

# 基于RoBERTa-CNN-BiLSTM-CRF的高中数学知识实体识别

赵梓宏, 单 菁, 王佳英

沈阳工业大学软件学院, 辽宁 沈阳

收稿日期: 2024年1月5日; 录用日期: 2024年2月23日; 发布日期: 2024年2月29日

## 摘 要

命名实体识别是自然语言处理中的一个重要研究步骤, 也是自然语言中许多下游任务的前置任务。传统的命名实体识别方法通常采用简单线性或非线性模型进行识别, 实体识别的准确率不高。随着深度学习的引入, 能够处理更为复杂的非线性问题, 使用神经网络模型来提高实体识别的准确率。本文提出一种基于RoBERTa-CNN-BiLSTM-CRF模型, 用于高中数学知识实体的识别任务。首先利用RoBERTa模型中的双向Transformer编码层对数据的特征进行提取与分析生成相应的词向量, 然后利用卷积神经网络(CNN)中的卷积层和池化层实现降维操作, 提取句子中的局部特征, 最后构建适合高中数学知识点实体识别的LSTM-CRF模型进行训练和处理。经过实验表明, 该模型具有较高的准确性。准确率、召回率和F1分别达到94.32%、94.58%和94.45%。

## 关键词

高中数学知识实体识别, RoBERTa模型, CNN, BiLSTM

# Entity Recognition of High School Mathematics Knowledge Based on RoBERTa-CNN-BiLSTM-CRF

Zihong Zhao, Jing Shan, Jiaying Wang

School of Software, Shenyang University of Technology, Shenyang Liaoning

Received: Jan. 5<sup>th</sup>, 2024; accepted: Feb. 23<sup>rd</sup>, 2024; published: Feb. 29<sup>th</sup>, 2024

## Abstract

Named entity recognition is an important research step in natural language processing, and it is

文章引用: 赵梓宏, 单菁, 王佳英. 基于 RoBERTa-CNN-BiLSTM-CRF 的高中数学知识实体识别[J]. 人工智能与机器人研究, 2024, 13(1): 121-129. DOI: 10.12677/airr.2024.131014

also the pre-research of many downstream tasks in natural language. The traditional method usually adopts a simple linear or nonlinear model for entity recognition, and its accuracy is not high. With the introduction of deep learning, it can deal with more complex nonlinear problems, and use neural network models to improve the accuracy of entity recognition. In this paper, a RoBERTa-CNN-BiLSTM-CRF model is proposed for recognition of mathematical knowledge entities in high school. Firstly, the bidirectional Transformer coding layer in RoBERTa model is used to extract and analyze data features and generate corresponding word vectors. Then, the convolution layer and pooling layer in a convolutional neural network (CNN) are used to achieve dimensionality reduction and extract local features in sentences. Finally, the LSTM-CRF model suitable for entity recognition of mathematics knowledge points in high school is constructed for training and processing. Experiments show that the model has high accuracy. Precision, recall and F1 reached 94.32%, 94.58% and 94.45%, respectively.

## Keywords

Entity Recognition of High School Mathematics Knowledge, RoBERTa Model, CNN, BiLSTM

Copyright © 2024 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. 引言

命名实体识别(NER)是指在中文文本中识别出实体的边界和类别。它是关系抽取、事件抽取、知识图谱、信息抽取、问答系统、句法分析、机器翻译等众多自然语言处理(NLP)任务的基础,被广泛应用于自然语言处理领域并发挥着重要作用,用于识别文本中的实体。命名实体识别通常从实体注释开始,然后识别单个实体,最后识别复合实体。

NER 技术主要有以下方法:基于规则和词典的方法、机器学习的方法和基于深度学习的方法。基于规则和词典的方法[1]因依赖于手工指定的原则从而不需要对数据进行标注,但容易出错。Collins 等人[2]发现未标记示例对于命名实体分类会产生影响,要想减少这些影响,需要制定大量规则用来提高实体识别的效率。随后许多机器学习算法被应用于监督 NER 中,包括 HMM [3]、SVM [4]和 CRF [5]等,这些模型利用监督学习算法对特征进行仔细标记。乐娟等人[6]将隐形马尔可夫模型(HMM)和 Viterbi 算法结合起来识别京剧机构命名实体。Bender 等人[7]在命名实体识别系统中利用最大熵模型(MEM)来提高实体识别的准确性。段少鹏等人[8]提出基于条件随机场(CRF)和支持向量机(SVM)的双层模型对老挝的组织名称进行识别。

最近几年,由于深度学习能够从文本中挖掘隐含的信息,在自然语言处理中得到了越来越多的关注。NER 算法能够充分发挥深度网络的非线性特性,实现对数据的有效提取。与现有方法(如 HMM 和 CRF)相比较,深度学习方法可以利用神经网络的非线性激活函数,从海量的数据中提取出复杂特征信息。与基于特征的方法相比,深度学习方法可以从输入中自动发掘信息以及学习信息的表示,从而不需要过于复杂的特征工程。深度 NER 模型采用端到端模型可以避免流水线(pipeline)类模型中模块之间的误差传播还可以承载更加复杂的内部设计,最终产出更好的结果。凌媛等人[9]在 LSTM-CRF 模型的基础上提出增强的 LSTM-CRF 疾病命名实体识别,在字符输入处理中添加了 CNN 与 LSTM,提高了疾病识别的有效性。由于单向长短期记忆网络只能从前往后读取信息,不能更好的理解词语间的修饰作用,所以提出了双向长短期记忆(BiLSTM [10])网络。万忠宝等人[11]提出融合双向长短时记忆网络(BiLSTN)、注意力机制

和条件随机场(CRF)的中文命名实体识别的模型来提高命名实体识别的性能和效率。这些方法虽然在 NLP 领域取得了一定的效果,但实体识别的效率上还需要进一步提升。谷歌在 2018 年提出 BERT 模型[12],具有强大的语言表示能力,可以对单词进行动态建模以获得更多上下文语义表示。李伟等人[13]利用 BERT-BiLSTM-CRF 模型预测未标记数据的实体标签得到可靠的数据扩展到初试训练数据集在重新训练,证明了模型的有效性,为教育舆论中的命名实体识别问题提供了解决方案。Liu 等人[14]将 BERT-BiLSTM-CRF 模型应用到了历史领域,用来提取非结构数据中的实体信息。胡薇薇等人[15]通过改进 BERT 并结合 CRF 将其注入到知识图谱中,提高了整体性能。李凯微等人[16]提出改进 BERT 模型,并使用 BiLSTM、IDCNN 和 CRF 来提高对 JAVA 知识点的实体识别。

## 2. RoBERTa-CNN-BiLSTM-CRF 模型

### 2.1. 模型概述

本文提出了 RoBERTa-CNN-BiLSTM-CRF 模型结构如图 1 所示,模型由 4 个模块组成分别为 RoBERTa 模块、CNN 模块、LSTM 模块和 CRF 模块。RoBERTa 模块对输入文本进行预训练提取数据的语义特征,将得到的语义特征作为 CNN 模块的输入提取局部特征并进行输出,然后输入到 LSTM 模块,LSTM 模块预测实体标签的概率分布。最后通过 CRF 模块将这些实体表情的概率分布进行解码操作,解码出相应的实体标签进行输出。该模型图如图 1 所示:

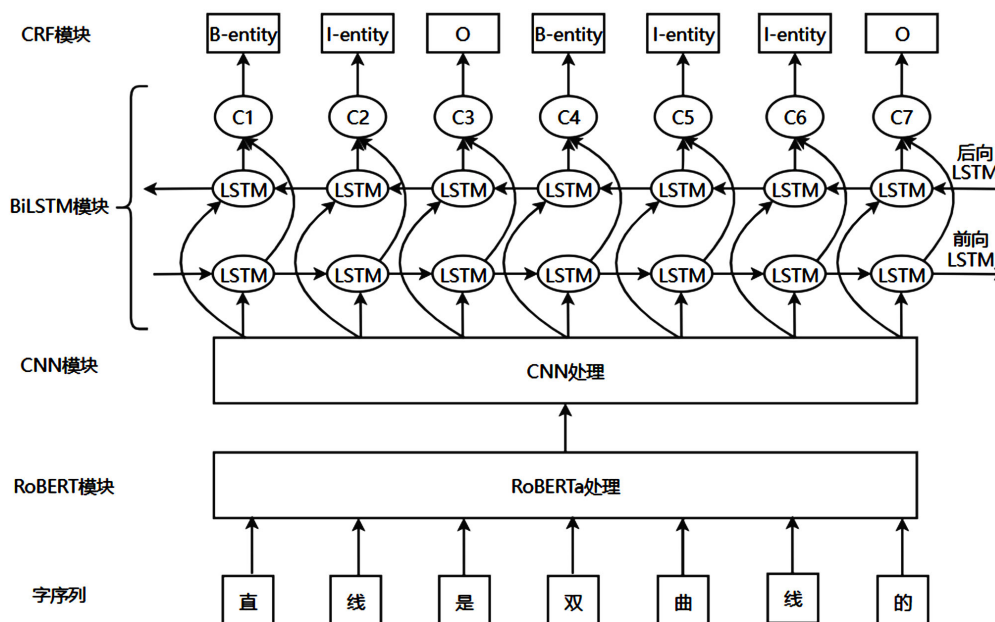


Figure 1. Structure diagram of the RoBERTa-CNN-LSTM-CRF model

图 1. RoBERTa-CNN-LSTM-CRF 模型结构图

算法描述如下:

输入: 中文语句

输出: 中文语句预测标签

算法步骤:

- (1) 将输入的文本分解成对应的字序列;
- (2) 将字序列输入到 RoBERTa 模块中,得到字符特征向量;

- (3) CNN 模块利用 CNN 获得每个字符的特征;
- (4) BiLSTM 模块通过 BiLSTM 计算输入隐藏信息;
- (5) 利用 CRF 模块对 BiLSTM 模块输出进行解码, 求解。

## 2.2. RoBERTa 模块

RoBERTa 模型是 BERT 的改进版。与 BERT 相比, RoBERTa 具有更大的训练集更大的数据样本数量、不需要使用下一句预测任务、使用更长的训练序列和动态掩码。这样可以在大量数据不断输入的过程中, 模型会逐渐适应不同的掩码策略, 学习不同的语言表征, 提高了模型输入数据的随机性, 最终提升了模型的学习能力。RoBERTa 模型主要对语言模型完成训练和处理, 确定各个句子语义之间的相互关系并进行处理, 整个模型架构分为输出层、编码层与输入层三层对数据进行处理。在利用 RoBERTa 模型对数据进行处理过程中, 模型中的双向 Transformer 编码层对数据的特征进行提取与分析起到了关键的核心作用。Transformer 编码层主要采用 Encoder 特征提取器通过计算注意力权重并进行归一化来计算语义向量, 计算公式如下:

$$Attention(Q, K, V) = \text{soft max} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (1)$$

其中  $Q, K, V$  向量为每个字向量与  $W_Q, W_K, W_V$  三个参数矩阵计算得到的,  $d_k$  为输入向量的维度。通过编码层处理后的字符特征向量由输出层输出, 作为 CNN 模块的输入向量。

## 2.3. CNN 模块

卷积神经网络(CNN)是一种专门用来处理具有类似网格结构的数据的神经网络, 它不仅广泛应用于计算机视觉领域, 也应用于自然语言处理的特征抽取中。在 NLP 任务中将句子或文章作为一个矩阵来输入给 CNN 网络, 可以用一个词语或者一个字符来表示矩阵中的每一行。再利用卷积层和池化层进行降维操作, 提取句子中的局部特征。卷积层和池化层的计算公式如下:

卷积层计算公式:

$$y_c = f \left( \sum_{k=1}^K \omega_{ck} * x_{ck} + b_c \right) \quad (2)$$

池化层计算公式:

$$y_p = f \left( \frac{1}{r_p^2} \sum_{i=1}^{r_p} \sum_{j=1}^{r_p} X_{p(i,j)} \right) \quad (3)$$

其中  $y_c$  为卷积层输出特征图,  $x_{ck}$  为输入特征图,  $\omega_{ck}$  为卷积核,  $b_c$  为偏置向量,  $K$  为卷积核的数量,  $y_p$  为池化层输出特征图,  $X_{p(i,j)}$  为输入特征图,  $r_p$  为池化窗口大小。卷积层使用卷积核滑过矩阵的整行, 再经过池化层提供一个固定大小的输出矩阵用于分类, 减少输出矩阵的维度, 同时保留句子中的特征。最后连接全连接层, 得到分类结果。

## 2.4. BiLSTM 模块

长短期记忆网络(LSTM)是对循环神经网络(RNN)进行了改进, 由于其加入了门结构, 更有选择地保存上下文信息。LSTM 由 3 个门单元模块组成分别为输入门、遗忘门及输出门。遗忘门决定了之前的记忆中需要被遗忘的信息; 输入门决定了新的输入信息对记忆的影响; 输出门决定了记忆中的信息如何被

传递给下一输出层。其计算公式如下：

遗忘门计算公式：

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (4)$$

输入门计算公式：

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (5)$$

$$\tilde{C}_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c) \quad (6)$$

输出门计算公式：

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (7)$$

长记忆计算公式：

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (8)$$

短记忆计算公式：

$$h_t = o_t * \tanh(C_t) \quad (9)$$

首先遗忘门将  $h_{t-1}$  和  $x_t$  作为输入，将  $C_{t-1}$  中的各个数字输出 0~1 的值用来表示保留信息的多少。其次由输入门层来决定更新某些值，并与  $\tanh$  层创建的  $\tilde{C}_t$  候选值向量相结合共同来更新信息。再将  $i_t$  和  $\tilde{C}_t$  相乘， $f_t$  和  $C_{t-1}$  相乘并相加得到更新后状态  $C_t$ 。最后将  $C_t$  的值进行规范化与输出门参数  $o_t$  进行点乘，得到过滤后的信息并输出。这使得 LSTM 能够有效地处理和记忆长距离依赖关系。

BiLSTM 在 LSTM 的基础上添加了反向计算，即将输入的序列进行反转，对反转后的序列重新按照 LSTM 的方式进行计算并输出，最后将正向 LSTM 的隐向量和反向 LSTM 的隐向量进行拼接，得到最终的隐向量。

所以将 CNN 和双向长短时记忆网络(BiLSTM)结合，即利用 CNN 编码得到每个段落的 embedding，再将其送到 LSTM 里拿到 LSTM 隐向量。CNN 能够有效地处理输入数据的局部特征，BiLSTM 的双向性能使模型同时考虑过去和未来的信息，从而更好地抓住序列数据中的上下文信息，这样能够学习到局部和全局特征，并且可以自适应地去提取特征，从而在处理序列数据时具有出色的表现。但 BiLSTM 只能挑出每个词的最大概率值的 label 进行输出，并不能学习到输出的各个标注之间是否存在转移依赖关系和序列标注的约束条件，所以引入了 CRF 层来学习序列标注的约束条件，确保预测出的结果的准确性。

## 2.5. CRF 模块

条件随机场(CRF)负责学习相邻实体标签之间的转移规则，当预测序列的每一个位置的输出状态，需要考虑到相邻位置的输出状态。CRF 的核心作用是建模标签之间的依赖关系，目标是学习隐变量到观测值的发射概率，以及当前观测值和下一个观测值之间的转移概率。CRF 的损失函数主要包括两个部分，分别为  $P_{total}$  和  $P_{RealPath}$ ，其中  $P_{total}$  表示所有路径分数  $P_{RealPath}$  表示真实路径分数，计算公式如下：

$P_{total}$  计算公式：

$$P_{total} = P_1 + P_2 + \dots + P_n = e^{S_1} + e^{S_2} + \dots + e^{S_n} \quad (10)$$

$$S_i = EmissionScore + TransitionScore \quad (11)$$

损失函数定义：

$$LossFunction = \frac{P_{RealPath}}{P_{total}} \quad (12)$$

其中 *EmissionScore* 为发射分数, *TransitionScore* 为转移分数。损失函数的值越大说明得到的序列越优, 那么该序列就是所有序列中的最优序列, 即识别的准确率越高。

### 3. 实验结果分析

#### 3.1. 实验数据集

本实验采用高中数学知识作为数据集, 数据集主要来源于高中数学人教版课本、百度百科和百度文库等包含高中数学知识点的相关语句。本实验包含了数据集样本, 实体, 训练集以及测试集。数据集样本的数量为 9271 个, 实体数量为 2225 个, 训练集数量为 6271 个, 验证集数量为 1500 个, 测试集数量为 1500 个。

#### 3.2. 数据集标注及评价指标

本实验采用 BIO 实体标注方法, B 表示每一个标记实体的开始, I 表示每一个实体的其余部分, O 表示非标注实体[17]。实验的主体主要分为两类: B-entity 和 I-entity, 标注结果(部分)如表 1 所示:

**Table 1.** BIO annotation results (partial) table

**表 1.** BIO 标注结果(部分)表

实体	标注标识
直	B-entity
线	I-entity
是	O
双	B-entity
曲	I-entity
线	I-entity
的	O
渐	B-entity
近	I-entity
线	I-entity
。	O

本实验选用精确率、召回率以及 *F1* 指数作为评测指标, 具体计算方式如下:

定义 *TP* 为实际为正例且预测正确的数量, *FP* 为预测为正例但实际为反例的数量, 因此有精确率:

$$PRE = \frac{TP}{TP + FP} * 100\% \quad (13)$$

定义 *FN* 为实际为正例但预测为反例的数量, 因此有召回率:

$$REC = \frac{TP}{TP + FN} * 100\% \quad (14)$$

*F1* 指数计算公式为:

$$F1 = \frac{2 * PRE * REC}{PRE + REC} * 100\% \quad (15)$$

### 3.3. 实验环境配置及参数

实验中使用设备配置如表 2 所示。

**Table 2.** Experimental environment configuration table

**表 2.** 实验环境配置表

实验环境	配置
CPU	Intel (R) Xeon (R) CPU E5-2678 v3 @ 2.50 GHz
GPU	3070 (8 G)
内存	32G
Python	3.7
Pytorch	1.12.1 + cu116

实验中使用的参数如表 3 所示。

**Table 3.** Experimental parameter table

**表 3.** 实验参数表

参数名称	值
Rnn_dim	128
Max_seq_length	64
Train_batch_size	64
Eval_batch_size	64
Gradient_accumulation_steps	1
Learning_rate	3e-5
Logging_steps	500

### 3.4. 实验结果与分析

为验证试验结果,使用高中数学知识数据集,分别使用 BERT-CRF 模型、BERT-BiLSTM-CRF 模型、RoBERTa-BiLSTM-CRF 模型以及 RoBERTa-CNN-BiLSTM-CRF 模型对高中数学知识数据集进行命名实体识别实验,结果如表 4 所示。

**Table 4.** Comparison table of model experiments

**表 4.** 模型实验对比表

模型	PRE	REC	F1
BERT-CRF	92.40%	94.66%	93.52%
BERT-BiLSTM-CRF	93.27%	94.76%	94.02%
RoBERTa-BiLSTM-CRF	93.27%	94.79%	94.02%
RoBERTa-CNN-BiLSTM-CRF	94.32%	94.58%	94.45%

由实验结果数据可知,文中使用的 RoBERTa-CNN-BiLSTM-CRF 模型在精确率、召回率和 F1 值都高于该实验模型中的其它模型,并且各个评测指标分别达到 94.32%、94.58%、94.45%。由下表可知本文中使用的 RoBERTa-CNN-BiLSTM-CRF 模型在处理高中数学知识命名实体的识别中具有很大的优势。

### 3.5. 模型测试

为验证 RoBERTa-CNN-BiLSTM-CRF 模型的准确性,对模型进行测试,选取导数应用这个知识点进

进行测试，测试结果如表 5 所示。

**Table 5.** Comparison table of model test results

**表 5.** 模型测试结果对比表

知识点描述	模型	实体
利用导数判断函数的单调性。设函数 $y=f(x)$ 在某个区间内可导， $f'(x) > 0$ ，那么 $f(x)$ 为增函数，如果 $f'(x) < 0$ ，那么 $f(x)$ 为减函数，如果在某个区间内恒有 $f'(x) = 0$ ，那么 $f(x)$ 为常数。	BERT-CRF	导数，函数，单调性，函数，区间，可导，函数，减函数，区间，常数
	BERT-BiLSTM-CRF	导数，函数，单调性，函数，区间，可导，函数，减函数，区间，常数
	RoBERTa-BiLSTM-CRF	导数，函数，单调性，函数，区间，可导，函数，函数，区间，常数
	RoBERTa-CNN-BiLSTM-CRF	导数，函数，单调性，函数，区间，可导，增函数，减函数，区间，常数

针对导数应用这个知识点的描述，分别利用四个模型进行测试，由上表发现 BERT-CRF 模型、BERT-BiLSTM-CRF 模型和 RoBERTa-BiLSTM-CRF 模型对高中知识点中的“增函数”和“减函数”这两个实体识别的不是非常准确，而 RoBERTa-CNN-BiLSTM-CRF 模型能够准确的识别出“增函数”和“减函数”这两个实体，具有较高的准确性。

#### 4. 结束语

本文是针对中文的文本实体进行识别本文提出 RoBERTa-CNN-BiLSTM-CRF 模型，RoBERTa 模块将输入文本进行预处理得到表示上下文语义特征信息的词向量，再通过 CNN 模块获得字的特征来补充词向量，然后通过 LSTM 模块对生成的词向量进行特征提取，最后通过 CRF 进行解码。实验结果表明，该模型在高中数学知识实体识别中取得了很高的精确率、召回率和 F1 值，分别达到 94.32%、94.58%、94.45%。对比实验结果表明，本文提出的模型在和其他模型对比中各方面的性能为最高。由于该实验的是在特定的数据集下取得了很好的效果，下一步的工作，考虑将其应用到其他数据集中，观察其性能，并对其进行相应的优化和改进。

#### 基金项目

辽宁省教育厅高校基本科研项目(JYTMS20231226)，面向多模态大数据的智能感知技术研究。

#### 参考文献

- [1] Li, Y., Chiticariu, L., Reiss, F., *et al.* (2010) Domain Adaptation of Rule-Based Annotators for Named-Entity Recognition Tasks. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Cambridge, October 2010, 1002-1012.
- [2] Collins, S. Y. (1999) Unsupervised Models for Named Entity Classification. <https://aclanthology.org/W99-0613/>
- [3] Morwal, S. (2012) Named Entity Recognition Using Hidden Markov Model (HMM). *International Journal on Natural Language Computing*, **1**, 15-23. <https://doi.org/10.5121/ijnlc.2012.1402>
- [4] Ju, Z., Wang, J. and Zhu, F. (2011) Named Entity Recognition from Biomedical Text Using SVM. *2011 5th International Conference on Bioinformatics and Biomedical Engineering*, Wuhan, 10-12 May 2011, 1-4. <https://doi.org/10.1109/icbbe.2011.5779984>
- [5] Song, S., Nan, Z. and Huang, H. (2017) Named Entity Recognition Based on Conditional Random Fields. *Cluster Computing*, **22**, 5195-5206. <https://doi.org/10.1007/s10586-017-1146-3>
- [6] 乐娟, 赵玺. 基于 HMM 的京剧机构命名实体识别算法[J]. *计算机工程*, 2013, 39(6): 266-271+286.
- [7] Bender, O., Och, F. J. and Ney, H. (2003) Maximum Entropy Models for Named Entity Recognition. *CONLL'03: Proceedings of the 7th Conference on Natural Language Learning at HLT-NAACL 2003*, Edmonton, 31 May 2003, 148-151.



- <https://doi.org/10.3115/1119176.1119196>
- [8] Duan, S.P., Zhou, L.J., Zhou, F., *et al.* (2017) Laos Organization Name Using Cascaded Model Based on SVM and CRF. *MATEC Web of Conferences*, **100**, Article No. 02051. <https://doi.org/10.1051/mateconf/201710002051>
- [9] Yuan, L., Oladimeji, F., *et al.* (2019) A Domain Knowledge-Enhanced LSTM-CRF Model for Disease Named Entity Recognition. *AMIA Joint Summits on Translational Science Proceedings*, **2019**, 761-770. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6568095/>
- [10] Graves, A. and Schmidhuber, J. (2005) Framewise Phoneme Classification with Bidirectional LSTM Networks. *IEEE International Joint Conference on Neural Networks*, **18**, 2047-2052. <https://doi.org/10.1016/j.neunet.2005.06.042>
- [11] Wan, Z., Xie, J., Zhang, W., *et al.* (2019) BiLSTM-CRF Chinese Named Entity Recognition Model with Attention Mechanism. *Journal of Physics: Conference Series*, **1302**, Article ID: 032056. <https://doi.org/10.1088/1742-6596/1302/3/032056>
- [12] Devlin, J., Chang, M.W., Lee, K., *et al.* (2018) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.
- [13] Li, W., Du, Y.J., Li, X.Y., *et al.* (2022) UD\_BBC: Named Entity Recognition in Social Network Combined BERT-BiLSTM-CRF with Active Learning. *Engineering Applications of Artificial Intelligence*, **116**, Article ID: 105460. <https://doi.org/10.1016/j.engappai.2022.105460>
- [14] Liu, S., Yang, H., Li, J., *et al.* (2021) Chinese Named Entity Recognition Method in History and Culture Field Based on BERT. 2021 *International Conference on Culture-Oriented Science & Technology (ICCST)*, Beijing, 18-21 November 2021, 181-186. <https://doi.org/10.1109/ICCST53801.2021.00047>
- [15] Hu, W.W., He, L., Ma, H.H., Wang, K. and Xiao, J.F. (2022) KGNER: Improving Chinese Named Entity Recognition by BERT Infused with the Knowledge Graph. *Applied Sciences*, **12**, 7702-7702. <https://doi.org/10.3390/app12157702>
- [16] 李凯微, 王佳英, 单菁. 基于多模融合的Java领域命名实体识别[J]. *计算机科学与应用*, 2022, 12(12): 2712-2724. <https://doi.org/10.12677/csa.2022.1212275>
- [17] Yang, Y.J., Shen, X.J. and Wang, Y.J. (2020) BERT-BiLSTM-CRF for Chinese Sensitive Vocabulary Recognition. In: Li, K.S., Li, W., Wang, H. and Liu, Y., Eds., *Artificial Intelligence Algorithms and Applications*, Springer, Singapore, 257-268. [https://doi.org/10.1007/978-981-15-5577-0\\_19](https://doi.org/10.1007/978-981-15-5577-0_19)