

# 使用TB-LSTM-CRF提高工业中文文本实体识别任务

李嘉正, 周佳乐, 程良伦

广东工业大学 自动化学院, 广东 广州  
Email: jzl@mail2.gdut.edu.cn

收稿日期: 2021年2月28日; 录用日期: 2021年3月22日; 发布日期: 2021年3月30日

## 摘要

由于缺乏自然定界符, 中文命名实体识别(NER)比英文更具挑战性。虽然中文分词(CWS)通常被认为是中文NER的关键和开放问题, 但其准确性对于下游模型训练至关重要, 而且经常遭受语音不足(OOV)的困扰。本文提出了一种改进的中文NER模型, 称为TB-LSTM-CRF, 该模型在LSTM-CRF之上引入了一个变压器块。带有Transformer Block的拟议模型利用self-attention机制从相邻字符和句子上下文中捕获信息。同时本文使用了一个全新的工业场景数据集, 在此数据集上, 与使用LSTM-CRF的基线相比, 实验结果表明, TB-LSTM-CRF方法具有竞争力, 几乎不需要任何外部资源, 例如参数迁移。

## 关键词

中文自然语言处理, 中文命名实体识别, 知识图谱

# An Improved Chinese Named Entity Recognition Method with TB-LSTM-CRF

Jiazheng Li, Jiale Zhou, Lianglun Cheng

School of Automation, Guangdong University Of Technology, Guangzhou Guangdong  
Email: jzl@mail2.gdut.edu.cn

Received: Feb. 28<sup>th</sup>, 2021; accepted: Mar. 22<sup>nd</sup>, 2021; published: Mar. 30<sup>th</sup>, 2021

## Abstract

Owing to the lack of natural delimiters, Chinese Named Entity Recognition (NER) is more challenging than it in English. While Chinese Word Segmentation (CWS) is generally regarded as key and open problem for Chinese NER, its accuracy is critical for the downstream models trainings

and it also often suffers from Out-of-Vocabulary (OOV). In this paper, we propose an improved Chinese NER model called TB-LSTM-CRF, which introduces a Transformer Block on top of LSTM-CRF. The proposed model with Transformer Block exploits the self-attention mechanism to capture the information from adjacent characters and sentence contexts. At the same time, this article uses a brand new industrial scene data set. On this data set, with the LSTM-CRF scale, the experimental results show that the TB-LSTM-CRF method is competitive and hardly requires any external resources, such as parameter migration.

## Keywords

Chinese Natural Language Processing, Chinese Named Entity Recognition, Knowledge Graph

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 介绍

工业知识自动化势在必行，工业知识图谱亟需大力推行，而命名实体识别(NER)既是自然语言处理(NLP)中的一项重要任务，也NER被广泛认为是知识图构建过程中的第一步，是知识图谱构建过程中的下游自然语言处理的重要基础例如实体链接，关系提取，事件提取和共引用解析。

就一般的NER而言，高性能方法主要包括递归神经网络(RNN)，支持向量机(SVM)或长短期记忆(LSTM)和门递归单元(GRU)，以获取单词级别的上下文信息，[1][2]所示其在英文领域取得了巨大成功，而由于中文句子没有与英文句子相同的自然定界符，而且中文句子中不同的分词可能导致句子的含义完全不同。汉语分词的问题还没有得到很好的解决。[3][4]可以看出，如果第一次错误地检测到边界，那么使用基于单词的NER模型进行正确提取就不可能了。最近，大多数用于中文NER的神经网络模型都严重依赖于出色的词嵌入表示形式和额外的词典，基于字符的方法比基于词汇的方法更适合中文NER[5]。

目前已有部分研究针对其他垂直邻域的实体抽取实验[6][7][8][9]，可是前提都需要大量的工作和极度依靠预训练模型性能。就工业领域NER而言，存在数据低资源和知识高门槛两大问题。数据低资源不止体现在工业机器人领域文本数据量少，还有专业质量不一，例如工程师甲描述某件事为“电机发热冒烟”，而工程师乙则描述为“电机转子铁芯硅钢片绝缘损坏，电机运行温度过高或冒烟”。工程师的描述不一致，同一件故障会有多种描述，甚至缺少实体，导致模型训练更难拟合，这种问题在缺少训练数据的情况下更为严重。数据集的质量差还体现在由于存在标准模糊，尺度难把握，认知水平存在差异等问题，甚至工程师撰写日志时图方便而自定义一套易记语言，或者时描述不清晰，缺少实体，都会导致数据的标注质量差。对于工业领域数据集的专业性，使用一般模型并不能很好地识别其中极具专业性的实体。例如“动圈铁芯”这样的实体一般很难出现在一般数据集上，而且模型很容易将其识别成“动圈”和“铁芯”两个实体，甚至其他，很明显这是错误的。

而在工业领域的NER研究少之又少，针对这一问题，在研究过程中利用工业生产场景下电动机工作日志作为我们本次实验的种种一个目标数据集。基于字符的NER无法充分利用显式单词和单词序列信息，这也会影响NER的性能。为了解决这个问题，我们提出了一种基于自我注意机制的变压器模块来表示字符并将潜在的词汇信息集成到基于字符的LSTM-CRF中。如图1所示，我们使用Transformer模块[10]搜索与上下文相关的字符，并输出包含与字符相关的信息的向量表示。门单元用于在包含多个字符信息的向量之间动态路由信息。最后，使用条件随机场(CRF)来整合和标准化在NER数据集上训练的预测，我们的方法可

以从上下文中自动有效地找到更多有用的单词，从而获得更好的性能。与基于字符和基于单词的NER方法相比，我们的模型可以自动获取上下文词汇信息，而无需使用其他外部资源。而且由于其加入了Transformer模块和LSTM，能对前后文的特征信息有更充分的学习，更有利于学习专业性强的长实体。在OntoNote [11]、Weibo NER [12]、Chinese Resume [13]和工业电机数据集上我们的方法在各项指标上都有具有竞争力的提升。

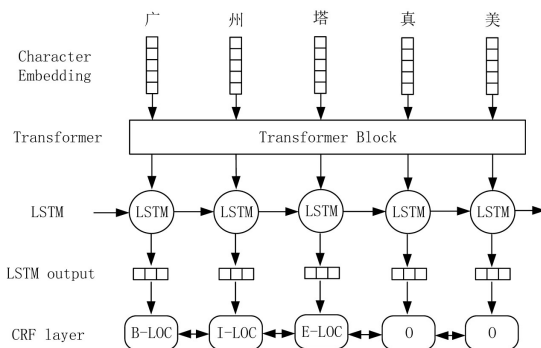


Figure 1. TB-LSTM-CRF model  
图 1. TB-LSTM-CRF 模型

## 2. TB-LSTM-CRF 模型

为了从相邻字符中捕获信息。我们采用 LSTM-CRF 作为主要网络结构，并研究了中文 NER 的Transformer 模块，该模块由一个自我注意层组成。

在本文中，我们使用 BIOES 标记方案 [14] 进行实验。我们将输入句子序列表示为  $X = \{x_1, x_2, x_3, \dots, x_m\}$ ，其中  $x_i$  表示句子  $X$  中第  $i$  个字符嵌入向量，而  $m$  表示最大句子长度。相应地，我们定义  $Y = \{y_1, y_2, y_3, \dots, y_m\}$  作为句子预测序列，其中  $y_i$  表示句子  $Y$  中第  $i$  个字符嵌入向量。目的是通过模型函数  $f_\theta : X \mapsto Y$  来获得实体类别序列。

### 2.1. Transformer 模块

Transformer 模块旨在对输入字符序列进行编码，并在上下文中隐式对含义相关的字符进行分类。第  $i$  字符的  $d$  维嵌入向量用  $x_i \in \mathbb{R}^d$  表示。模块整体框架如图 2 所示。

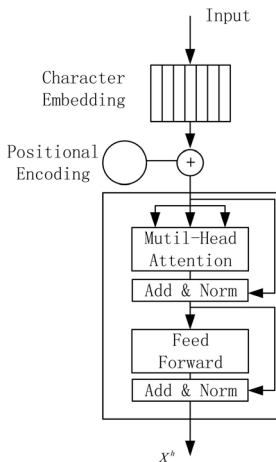


Figure 2. The overall framework of the transformer block  
图 2. Transformer 模块整体框架

### 2.1.1. 位置编码

Transformer 模块按照位置编码来理解语言的前后文，并使用 self-attention 机制和完全连接。Transformer 模型没有递归神经网络的迭代操作，因此我们需要将每个单词的位置信息提供给 Transformer 模块，以便该模型可以识别句子中字符的顺序关系。位置嵌入被初始化为一个向量，其维度为(字典大小，嵌入维度)，并按以下两个部分计算：

$$PE_{pos,2i} = \sin\left(\frac{pos}{10000^{2i/d}}\right) \quad (1)$$

$$PE_{pos,2i+1} = \cos\left(\frac{pos}{10000^{2i/d}}\right) \quad (2)$$

在上面公式里， $pos$  表示字符在句子中的位置，取值范围是[0, 最大句子长度]。 $i$  表示字符向量的第  $i$  个维度，取值范围是[0, 最大嵌入维度]。 $d$  表示字符向量的维度数量。这里有一对正弦和余弦函数，它们是与嵌入向量维度相对应的一组奇数和偶数的维。

然后将字符嵌入和位置编码对应相加，最后获得了包含位置信息的新字符嵌入向量  $\Lambda_i \in \mathbb{R}^{\text{embed.dim}}$ 。 $PE_{pos}$  表示第  $pos$  个字符的位置编码。

$$\Lambda_i = x_i + PE_{pos} \quad (3)$$

### 2.1.2. Self-Attention 机制

为了有效地学习句子中单词的依存关系并捕获句子的内部结构[15]，我们使用多头注意力(multi-head attention)来处理先前的特征向量  $\Lambda \in \mathbb{R}^{\text{batch size} \times \text{seq.len} \times \text{embed.dim}}$ 。我们分配三个权重矩阵  $W_Q, W_K, W_V \in \mathbb{R}^{\text{embed.dim} \times \text{embed.dim}}$  并根据以下公式线性映射后获得三个权值矩阵  $Q, K, V \in \mathbb{R}^{\text{embed.dim} \times \text{embed.dim}}$ ，和通过下面线性映射公式得到三个特征矩阵  $Q, K, V$ ：

$$Q = \Lambda W_Q \quad (4)$$

$$K = \Lambda W_K \quad (5)$$

$$V = \Lambda W_V \quad (6)$$

然后，我们可以通过以下公式得到注意力矩阵：

$$H = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (7)$$

$QK^T$  表示注意力矩阵， $\sqrt{d_k}$  的目的是使注意力矩阵遵循标准正态分布，从而使 softmax 层的结果更稳定。

### 2.1.3. Add & Layer Normalization

上一小节获得了注意矩阵  $H$ 。我们对其进行转置以使其与特征矩阵  $\Lambda$  的尺寸一致，然后由于元素的尺寸相同，直接将这两个矩阵相加。

$$H = \Lambda + H \quad (8)$$

Layer Normalization 的作用是将神经网络中的隐藏层归一化为标准正态分布，从而加快训练和收敛速度：

$$H = \text{LayerNorm}(H) = \alpha \odot \frac{h_{ij} - \mu_i}{\sqrt{\sigma_i^2 + \varepsilon}} + \beta \quad (9)$$

其中  $\sqrt{d_k}$  是注意力矩阵  $H$  的第  $i$  行和第  $j$  列， $\varepsilon$  是一个很小的数字，可以防止方程中被零除。引入两个

可训练的参数  $\alpha$  和  $\beta$  来弥补归一化期间丢失的信息。

注意,  $\odot$  表示元素乘法而不是点积。通常, 我们将  $\alpha$  初始化为 1, 将  $\beta$  初始化为 0。  $\mu_i$  和  $\sigma_i$  分别代表每行的均值和方差。计算公式如下, 其中  $m$  表示矩阵  $H$  中的列数。

$$\mu_i = \frac{1}{m} \sum_{j=1}^m h_{ij} \quad (10)$$

$$\sigma_i^2 = \frac{1}{m} \sum_{j=1}^m (h_{ij} - \mu_i)^2 \quad (11)$$

#### 2.1.4. Feedforward

在前馈网络部分, 此处使用两级线性映射, 并通过 ReLU 激活函数进行激活。在前馈网络之后, 后续的剩余连接和层规范化与前一小节中的类似, 不同之处在于前一个  $H$  由前馈网络的输出代替。

## 2.2. Long Short-Term Memory (LSTM)

LSTM 的目的是进一步处理上下文序列关系, 其输入是 Transformer 模块输出的特征向量。

## 2.3. Conditional Random Field (CRF)

为了提高序列标签预测的准确性, 我们需要参考相邻字符的预测结果。这是序列标记的问题, 也是 CRF 可以完美适配的工作。在 LSTM 输出该句子中每个字符的隐藏状态  $h_1, h_2, \dots, h_m$  之后, 我们利用标准的 CRF 层来接收它。

## 2.4. Training and Inference

一阶维特比算法用于在基于字符的输入序列上找到得分最高的标签序列。我们使用句子级别的对数似然损失和  $L_2$  正则化来训练模型:

$$L = \sum_{i=1}^N \log(P(y_i | s_i)) + \frac{\lambda}{2} \|\theta\|^2 \quad (10)$$

此处  $\lambda$  和  $\theta$  分别是为  $L_2$  正则的自定义设置参数。

## 3. 实验和结果

为了验证我们提出的方法的有效性, 我们对涉及不同领域的几个中文 NER 数据集进行了一些实验。我们采用以下评估指标, 分别采用标准精度(P), 召回率(R)和 F1 得分(F1)。实验数据集的详细信息将在表 1 中显示。

我们的实验中使用了三个数据集, 分别包含在 OntoNote [11], Weibo NER [12] 和 Chinese Resume [13]。OntoNote 数据集和 WeiboNER 数据集均包含 4 个命名实体类别: PER (人), ORG (组织), LOC (位置) 和 GPE (地缘政治实体)。中文简历数据集使用八种类型的命名实体进行注释: CONT (国家/地区), EDU (教育机构), LOC, PER, ORG, PRO (专业), RACE (种族背景) 和 TITLE (职位)。

使用以上三个数据集, 我们将 TB-LSTM-CRF 方法的性能与最新技术水平进行了比较, 并形成了便于参考的表格。LSTM + CRF 是基线方法。

### 3.1. OntoNote 数据集

从表 2 的显示中可以看出, 在包含 4 个实体分类的 OntoNote 数据集上, 我们的方法在准确性和 F1 得分指标上取得了最佳结果。我们的方法比最近的 Zhang *et al.* (2018) [13] 的准确性高 2.43%, 在 F1 分数

上比最好的 Zhang *et al.* (2018) [13]高 0.7%，比 Zhu *et al.* (2019) [16]低 1.07%。与基线相比，我们提出的新方法已经取得了全面的改进。

**Table 1.** Statistics of datasets

**表 1.** 数据集信息

数据集	类型	训练集	验证集	测试集
OntoNotes	句子	15.7 k	4.3 k	4.3 k
	字符	491.9 k	200.5 k	208.1 k
	实体	13.4 k	6.95 k	7.7 k
Weibo	句子	1.4 k	0.27 k	0.27 k
	字符	73.5 k	14.5 k	14.8 k
	实体	1.89 k	0.39 k	0.42 k
Resume	句子	3.8 k	0.46 k	0.48 k
	字符	124.1 k	13.9 k	15.1 k
	实体	1.34 k	0.16 k	0.15 k
工业电机数据集	句子	0.54 k	0.18 k	0.18 k
	字符	17.124 k	4.46 k	4.81 k
	实体	4.109 k	0.87 k	0.881 k

**Table 2.** Result on OntoNote

**表 2.** OntoNote 数据集上的结果

模型	准确率	召回率	F1
[17] Yang <i>et al.</i> (2016)	65.59	71.84	68.57
[13] Zhang <i>et al.</i> (2018)	76.35	71.56	73.88
[16] Zhu <i>et al.</i> (2019)	75.05	<b>72.29</b>	73.64
基线模型	67.49	70.13	67.47
TB-LSTM-CRF	<b>78.78</b>	71.22	<b>74.58</b>

### 3.2. WeiBo 数据集

在 WeiBo 数据集上，与最近的高级方法相比，我们提出的方法仍在三个指标上进行了改进。与基线方法相比，F1 得分甚至提高了 8.48% (表 3)。

**Table 3.** Result on Weibo

**表 3.** WeiBo 数据集上的结果

模型	准确率	召回率	F1
[18] Peng and Dredze (2016)	55.25	62.97	58.99
[13] Zhang <i>et al.</i> (2018)	53.04	62.25	58.79
[16] Zhu <i>et al.</i> (2019)	55.38	62.98	59.31
基线模型	51.71	57.29	52.53
TB-LSTM-CRF	<b>57.34</b>	<b>64.18</b>	<b>61.01</b>

### 3.3. Chinese Resume 数据集

与之前的数据集相比, Chinese Resume 数据集具有较少的命名实体和更多的命名实体类别。我们的方法在此数据集上的性能比前两种方法稍差。Zhu *et al.* (2019) [16]在该数据集上的准确性和召回率均达到了最佳性能。我们的模型在 F1 得分上得分最高(表 4)。

**Table 4.** Result on Chinese resume dataset

**表 4.** 中文 Resume 数据集上的验证结果

模型	准确率	召回率	F1
[13] Zhang <i>et al.</i> (2018)	94.81	94.11	94.46
[16] Zhu <i>et al.</i> (2019)	<b>95.05</b>	<b>94.82</b>	94.94
基线模型	91.34	90.58	89.33
TB-LSTM-CRF	94.23	94.25	<b>95.22</b>

### 3.4. 工业电机数据集

我们提供了一个全新的数据集, 工业电机数据集。其中包含工业场景里电动机的运行状态及诊断信息, 包含了大量富有专业性的长实体。由于网上极少公开工业相关的实体识别数据集, 本次数据集集合极有可能是第一个相关领域实体识别数据集。我们的电动机工业数据集的静态统计如表 5。

**Table 5.** Statistics of industrial motor data set

**表 5.** 工业电机数据集静态统计信息

数据集名称	NER 标签类别	实体数量
工业电机数据集	设备	1504
	子设备	422
	部件	863
	零件	333
	属性	653
	属性值	646
	状态值	1439

在该数据集上, 我们的方法在各项性能上都处于领跑地位(表 6)。

**Table 6.** Result on industrial motor dataset

**表 6.** 工业电机数据集上的验证结果

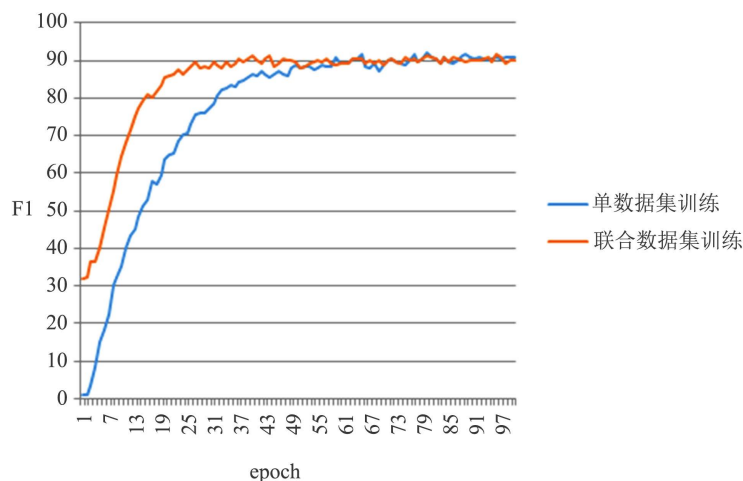
模型	准确率	召回率	F1
[13] Zhang <i>et al.</i> (2018)	79.64	84.82	84.64
[16] Zhu <i>et al.</i> (2019)	82.43	86.28	86.87
基线模型	80.67	81.27	84.77
TB-LSTM-CRF	<b>88.36</b>	<b>86.94</b>	<b>91.03</b>

### 3.5. 联合数据集

为了验证我们的模型是否需要大量的外部资源, 我们除了在单个数据集(工业电机数据集)中进行训练



如 3.4 节实验外, 我们还将模型从其余三个数据集上先预训练, 保留训练结果后模型参数, 最后将各参数作为初始值, 重做一遍实验。得出结果如表所示。其 F1 分数值随迭代次数增加折线图如图 3。



**Figure 3.** F1 score values of TB-LSTM-CRF experiments on the industrial motor dataset and the joint dataset  
**图 3.** TB-LSTM-CRF 分别在工业电机数据集和联合数据集上实验的 F1 分数值

结果表明, 虽然我们的模型在 F1 分数提升速度上比具有丰富资源的联合数据集方法训练模型时增长的快, 可是最终两者的 F1 分数最终都趋于统一。表明, 我们的模型在缺少训练样本, 或外部特征资源时仍然具有竞争力。

## 4. 结论

在本文中, 我们使用了 Transformer 模块和 LSTM 结合来提高中文 NER 模型的性能, 在 3.4 节实验中说明我们的模型几乎无需依靠其他额外资源也能快速得到优异的 F1 分数值, 而且我们的模型更加微型和实用。在 3.1 到 3.3 节实验表明, 在不同的域数据集中, 我们的模型可以比其他方法效果更好。

对于将来的工作, 我们将继续改进我们的模型, 使其更适合具有更多标签类别的任务。我们甚至将其扩展到实体关系的联合抽取任务上。

## 基金项目

《工业过程数据实时获取与知识自动化》, 国家自然科学基金委员会资助项目, 项目编号: U17012621006336。《广东省新兴海洋经济产业地图与大数据平台》, 广东省促进经济高质量发展专项资金项目。

## 参考文献

- [1] Yadav, V. and Bethard, S. (2019) A Survey on Recent Advances in Named Entity Recognition from Deep Learning Models. arXiv:1910.11470.
- [2] Liu, L., Shang, J., Ren, X., Xu, F.F., Gui, H., Peng, J. and Han, J. (2018) Empower Sequence Labeling with Task-Aware Neural Language Model. *32nd AAAI Conference on Artificial Intelligence*, New Orleans, 2-7 February 2018, 5253-5260.
- [3] Yao, L., Ge, J., Li, C., Yao, Y., Li, Z., Zeng, J., et al. (2019) Word Segmentation for Chinese Judicial Documents. *International Conference of Pioneering Computer Scientists, Engineers and Educators*, 20-23 September 2019, Guilin, 466-478. [https://doi.org/10.1007/978-981-15-0118-0\\_36](https://doi.org/10.1007/978-981-15-0118-0_36)
- [4] Chen, X., Shi, Z., Qiu, X. and Huang, X. (2017) Adversarial Multi-Criteria Learning for Chinese Word Segmentation.



- Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, 30 July-4 August, 1193-1203. <https://doi.org/10.18653/v1/P17-1110>
- [5] Li, H., Hagiwara, M., Li, Q. and Ji, H. (2014). Comparison of the Impact of Word Segmentation on Name Tagging for Chinese and Japanese. *Proceedings of the 9th International Conference on Language Resources and Evaluation*, Reykjavik, 26-31 May 2014, 2532-2536.
- [6] Yang, Y. and Katiyar, A. (2020) Simple Few-Shot Named Entity Recognition with Structured Nearest Neighbor Learning. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online, November 2020, 6365-6375. <https://doi.org/10.18653/v1/2020.emnlp-main.516>
- [7] Al-Smadi, M., Al-Zboon, S., Jararweh, Y. and Juola, P. (2020) Transfer Learning for Arabic Named Entity Recognition with Deep Neural Networks. *IEEE Access*, **8**, 37736-37745. <https://doi.org/10.1109/ACCESS.2020.2973319>
- [8] Popovski, G., Seljak, B.K. and Eftimov, T. (2020) A Survey of Named-Entity Recognition Methods for Food Information Extraction. *IEEE Access*, **8**, 31586-31594. <https://doi.org/10.1109/ACCESS.2020.2973502>
- [9] Prakash, R. and Dubey, R.K. (2020) Transfer Learning and Domain Adaptation for Named-Entity Recognition. In: Johri, P., Verma, J. and Paul, S., Eds., *Applications of Machine Learning*, Springer, Singapore, 67-73. [https://doi.org/10.1007/978-981-15-3357-0\\_5](https://doi.org/10.1007/978-981-15-3357-0_5)
- [10] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., et al. (2017) Attention Is All You Need. *31st Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 5998-6008.
- [11] Weischedel, R., Pradhan, S., Ramshaw, L., Palmer, M., Xue, N., Marcus, M., et al. (2011) OntoNotes Release 4.0. LDC2011T03. Linguistic Data Consortium, Philadelphia.
- [12] Peng, N. and Dredze, M. (2015) Named Entity Recognition for Chinese Social Media with Jointly Trained Embeddings. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, September 2015 548-554. <https://doi.org/10.18653/v1/D15-1064>
- [13] Zhang, Y. and Yang, J. (2018) Chinese NER Using Lattice LSTM. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, July 2018, 1554-1564. <https://doi.org/10.18653/v1/P18-1144>
- [14] Ratnoff, L. and Roth, D. (2009) Design Challenges and Misconceptions in Named Entity Recognition. *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, June 2009, 147-155. <https://doi.org/10.3115/1596374.1596399>
- [15] Geng, X., Wang, L., Wang, X., et al. (2020) How Does Selective Mechanism Improve Self-Attention Networks? *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, July 2020, 2986-2995. <https://doi.org/10.18653/v1/2020.acl-main.269>
- [16] Zhu, Y. and Wang, G. (2019) CAN-NER: Convolutional Attention Network for Chinese Named Entity Recognition. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1 (Long and Short Papers), Minneapolis, 2-7 June 2019, 3384-3393.
- [17] Yang, J., Teng, Z., Zhang, M. and Zhang, Y. (2016) Combining Discrete and Neural Features for Sequence Labeling. In: Gelbukh, A., Ed., *Computational Linguistics and Intelligent Text Processing*, Springer, Cham, 140-154. [https://doi.org/10.1007/978-3-319-75477-2\\_9](https://doi.org/10.1007/978-3-319-75477-2_9)
- [18] Peng, N. and Dredze, M. (2016) Improving Named Entity Recognition for Chinese Social Media with Word Segmentation Representation Learning. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin, August 2016, 149-155. <https://doi.org/10.18653/v1/P16-2025>