码阶段使用CRF：

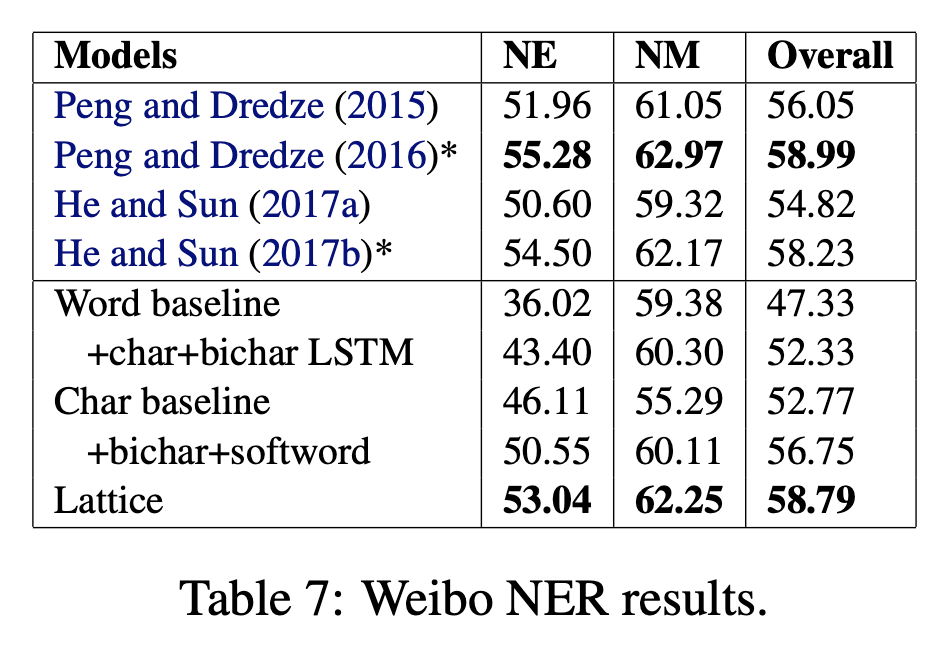
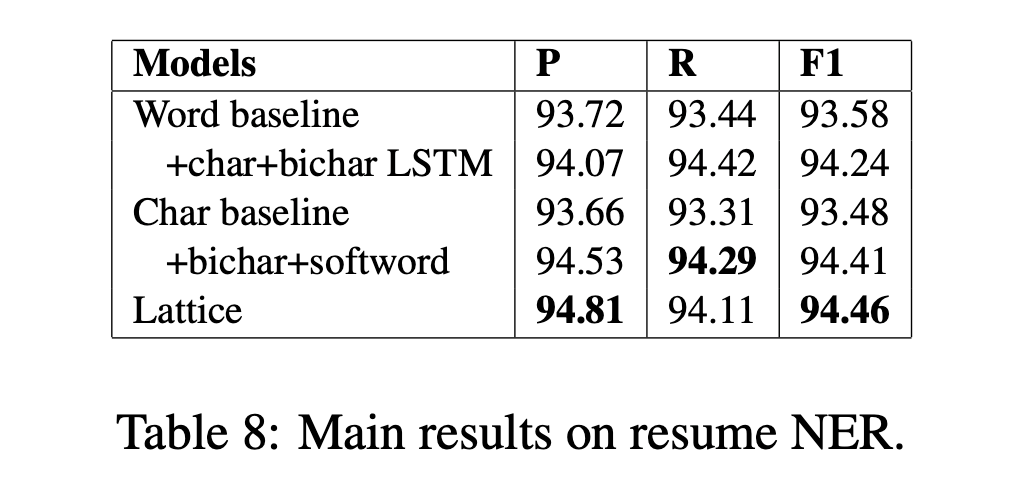


加入L2正则化的损失函数：

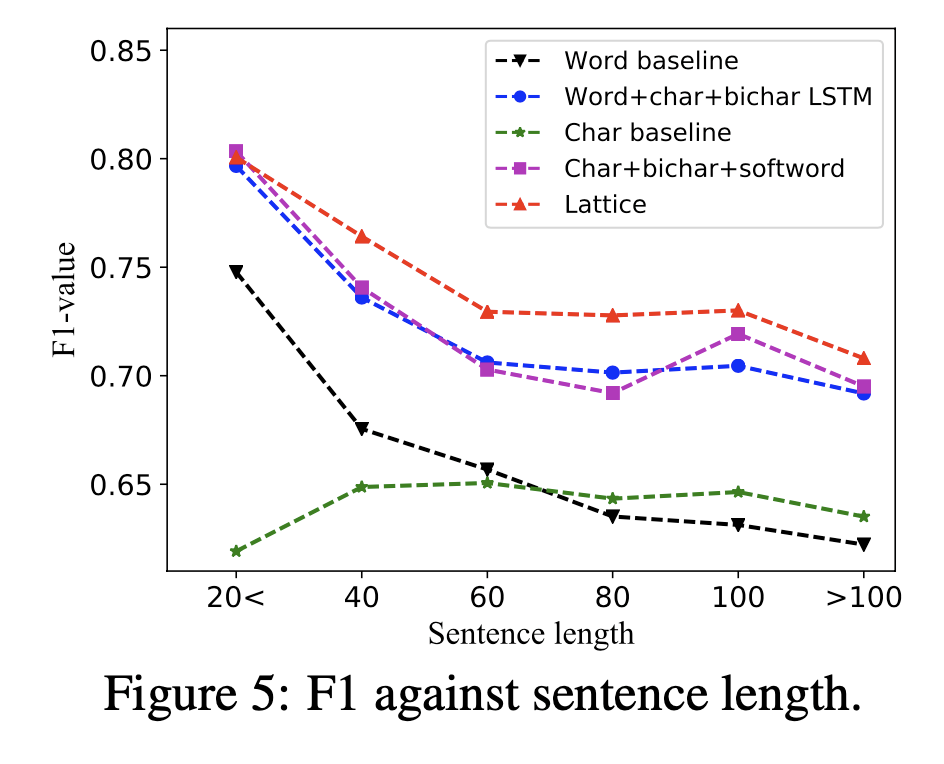


1. 实验结果

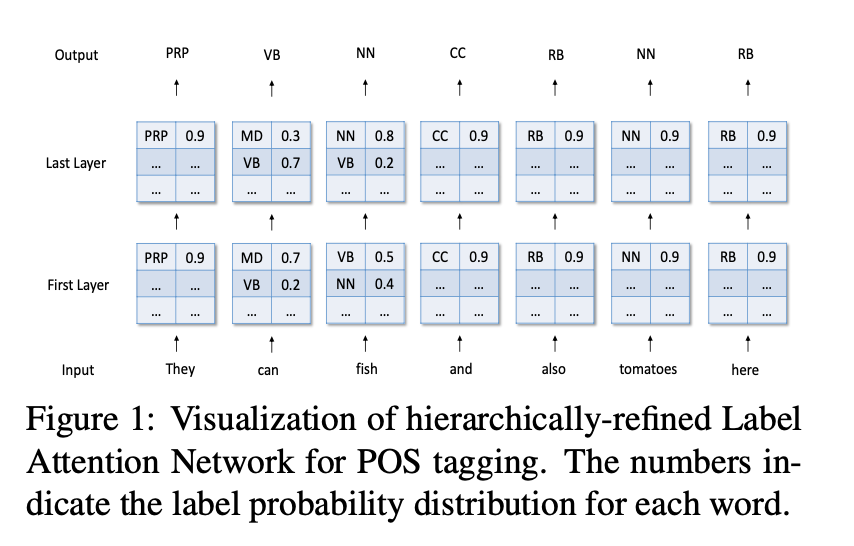
F1分数随句子长度变化的变化：



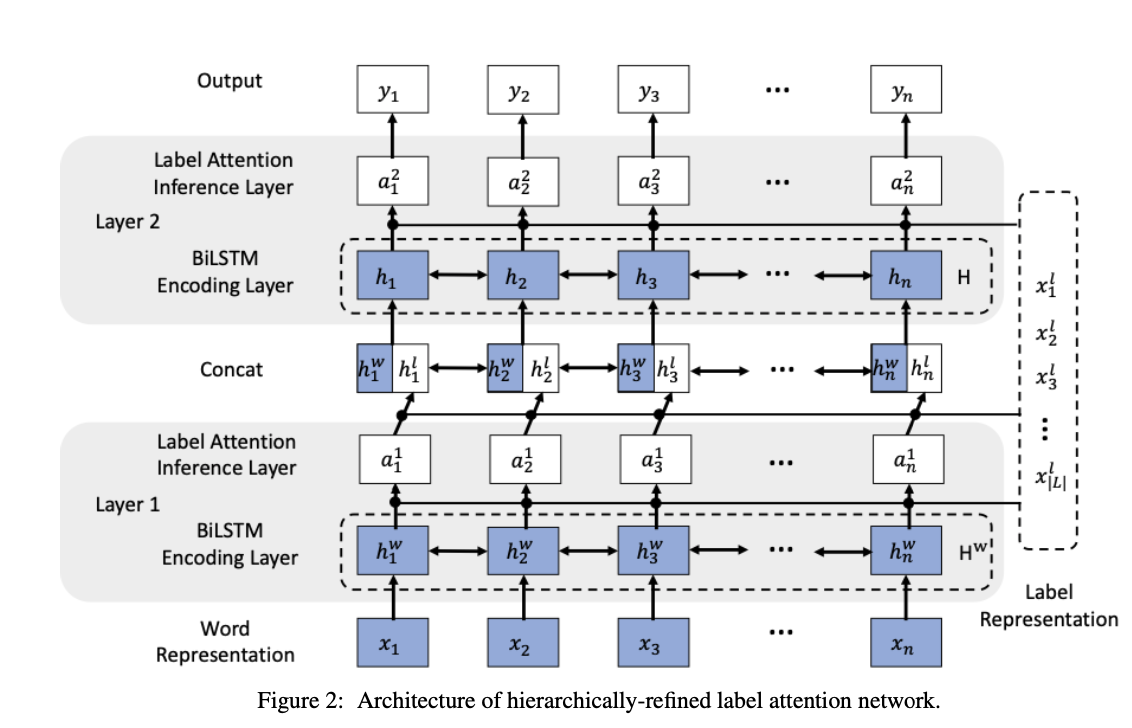
论文阅读笔记三

Hierarchically-Refined Label Attention Network for Sequence Labeling

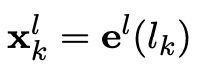
1. 贡献
2. 提出分级改进的标签注意力网络，显示地利用标签嵌入。
3. 给予每个单词逐步精细化的标签分布来捕获长序列标签依赖性。
4. 构建了一个用于序列表示分层的注意力神经网络，主要研究标签序列编码。使用动态精炼的序列表示网络对深层标签交互进行建模。

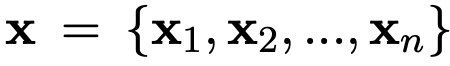


1. 方法：标签注意力网络（label attention network）



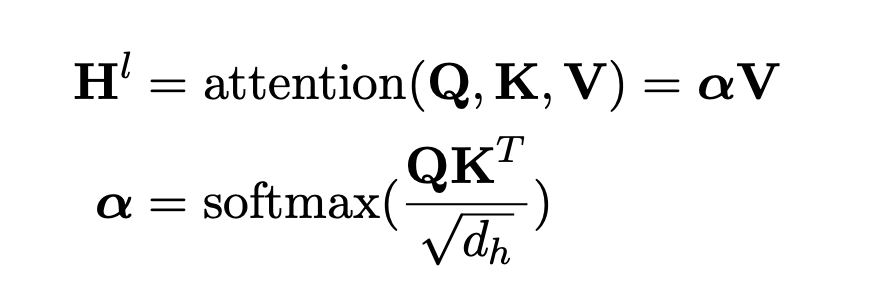
1. 标签表示：将每个标签都用词向量进行表示，

其中，el是标签词嵌入表，标签嵌入被随机初始话，在模型训练过程中更新。

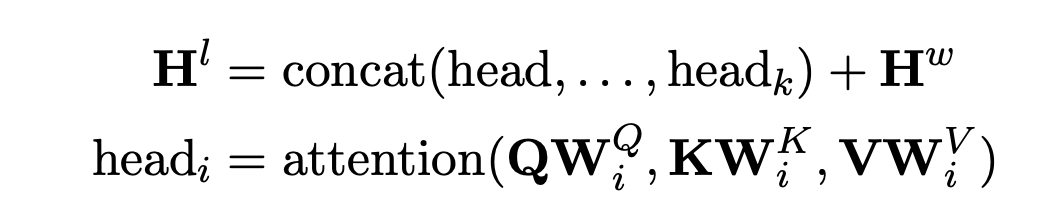
将输入表示为，BiLSTM用于计算隐藏层向量Hw∈Rn×dh，其中n和dh分别表示单词序列长度和BiLSTM隐藏大小（与标签嵌入尺寸相同）。

1. BiLSTM-LAN层：BiLSTM编码自层和标签注意力推理子层。前者用BiLSTM对单词进行编码，后者用多头注意力机制联合编码单词和标签词嵌入。对每个注意力头，模型都用一个缩放因子进行点积注意力的计算。

定义Q=Hw，K=V=xl，xl∈R|L|×dh 是标签集的表示，|L|是总的标签数目，标签层的输出为：



模型使用了多头注意力机制：



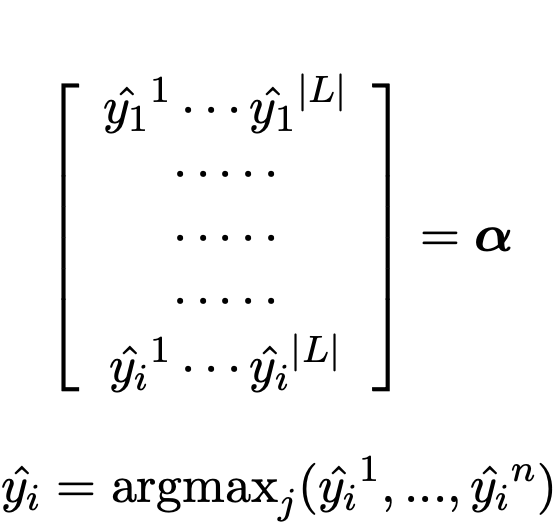
其中∈，∈,∈是待学习的参数。

最终的BILSTM-LAN层表示是隐藏层状态和注意力输出的拼接：



H是BILSTM-LAN层的输入。

模型的输出层根据注意力权重直接预测每个单词的标签。



其中，表示第i个词的第j个候选标签；是第i个词的标签。

模型使用对数似然函数通过标准反向传播训练BiLSTM-LAN，是的交叉熵损失最小。

1. 实验结果

数据集：WSJ、University Dependencies、OntoNotes 5.0、CCHBank

