

以下F1 分数基于Conll2003数据集

- 1.数据集分类：新闻等数据集、医学数据集、毒药数据集
- 2.NER模型分类
- 1基于已有知识系统 毒药等命名规则 很少用了现在
- 2无监督 浅层句法知识 逆文档概率

以下模型基本都有外部知识 比如词典、地名索引（如何实现的？）

- 3监督 CRF HMM F1 90.90

常用CRF特征工程整理如下：

Feature engineering	
Features Categories	Feature
Morphological	n-gram character, n-gram word, suffixes and prefixes
Orthographic	capitalization, symbols
Linguistic	lemmatization, stemming, POS, chunking, syntactic parsing
Context	windows, conjunctions
Domain knowledge	lexicons, existing NER tools

- 4半监督 Agerri和Rigau（2016）通过提供NER分类器开发了一个半监督系统，其中包括正字法，字符n-gram，词典，前缀，后缀，双字母，三元组和布朗语料库中的非预期聚类特征，Clark 使用词嵌入的开放文本的语料库和k均值聚类

F1 91.36

- 5“特征推断”神经网络

word2vec预训练非常重要 以下的分类依据:表示可以基于单词，字符，其他子单词单元或这些单元的任意组合

(1)词 bilstm+crf conll2003 F1 84.26

(2)字符 bilstm+crf conll2003 F1 84.52

（2016）在卷积神经网络（CNN）上使用高速公路网络对单词的字符序列，然后使用另一层LSTM + softmax进行最终预测

seq2seq 定长60字符非长句 F1 86.5

(3)词+字符 两大类

第一种 单词嵌入和单词的字符卷积或attention的组合

①F1 91.21 （若出现out of vocabulary 效果不好）

②加入大写特征 加大词汇表 F1 91.62

第二种 拼接

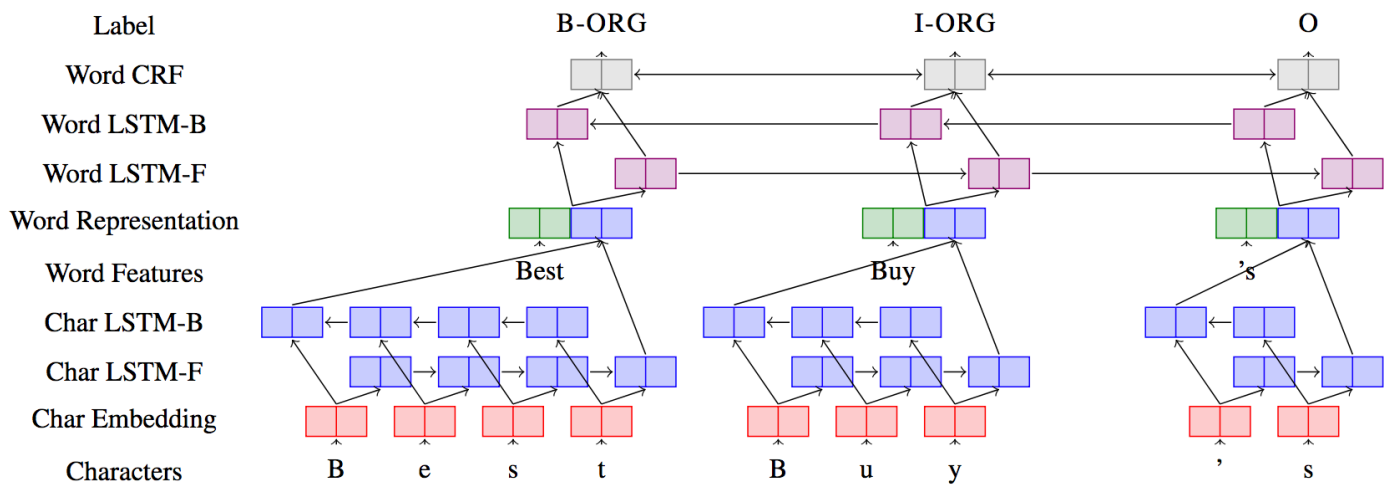


Figure 3: Word+character level NN architecture for NER

①上图F1 90.94

②多任务、多语言迁移学习 F1 91.26

(4)词+字符+前后缀

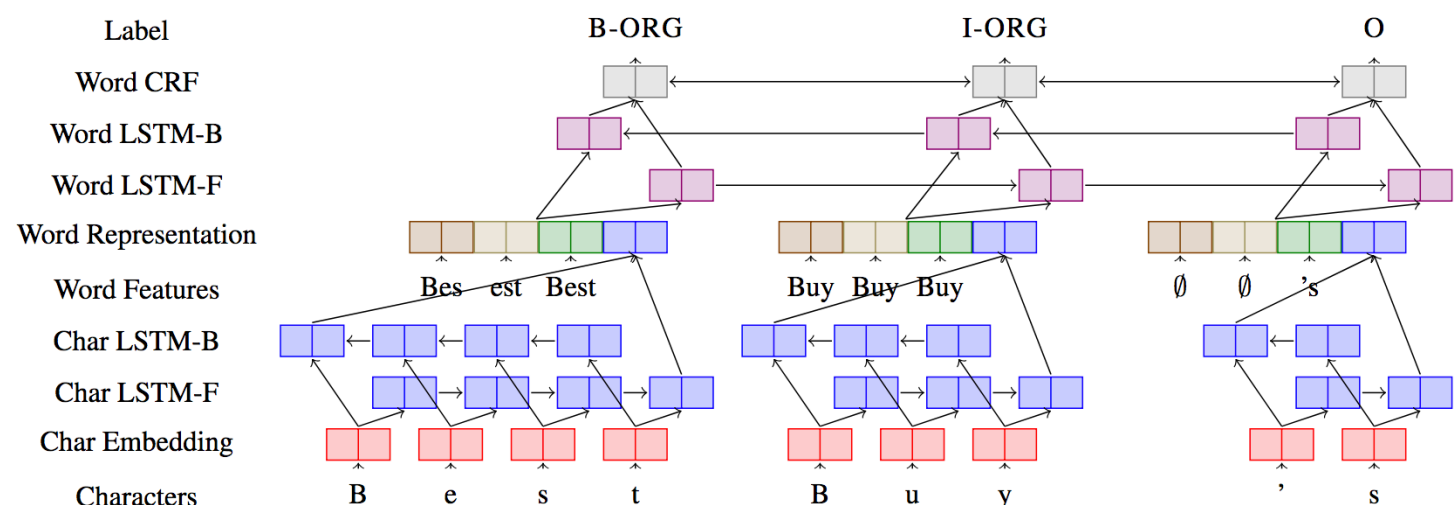


Figure 4: Word+character+affix level NN architecture for NER

Word representation处使用CNN提取前后缀 F1 91.21

(5)字符+词+attention

①词向量与字符向量的拼接改进为了权重求和 F1 89.91

②加入了音韵特征，并在字符向量上使用 attention 机制来学习关注更有效的字符 没有做Conll2003

这两篇都是16年的文章 没有提及最近的attention

总结：16年后词向量+字符是主流

Feature-engineered machine learning systems	Dict	SP	DU	EN	GE
Carreras et al. (2002) binary AdaBoost classifiers	Yes	81.39	77.05	-	-
Malouf (2002) - Maximum Entropy (ME) + features	Yes	73.66	68.08	-	-
Li et al. (2005) SVM with class weights	Yes	-	-	88.3	-
Passos et al. (2014) CRF	Yes	-	-	90.90	-
Ando and Zhang (2005a) Semi-supervised state of the art	No	-	-	89.31	75.27
Agerri and Rigau (2016)	Yes	84.16	85.04	91.36	76.42
Feature-inferring neural network word models					
Collobert et al. (2011) Vanilla NN +SLL / Conv-CRF	No	-	-	81.47	-
Huang et al. (2015) Bi-LSTM+CRF	No	-	-	84.26	-
Yan et al. (2016) Win-BiLSTM (English), FF (German) (Many fets)	Yes	-	-	88.91	76.12
Collobert et al. (2011) Conv-CRF (SENNA+Gazetteer)	Yes	-	-	89.59	-
Huang et al. (2015) Bi-LSTM+CRF+ (SENNA+Gazetteer)	Yes	-	-	90.10	-
Feature-inferring neural network character models					
Gillick et al. (2015) – BTS	No	82.95	82.84	86.50	76.22
Kuru et al. (2016) CharNER	No	82.18	79.36	84.52	70.12
Feature-inferring neural network word + character models					
Yang et al. (2017)	Yes	85.77	85.19	91.26	-
Luo (2015)	Yes	-	-	91.20	-
Chiu and Nichols (2015)	Yes	-	-	91.62	-
Ma and Hovy (2016)	No	-	-	91.21	-
Santos and Guimaraes (2015)	No	82.21	-	-	-
Lample et al. (2016)	No	85.75	81.74	90.94	78.76
Bharadwaj et al. (2016)	Yes	85.81	-	-	-
Dernoncourt et al. (2017)	No	-	-	90.5	-
Feature-inferring neural network word + character + affix models					
Re-implementation of Lample et al. (2016) (100 Epochs)	No	85.34	85.27	90.24	78.44
Yadav et al. (2018)(100 Epochs)	No	86.92	87.50	90.69	78.56
Yadav et al. (2018) (150 Epochs)	No	87.26	87.54	90.86	79.01