论文阅读笔记一

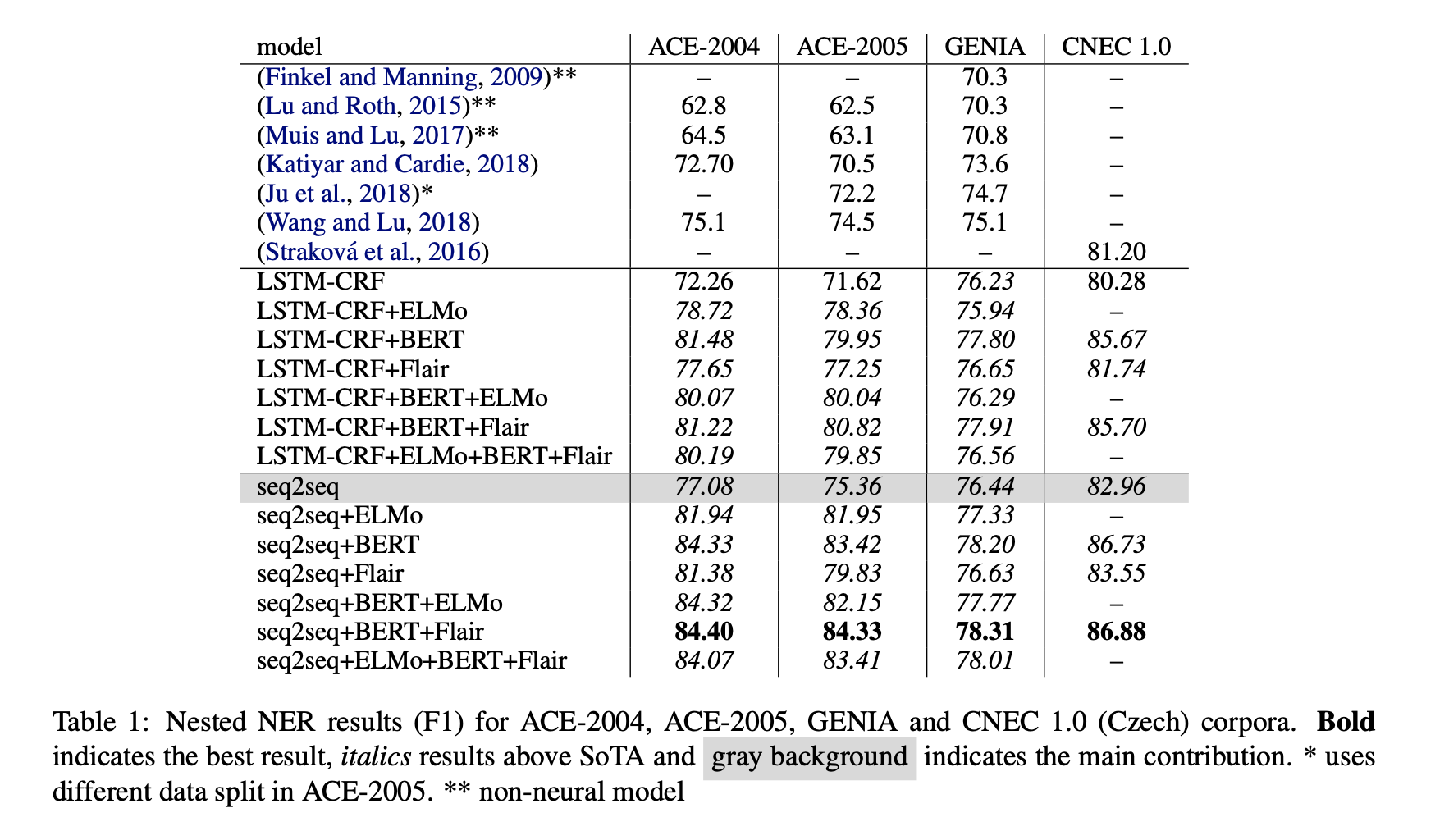
Hierarchically-Refined Label Attention Network for Sequence Labeling

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

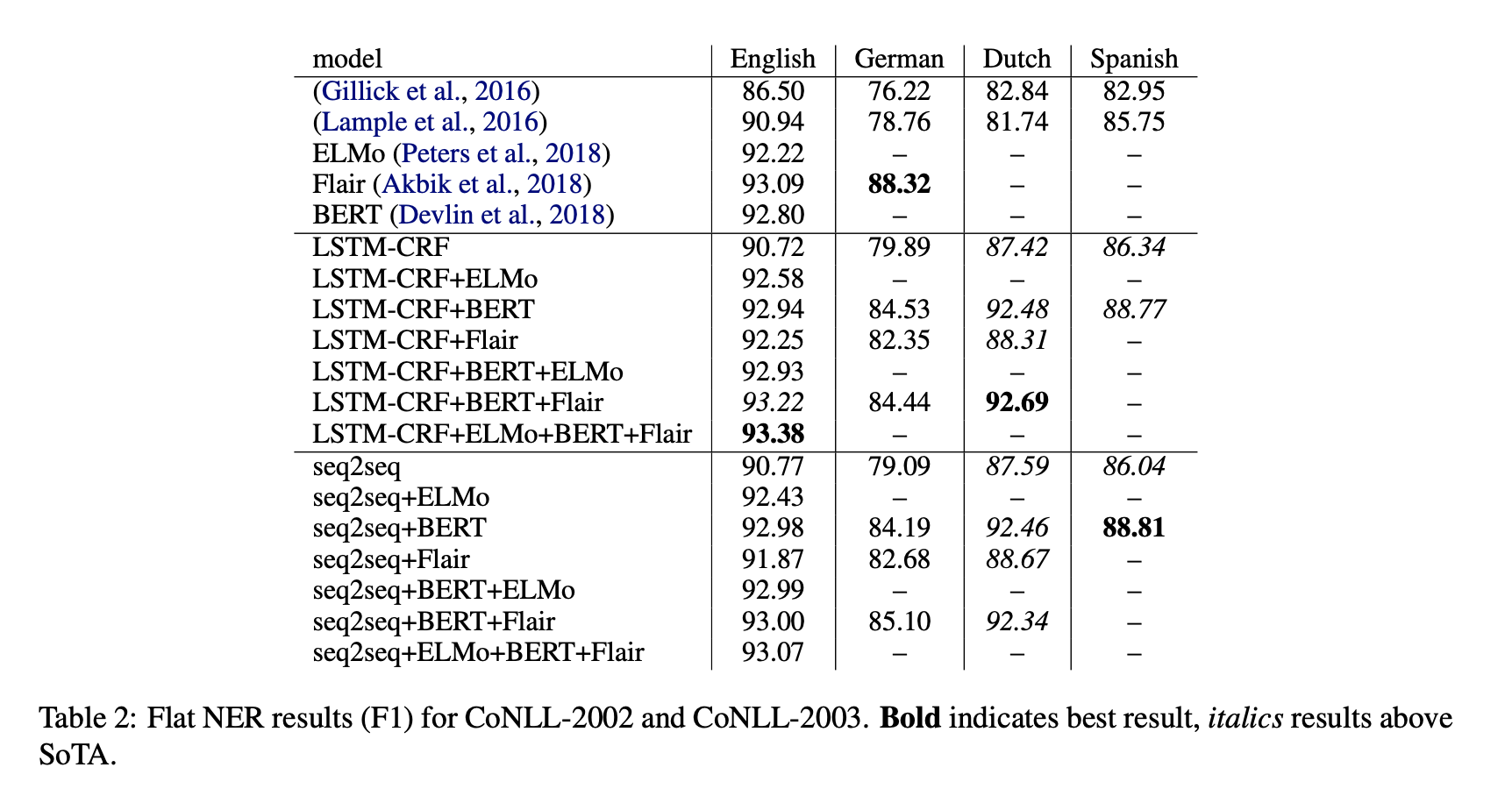
BILOU标注：B- (beginning), I- (inside), U- (unit-length entity), L- (last) or O (outside) labels

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

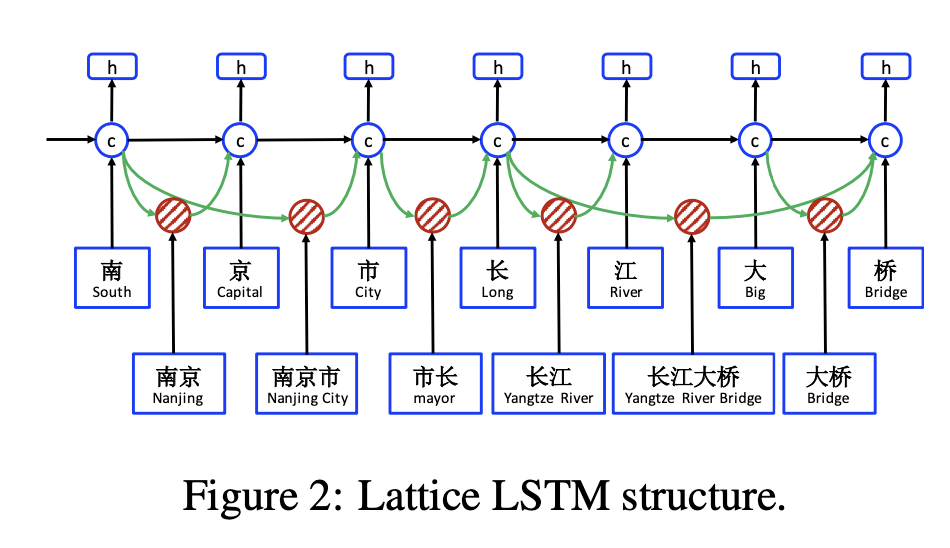
1. 贡献

引入中文分词信息避免分词错误

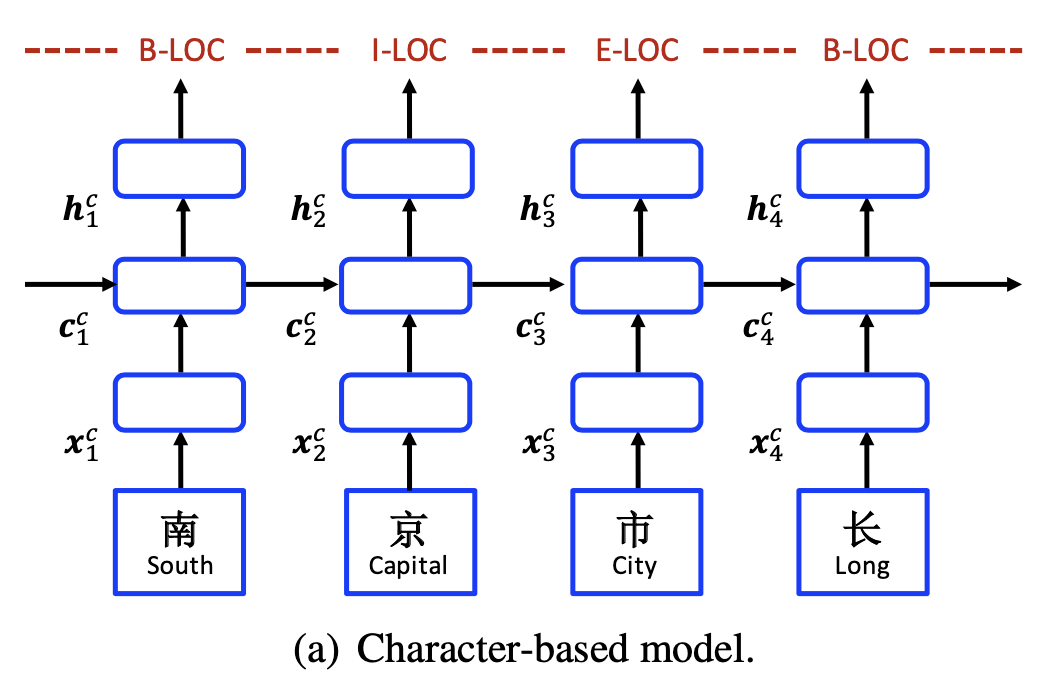
若当前字符与之前字符构成词汇，则从词汇中提取信息，联合更新记忆状态。

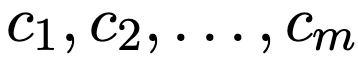
1. 方法

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

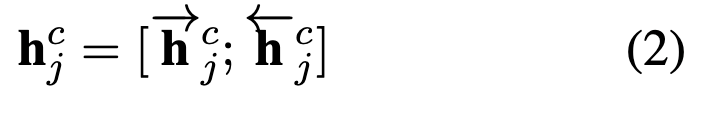


1. 对字符序列，每个字符同通过LSTM-CRF表示为向量：

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：



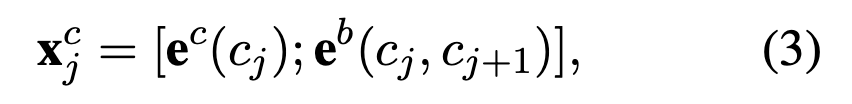
然后用BiLSTM进行训练：



CRF用于序列标注。

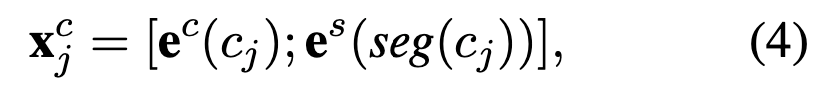
1. 字符+双字符组

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：

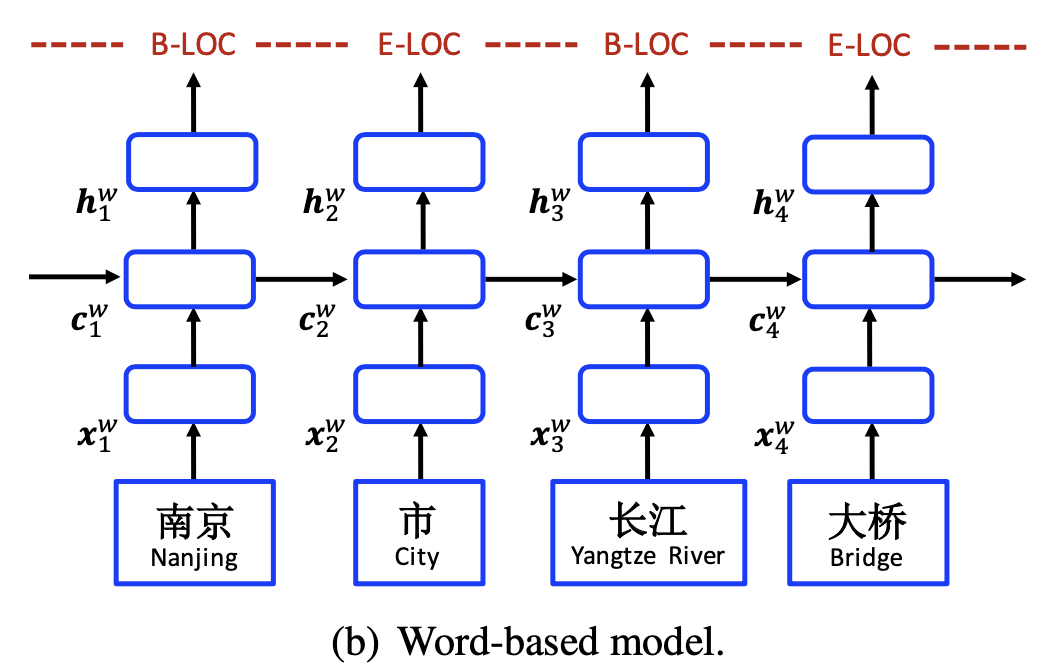


1. 字符+softword

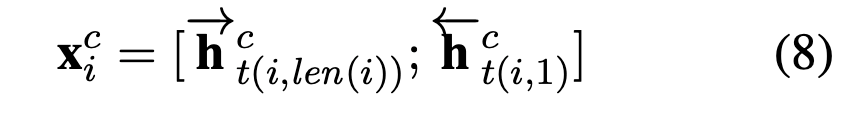
将分段信息作为字符模型的软特征，将分段标签嵌入和字符嵌入连接，用分段信息来增强字符表示，其中es表示段标签编码词嵌入表：



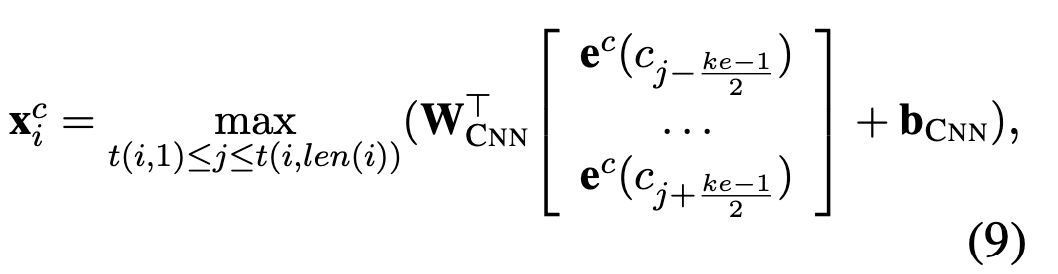
1. 基于单词的模型



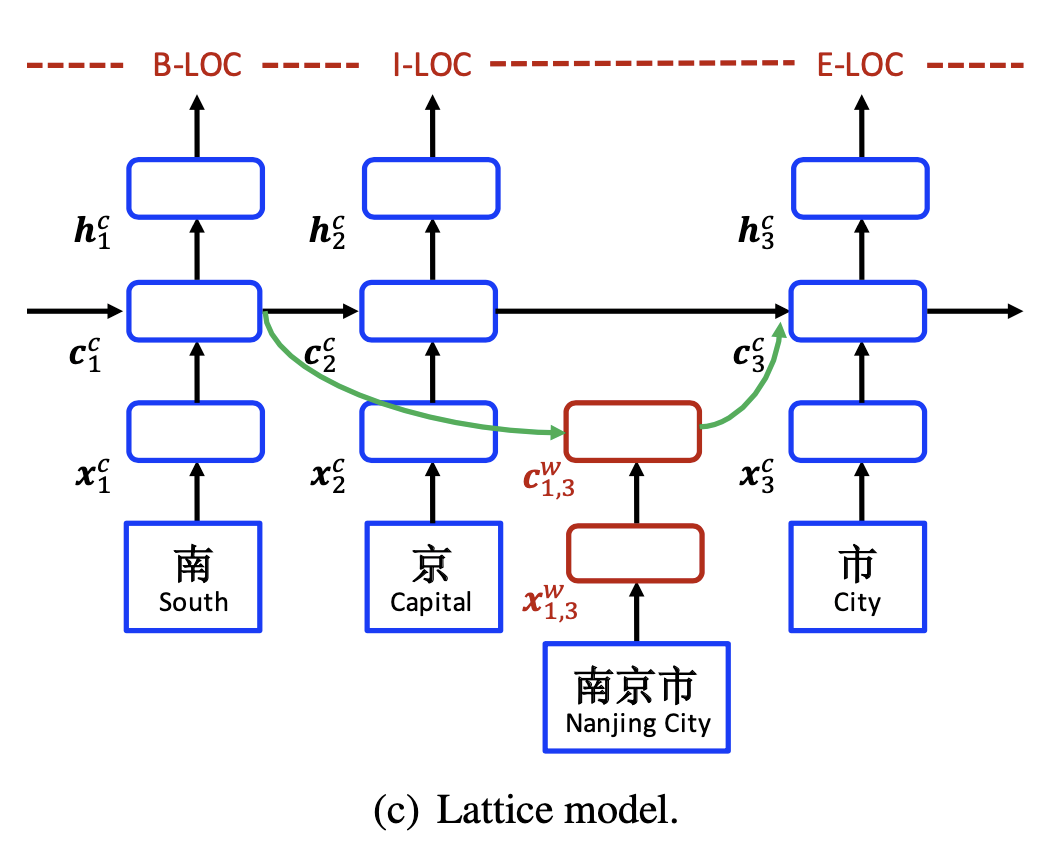
1. 词语和LSTM字符词嵌入，其中下标len(i)表示第i个词的字符的个数：



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

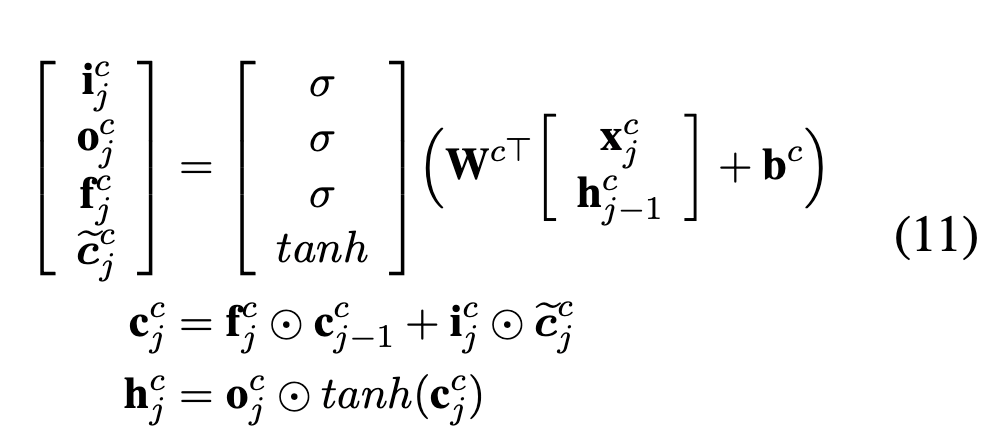


模型的输入由四个部分组成，分别是输入向量、输出隐藏层向量、细胞向量和门向量。

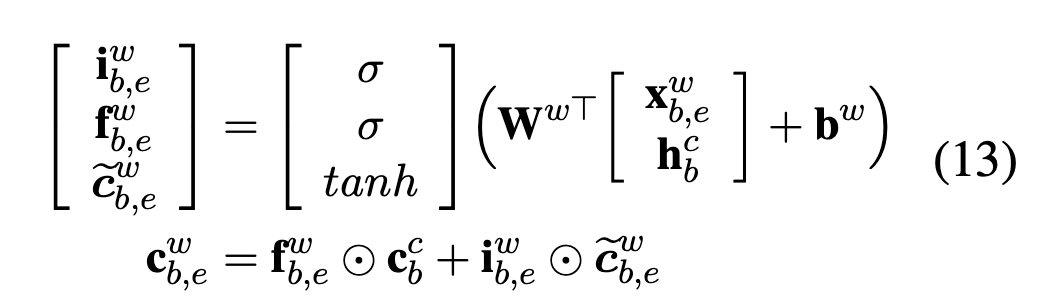
输入向量为：



LSTM提取的字符特征：

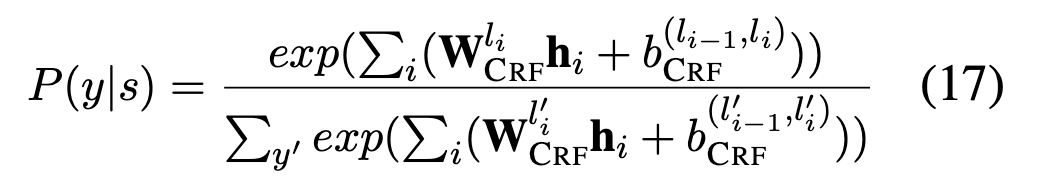


CNN提取的字符特征：

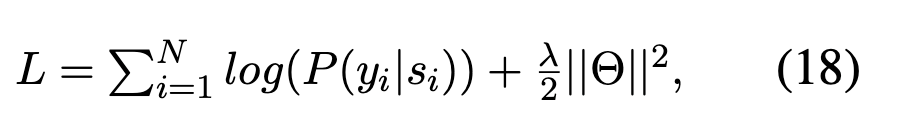


1. 解码和训练

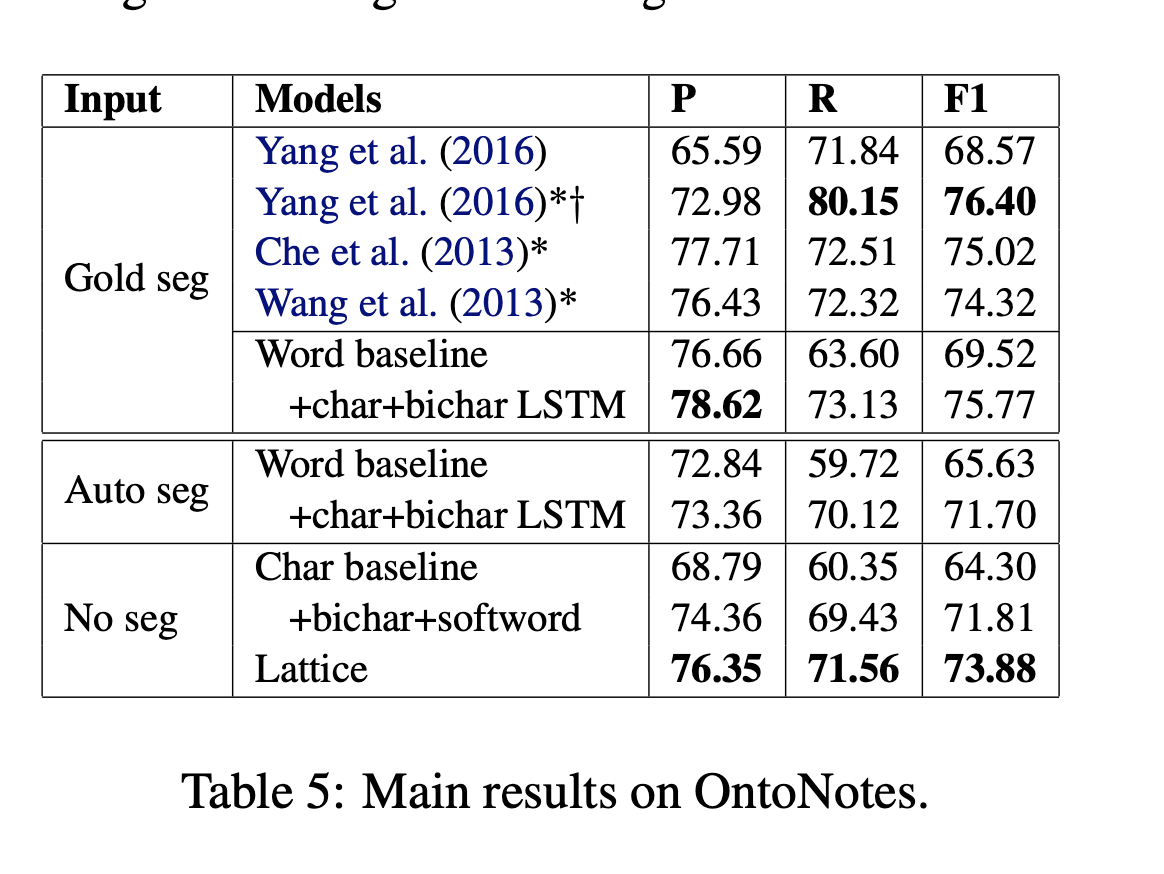
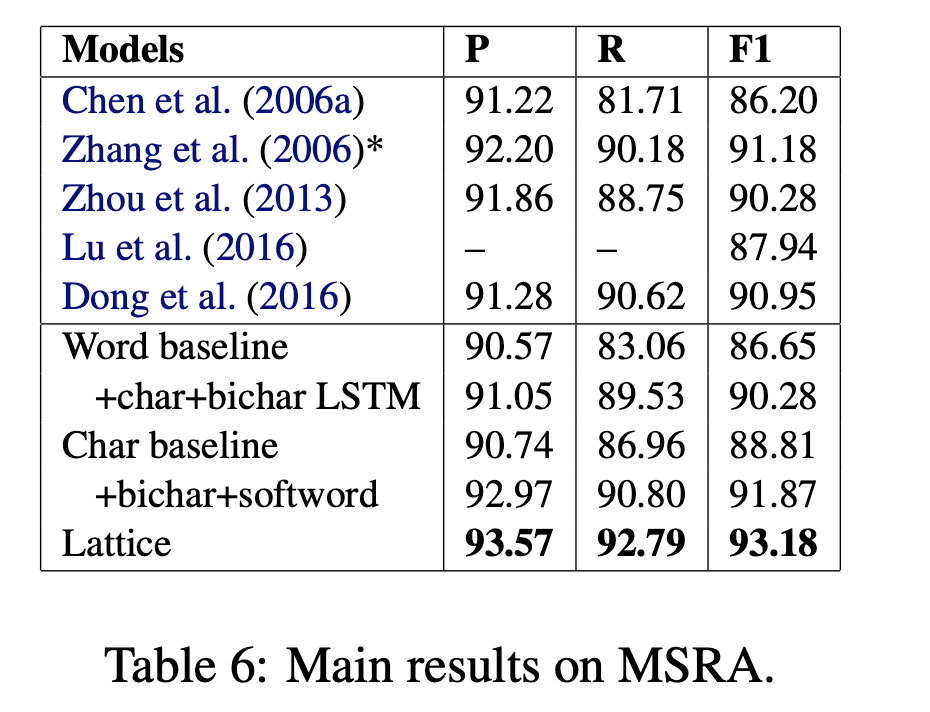
解码阶段使用CRF：

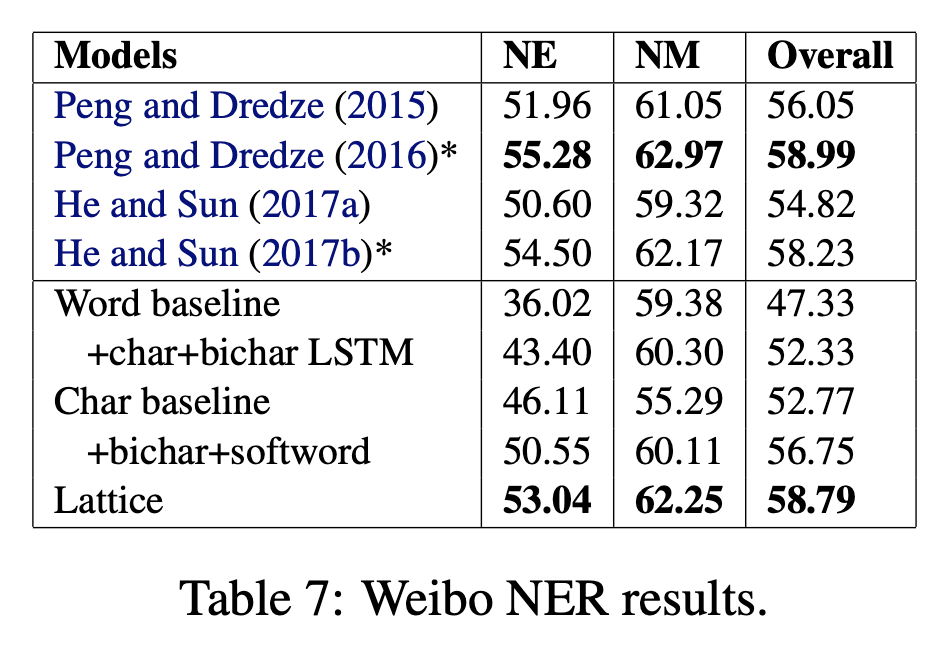
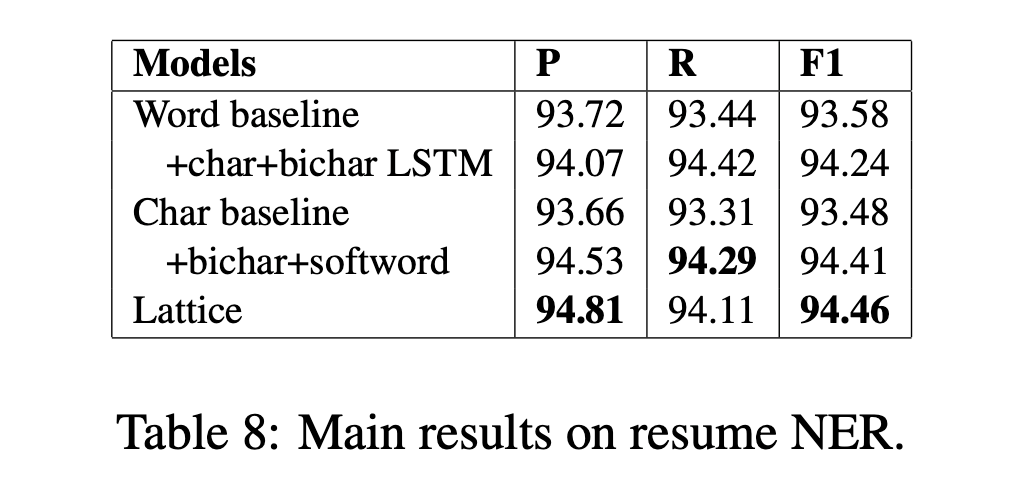


加入L2正则化的损失函数：

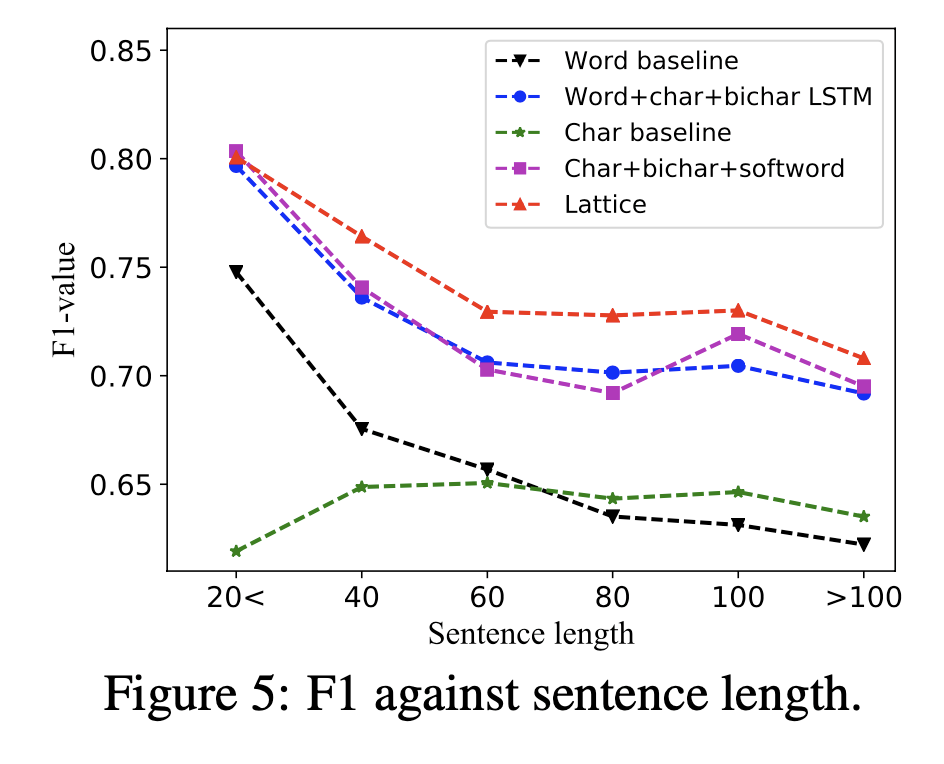


1. 实验结果

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