

# 命名实体识别研究综述

胡德洲, 李贯峰

(宁夏大学信息工程学院, 宁夏银川 750021)

**摘要:** 命名实体识别是机器翻译、问答系统等应用领域的重要基础工具, 一直是自然语言处理领域的研究热点之一。首先, 介绍命名实体识别定义, 整理命名实体识别任务中常用的实现工具、数据集和评估标准; 其次, 按照其发展历程总结现有命名实体识别方法, 将其分为传统方法和深度学习方法; 再次, 归纳总结命名实体识别各方法的关键思想、优缺点, 给出基于深度学习的命名实体方法的主要流程, 并按照流程顺序进行综述; 最后, 展望命名实体识别未来的研究方向, 为后续研究提供思路。

**关键词:** 自然语言处理; 命名实体识别; 机器学习; 深度学习

**DOI:** 10.11907/rjdk.231923

**中图分类号:** TP391.1

**文献标识码:** A

**开放科学(资源服务)标识码(OSID):**

**文章编号:** 1672-7800(2024)009-0001-09



## Review of Named Entity Recognition Research

HU Dezhou, LI Guanfeng

(School of Information Engineering, Ningxia University, Yinchuan 750021, China)

**Abstract:** Named entity recognition is an important foundational tool in application fields such as machine translation and question answering systems, and has always been a research hotspot in the field of natural language processing. Firstly, introduce the definition of named entity recognition, and organize the commonly used implementation tools, datasets, and evaluation criteria in named entity recognition tasks; Then, based on their development history, existing named entity recognition methods are summarized and divided into traditional methods and deep learning methods; Next, summarize the key ideas, advantages, and disadvantages of various methods for named entity recognition, provide the main process of deep learning based named entity methods, and summarize them in the order of the process; Finally, looking forward to the future research directions of named entity recognition, providing ideas for subsequent research.

**Key Words:** natural language processing; named entity recognition; machine learning; deep learning

### 0 引言

命名实体识别(Named Entity Recognition, NER)旨在识别文本中的命名实体, 并将其归类为预先定义好的实体类别, 例如人名、地名、组织名等<sup>[1]</sup>。NER不仅是信息抽取的基础任务, 而且在文本理解、机器翻译、文本摘要等多种自然语言处理(Natural Language Processing, NLP)任务中也发挥着重大作用。

在新闻领域, 准确提取人物、时间、机构等命名实体是智能化新闻推荐的基础工作; 在生物医学领域, 面对大量

生物医学文献, 生物医学命名实体识别是生物医学文本分类、关系抽取等任务的重要基础; 在社交媒体领域存在固有的噪声信息和不当的语法结构, 命名实体识别的结果好坏直接影响用户的使用体验。虽然, NER目前应用十分广泛, 但仍存在跨领域适应性差、数据不平衡、未登录词等问题, 要解决这些挑战需要深入研究, 包括使用更复杂的模型结构、跨领域数据集收集及改进数据集标注工具等。因此, 有效提高命名实体识别的效果具有重大的研究意义。

从时间上来看, NER起源可追溯到1950年代, 当时主要从论文和医疗记录中提取结构化实体, 1980年代应用范围扩展至新闻报道领域, 直到1996年命名实体识别这一

收稿日期: 2023-08-30

扫描二维码阅读全文:



**基金项目:** 国家自然科学基金项目(62066038); 宁夏自然科学基金项目(2022AAC03026); 宁夏大学研究生创新项目(CXXM202356)

**作者简介:** 胡德洲(2000-), 男, CCF学生会会员, 宁夏大学信息工程学院硕士研究生, 研究方向为自然语言处理; 李贯峰(1979-), 男, 博士, 宁夏大学信息工程学院副教授、硕士生导师, 研究方向为语义计算、知识图谱。本文通讯作者: 李贯峰。

术语在第六届信息理解会议中被正式提出<sup>[2]</sup>。最初的实现方法主要依靠专家手工制定规则和词典,再后来凭借传统机器学习得到大幅度提升,近年来随着大语言模型和神经网络快速发展,深度学习逐渐主导NER任务研究,命名实体识别技术取得了重大突破。

目前,命名实体识别已蓬勃发展了几十年,本文调研了命名实体识别研究工作中具有代表性的综述论文。Saju等<sup>[3]</sup>发表了针对银行、医疗等垂直领域的命名实体识别方法综述。Dai等<sup>[4]</sup>针对复杂命名实体识别方法进行了总结和展望。Li等<sup>[5]</sup>详细总结了命名实体识别方法的深度学习学习方法。潘俊等<sup>[6]</sup>系统梳理了深度学习在中文命名实体识别中的广泛应用。祁鹏年等<sup>[7]</sup>针对基于深度学习的中文NER模型进行归纳分类。高翔等<sup>[8]</sup>概述了命名实体识别领域较主流的研究方向。然而,以上工作都只是对命名实体识别的部分方法进行了总结,未对NER方法进行全面梳理,且最新的深度学习方法也并未涉及。

本文首先基于国内外NER领域文献介绍了NER的定义,通过整理NER常用的实现工具、数据集和评估指标分析了NER的传统方法和深度学习方法,总结了每种方法的优缺点;然后按实现流程总结NER深度学习方法;最后最后探讨了NER未来的研究方向,以期后续命名实体识别研究提供思路。

## 1 NER概述

### 1.1 命名实体识别定义

Rau<sup>[9]</sup>提出一种从文本中抽取公司名的方法——命名实体识别,但当时抽取的只有公司名一种类型,直到1996年在MUC-6会议上命名实体识别这一定义被正式提出,将其定义为实体的唯一标识符<sup>[2]</sup>。命名实体识别可分为:①通用领域NER,需要识别人、组织、位置等命名实体;②垂直领域NER,以生物医学领域为例识别药物、疾病、基因等命名实体。

NER是将命名实体定位并分类为预定义的实体类型的过程。从形式上而言,对于给定的文本序列 $S=\langle w_1, w_2, w_3, \dots, w_n \rangle$ ,经过命名实体识别后会得到 $\langle I_b, I_e, t \rangle$ 的三元组列表。其中, $I_b \in [1, n]$ , $I_e \in [1, n]$ , $I_b$ 代表开始标签, $I_e$ 代表结束标签, $t$ 为实体所属类别,NER系统识别命名实体实例如图1所示。对于给定的文本序列,经过NER系统得到所属实体类别,根据结果可知Li Ming为人物类实体(Person类实体),而另外两个实体为地点类实体(Location类实体)。

### 1.2 命名实体识别工具与数据集

目前有许多权威的命名实体识别工具,例如哈尔滨工业大学研发的LTP平台和百度研发的LAC工具包。Zshot是目前最新用于零样本命名实体识别的开源框架,对常用的NER实现工具进行了整理,如表1所示。高质量的数据集对NER模型的训练和评估有着至关重要的作用。NER

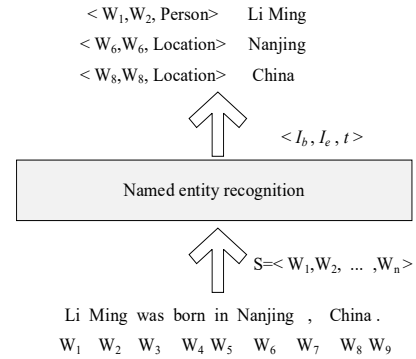


Fig. 1 NER task instance

图1 NER任务实例

领域使用最广泛的数据集包括MUC-6、GENIA、FSU-PRGE、NCBI Disease、Few-NERD等。本文按照数据集的时间顺序进行了汇总整理,具体如表2所示。

Table 1 Summary of NER implementation tools

表1 NER实现工具汇总

平台名称	引用地址
StanfordCoreNLP	<a href="https://stanfordnlp.github.io/CoreNLP/">https://stanfordnlp.github.io/CoreNLP/</a>
NeuroNER	<a href="http://neuroner.com/">http://neuroner.com/</a>
NERsuite	<a href="http://nersuite.nlplab.org/">http://nersuite.nlplab.org/</a>
NLTK	<a href="https://www.nltk.org">https://www.nltk.org</a>
Stanza	<a href="https://stanfordnlp.github.io/stanza/">https://stanfordnlp.github.io/stanza/</a>
LTP	<a href="http://ltp.ai/download.html">http://ltp.ai/download.html</a>
LAC	<a href="https://github.com/baidu/lac">https://github.com/baidu/lac</a>
OpenNLP	<a href="https://opennlp.apache.org/">https://opennlp.apache.org/</a>
Zshot	<a href="https://ibm.github.io/zshot/">https://ibm.github.io/zshot/</a>

### 1.3 命名实体识别的评估标准

模型构建完成后的评估也十分重要,不仅是为了检验模型能否达到预期效果,更重要的是需要根据实验结果对模型参数加以修正,从而优化模型。在命名实体识别领域,通常使用TP、FP和FN计算准确率(P)、召回率(R)、F1值3项评估指标,具体计算公式如式(1)一式(3)所示。其中,TP代表将正样本预测为正,FP代表将负样本预测为正,FN代表将正样本预测为负。

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (3)$$

除了使用P、R、F1外,还通过Micro-F1、Macro-F1辅助评估模型的识别效果。假设实体的类别共 $n$ 种,Micro-F1计算方法是先计算所有实体类别总的P、R,然后代入F1公式得到Micro-F1。具体计算公式如式(4)一式(6)所示。由此可见,Micro-F1综合考虑了各种类型实体数量,当实体数量不均匀时Micro-F1可较好描述模型效果。

$$P_{micro} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FP_i} \quad (4)$$

Table 2 Summary of NER datasets

表 2 NER数据集汇总

数据集	年份	语言	文本来源	实体类型数	地址
MUC-6	1996	英文	新闻	7	<a href="https://catalog.ldc.upenn.edu/LDC2003T13">https://catalog.ldc.upenn.edu/LDC2003T13</a>
MUC-7	1997	英文	新闻	7	<a href="https://catalog.ldc.upenn.edu/LDC2001T02">https://catalog.ldc.upenn.edu/LDC2001T02</a>
1998人民日报数据集	1998	中文	人民日报	3	<a href="https://github.com/GuocaiL/nlp_corpus/tree/main/open_ner_data/people_daily">https://github.com/GuocaiL/nlp_corpus/tree/main/open_ner_data/people_daily</a>
CoNLL03	2003	英文	新闻	4	<a href="https://www.clips.uantwerpen.be/conll2003/ner">https://www.clips.uantwerpen.be/conll2003/ner</a>
GENIA	2004	英文	临床文本	36	<a href="http://www.geniaproject.org/home">http://www.geniaproject.org/home</a>
MSRA	2006	中文	新闻	3	<a href="https://github.com/InsaneLife/ChineseNLPCorpus/tree/master/NER">https://github.com/InsaneLife/ChineseNLPCorpus/tree/master/NER</a>
WikiGold	2009	英文	维基百科	4	<a href="https://figshare.com/articles/Learning_multilingual_named_entity_recognition_from_Wikipedia/5462500">https://figshare.com/articles/Learning_multilingual_named_entity_recognition_from_Wikipedia/5462500</a>
FSU-PRGE	2010	英文	PubMed MEDLINE	5	<a href="https://julielab.de/Resources/FSU_PRGE.html">https://julielab.de/Resources/FSU_PRGE.html</a>
NCBI Disease	2014	英文	PubMeb	790	<a href="https://www.ncbi.nlm.nih.gov/CBBresearch/Dogan/DISEASE">https://www.ncbi.nlm.nih.gov/CBBresearch/Dogan/DISEASE</a>
CCKS2017	2017	中文	电子病历	5	<a href="https://www.biendata.xyz/competition/CCKS2017_2/">https://www.biendata.xyz/competition/CCKS2017_2/</a>
Weibo NER	2018	中文	社交媒体	4	<a href="https://github.com/hltcoe/golden-horse/tree/master/data">https://github.com/hltcoe/golden-horse/tree/master/data</a>
CCKS2019	2019	中文	电子病历	6	<a href="http://openkg.cn/dataset/yidu-s4k">http://openkg.cn/dataset/yidu-s4k</a>
CLUENER2020	2020	中文	新闻	10	<a href="https://github.com/CLUEbenchmark/CLUENER2020">https://github.com/CLUEbenchmark/CLUENER2020</a>
Few-NERD	2021	英文	维基百科	66	<a href="https://ningding97.github.io/fewnerd/">https://ningding97.github.io/fewnerd/</a>
GAHC	2022	中文	商品信息	52	<a href="https://github.com/pinskyrobin/GAHC_NER">https://github.com/pinskyrobin/GAHC_NER</a>

$$R_{micro} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FN_i} \quad (5)$$

$$F1_{micro} = 2 \times \frac{P_{micro} \times R_{micro}}{P_{micro} + R_{micro}} \quad (6)$$

Macro-F1 将所有类别的 P、R 求平均数后代入 F1 公式进行计算,具体计算公式如式(7)~式(9)所示。由此可见, Micro-F1 并未考虑各种类型实体的具体数量,平等看待每一种实体类别,但最终结果容易被具有较高 P 和 R 的实体种类影响。

$$P_{macro} = \frac{\sum_{i=1}^n P_i}{n} \quad (7)$$

$$R_{macro} = \frac{\sum_{i=1}^n R_i}{n} \quad (8)$$

$$F1_{macro} = 2 \times \frac{P_{macro} \times R_{macro}}{P_{macro} + R_{macro}} \quad (9)$$

## 2 NER传统方法

### 2.1 基于规则

基于规则的 NER 方法具备易实现、无需训练的优点,在早期 NER 任务中取得了不错的效果。针对不同领域的数据集,领域专家们需要构造不同规则,但一般都根据关键字、位置词、指示词、标点符号等信息来构造,在此基础上凭借模式匹配达到实体识别效果。Rua<sup>[9]</sup>首次结合人工规则和启发式思想自动识别公司名称。袁金斗等<sup>[10]</sup>基于规则和词典对用电安全领域的命名实体进行识别,制定的规则模板包含基于字符构词规则的识别和基于词性组合规则的识别两大类,在测试样本较少的情况下 F1 值最高达到 90%,但面对海量数据时会出现规则模版冗余、识别率下降的问题。Shaanan 等<sup>[11]</sup>利用上下文特征和地名词典构造规则,使用一种过滤机制处理格式不正确的命名实体

并消除命名实体的歧义,实现了对阿拉伯语的命名实体识别。闫萍<sup>[12]</sup>根据姓名构造命名实体识别规则,利用内部规则计算分词得到的候选姓名概率估值,并从语料中提取满足条件的阈值进行判断,最后依据外部规则再次筛选得到最终结果,以有效提升中文人名识别的准确率。包振山等<sup>[13]</sup>针对中医古籍数字化程度低、标注语料少的问题,提出一种基于规则和半监督学习的命名实体识别方法, F1 值达到 83.18%。刘合兵等<sup>[14]</sup>针对小麦病虫害领域实体结构复杂、分布不均匀等问题,提出一种基于规则修正的实体识别模型,融合规则后 P、R、F1 值分别提升 1.71%、0.34%、1.03%。Quimbaya 等<sup>[15]</sup>提出一种基于字典的组合方法识别电子健康记录的命名实体,证明了提升召回率对命名实体识别精度的影响有限。

基于规则的 NER 方法主要基于人工制定的语义和句法规则识别实体,只有字典内容详尽时才能取得较好的效果,且特定领域规则十分依赖行业专家、可迁移性较差、词典也需要不断维护,不断衍生的新实体和实体不规则性使得词典维护难度较大,造成基于规则的 NER 方法虽然准确率高,但可迁移性较差。

### 2.2 基于统计机器学习

随着机器学习在 NLP 领域的发展,越来越多研究者开始转变思路,从制定规则和词典到利用机器在大规模语料库中训练以自动识别命名实体。特征工程是该方法的关键,主流机器学习算法包括隐马尔可夫模型(Hidden Markov Models, HMM)、最大熵模型(Maximum Entropy Models, MEM)、支持向量机(Support Vector Machines, SVM)等。

#### 2.2.1 HMM

HMM 是经典的机器学习序列模型之一,假设状态序列在某时刻的值只与前一时刻状态序列取值有关。Bikel 等<sup>[16]</sup>提出基于 HMM 的 NER 系统 IdentiFinder,但由于

HMM模型无法充分考虑上下文信息,因此难以捕捉远距离信息,只实现了对姓名、数字数量等实体的识别。实验结果表明该系统性能相较于基于规则的方法更具竞争力。Zhou等<sup>[17]</sup>基于HMM提出一种组块标记器的NER方法,通过整合4种类型的内部证据和外部证据,在MUC-6和MUC-7数据集上F1值分别达到96.6%和94.1%,并始终优于基于规则的方法。俞鸿魁等<sup>[18]</sup>提出层叠HMM的NER方法,将多种类型命名实体识别融入一个相对统一的理论模型中,在中文命名实体识别中效果不错。张华平等<sup>[19]</sup>结合HMM提出一种基于角色标注的中文人名识别方法,主要思想采取viterbi算法对分词结果进行角色标注,同时结合最大匹配算法识别命名实体,有效解决了姓名丢失、内部成词等问题,实验表明该方法召回率接近98%。

### 2.2.2 MEM

基于MEM的NER方法主要原理是在根据约束条件筛选出的概率模型中选取熵最大的作为最终模型,舍弃了HMM中不合理的独立性假设。Borthwick等<sup>[20]</sup>首次将MEM用于英文命名实体识别任务,提出了一个名为最大熵命名实体的新系统。该系统框架遵守最大熵理论,且利用词汇特征、文本类型等多种信息作出标记决策。Bender等<sup>[21]</sup>在文献[20]方法的基础上进行优化,构造了一个基础的命名实体识别器,从非注释数据中提取命名实体及其上下文信息,并将所识别的信息融入原有识别器以进一步提升识别准确率。王江伟<sup>[22]</sup>以MEM为基本框架,结合局部

最优解码算法和全局最优解码算法,实现了中文命名实体识别,并针对最大熵模型的解码问题提出树一栅格解码算法,有效解决了命名实体识别过程中潜在的冲突问题。张玥杰等<sup>[23]</sup>提出融合局部特征和全局特征的最大熵汉语命名实体识别模型,同时引入启发式规则进行多组消融实验,实验表明该方法对不同测试数据源具有一致性,F1值最高达到86.31%。

### 2.2.3 SVM

SVM是一种二分类模型,学习策略是间隔最大化,本质上就是求解凸二次规划问题的最优化算法。Isozaki等<sup>[24]</sup>基于SVM提出一种特征选择方法和高效训练算法有效提升了系统训练速度,还适用于分块和词性标记任务。郑亚南等<sup>[25]</sup>将SVM与Glove词向量相结合,利用Glove词向量模型提取语义特征,再使用SVM进行文本分类,模型构造完毕后与Word2vec词向量结合SVM的实验表明,Glove作为嵌入式向量具有一定的稳定性。陈霄等<sup>[26]</sup>在应用SVM的基础上设计了一种分布递增式学习方法,利用主动学习策略逐步提升训练样本规模,有效提高了组织机构名识别精度,证明了主动学习策略的有效性。

综上所述,以上3种基于统计机器学习的NER方法各有优缺点(见表3),虽然该方法缓解了一词多义等问题,但模型复杂度较高。随着计算机技术快速发展,NER深度学习模型取得了更好的效果,学者们都逐渐转向该方法。

Table 3 Comparison of NER methods based on statistical machine learning

表3 基于统计机器学习的NER方法比较

模型	原理	优点	缺点
HMM	通过转移概率和生成概率对序列进行建模和分析	可解释性强	模型计算量较大
MEM	在所有可能的概率模型中选取熵最大的一个模型	准确率较高	时间复杂度较高
SVM	特征空间上的间隔最大的线性分类器	利用内核积函数代替一般的非线性映射	难以训练大规模的样本

## 3 基于NER的深度学习方法

### 3.1 深度学习优势

深度学习模型由多个处理层构成,每一层都由前向传播和反向传播组成的神经网络,神经网络通过前向和后向传播不断调整神经网络权重,当训练结果较好时停止迭代<sup>[27]</sup>。深度学习主要优势在于其表征学习能力及神经处理所带来的语义组合能力,这些优势允许机器输入原始数据并自动挖掘所需的潜在信息。

目前,NER深度学习方法与传统方法存在着较大区别,各方法优缺点如表4所示。深度学习技术应用于NER的优势包括以下3点:①NER模型得益于深度学习的非线性转换,可生成从输入到输出的非线性转换,相较于线性模型而言,基于深度学习的NER模型可学习到更复杂的特征;②特征工程需要大量的领域知识和工程技能,而深度学习模型可节省设计特征工程的时间,相较于人工构造特征可自动学习有用的文本信息;③深度NER模型可通过梯度下降在端到端的架构下训练,从而设计出更复杂的NER系统。

Table 4 Comparison of NER methods

表4 NER各方法比较

方法名称	工作原理	优点	缺点
基于规则的NER方法	基于领域专家人为设定的规则模版和词典,需要较少的标注数据,需要语言学相关知识	需要较少的标注数据,有较高的解释性,适用于特定领域的NER任务	不够灵活,可迁移性较差,词典需要不断进行更新维护
基于统计机器学习的NER方法	使用统计模型来学习文本中命名实体的模式	摆脱了对语言学知识的依赖,从标注数据中自主学习命名实体的相关特征	比较依赖标注的训练数据,需要大量人为设计的特征
基于深度学习的NER方法	借助神经网络模型,从大规模数据中学习命名实体的表示	不需要太多额外的特征,在大规模数据集上表现最好,具有较强的泛化能力	需要更多的计算资源和训练数据,可解释性较弱

### 3.2 基于深度学习的NER流程

基于深度学习的命名实体识别任务的主要流程如图 2 所示,分为以下 3 层:

(1)嵌入表示层。将输入文本序列转化为固定长度的向量表示,具体实现往往借助基于静态词向量的 Word2vec、Glove 等,或基于动态词向量的 Cove、ELMo 等。2018 年,以 GPT<sup>[28]</sup>、BERT<sup>[29]</sup> 为首的基于深层 Transformer 的语言表示模型出现后成为了词嵌入的主流方法。

(2)序列编码层。对词嵌入层向量进行语义编码,借助 CNN、RNN、BiLSTM、Transformer 等特征提取器进一步提取文本特征。

(3)标签解码层。将文本编码层向量输入解码网络以获得最佳的标签序列。

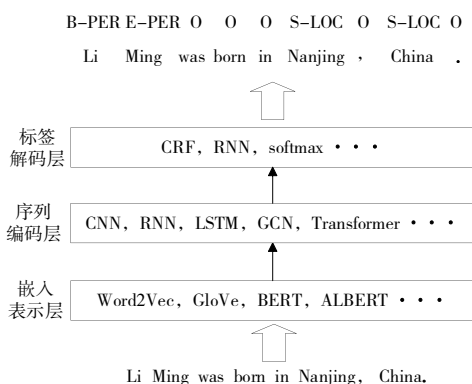


Fig. 2 NER process based on deep learning

图 2 基于深度学习的NER流程

### 3.3 嵌入表示层

对于 NER 任务,计算机无法直接识别输入语句中单词、词语和字符等标识符,所以应将标识符进行数值化表示。最早使用独热编码(One-Hot)表示词向量,这种编码方式原理十分简单,把词在词典中的位置作为向量,但词数量很大时维度也很高,并且每个向量之间彼此独立,所以实际应用较少。

神经网络最先在图像和语音识别等领域取得了良好的成绩,NLP 领域借鉴了该思想。Bengio 等<sup>[30]</sup>使用神经网络语言模型训练生成词向量,在一定程度上缓解了数据稀疏问题,但从词向量学习角度而言,基于神经网络语言模型的预训练方法对  $t$  时刻词进行预测时只考虑了历史词序列,无法很好地考虑上下文之间的语义信息。Mikolov 等<sup>[31]</sup>提出了新的词向量训练模型 Word2vec,基于词与词之间的信息实现词向量学习,提供了 CBOW 和 Skip-gram 两种训练方法。Pennington 等<sup>[32]</sup>针对 Word2vec 只能参考窗口上下文信息的缺点,结合矩阵分解思想提出了 Glove 模型。郑亚南等<sup>[25]</sup>在 Glove 模型后接入 SVM,实验表明训练得到的词向量效果更优。

在 NLP 领域,Word2vec、Glove 两种词向量称为静态词向量,因为生成的词向量矩阵无法根据具体语义动态修

改,因此无法解决一词多义的难题。此时,预训练语言模型(Pre-trained Language Model, PLM)应运而生,其生成的动态词向量有效弥补了静态词向量的缺点,目前 PLM 已成为目前主流的文本表示模型。Peters 等<sup>[33]</sup>提出 ELMo 预训练语言模型,使用 BiLSTM 抽取文本的语义信息。2018 年,OpenAI 公司提出 GPT 模型<sup>[28]</sup>,使用 Transformer 作为特征抽取器抽取文本信息,由于 Transformer 具有较高的并行程度,因此 GPT 相较于 ELMo 模型的信息抽取效率更优,但只实现了单向信息抽取<sup>[34]</sup>。Devlin 等<sup>[29]</sup>提出 BERT 模型有效结合了 ELMo 和 GPT 优点,BERT 使用 Transformer 编码器双向抽取文本语义特征,预训练任务包括掩码语言模型(Masked Language Model, MLM)和下一个句子预测(Next Sentence Predict, NSP),训练效果显著优于 GPT 和 ELMo。

此外,以 BERT 和 GPT 为代表的预训练语言模型为 NLP 领域带来了巨大变革,该类模型通常规模巨大,BERT 模型含有上亿个参数,GPT-3 模型参数量达到千亿级别<sup>[35]</sup>。面对参数量爆炸增长,研究者们开始思考如何在保障预训练模型效率的同时有效降低所需资源。Liu 等<sup>[36]</sup>对 BERT 模型进行优化提出了 RoBERTa 模型,在 BERT 模型的基础上引入动态掩码技术,采用跨文档整句输入并舍弃了 NSP 任务。Lan 等<sup>[37]</sup>针对 BERT 模型参数量过大的问题提出 ALBERT 模型,通过词向量因式分解、跨层参数共享技术有效降低了内存消耗,提升了模型训练速度,同时 ALBERT 使用句子顺序预测(Sentence Order Prediction, SOP)代替了 BERT 中的 NSP,使模型可学习到更多细微语义信息。后续, TinyBERT<sup>[38]</sup>、SpanBERT<sup>[39]</sup> 等预训练语言模型也相继被提出。

### 3.4 文本编码层

基于深度学习的 NER 模型的文本编码层主要基于卷积神经网络(Convolutional Neural Network, CNN)、循环神经网络(Recurrent Neural Network, RNN)、图卷积网络(Graph Convolutional Network, GCN)、Transformer 等。

#### 3.4.1 CNN

早期,CNN 在图形图像领域取得了重大突破,后面被研究者们引入了 NLP 领域。Collobert 等<sup>[40]</sup>基于 CNN 提出一种自然语言处理模型,采用无监督方式学习内部表示,无需人工构造特征即可处理 NER 在内的多种任务,但只使用了前馈神经网络提取每个单词周围固定窗口大小上下文信息,无法利用后缀等字符级特征。Yao 等<sup>[41]</sup>利用 CNN 实现生物医学领域的 NER,模型达到了较高的准确率,证明了深度学习方法在生物医学 NER 领域的有效性。Zhu 等<sup>[42]</sup>将 CNN 与注意力机制结合,分别抽取文本语义中的局部特征和全局特征,最后经过 BiGRU-CRF 层得到序列标注结果,在 MSRA 数据集的实验表明,该模型相较于 Baseline 模型 F1 值提升 2.65%。Gui 等<sup>[43]</sup>利用 CNN 对字符特征进行编码,引入一种重新思考机制有效解决了词汇冲突问题,并且模型推理速度也较优。刘宇鹏等<sup>[44]</sup>利用

CNN抽取文本特征提出基于BiLSTM-CNN-CRF的中文NER方法,通过CNN提取字级别的向量表示,使用BiLSTM提取上下文信息,最终基于CRF层解码得到标签序列,在CCKS2017医疗数据集的实验表明所提模型F1达到90.97%。卷积神经网络的主要优势是并行能力较强,每个时间状态不受之前时间状态影响,但无法很好地抽取文本的序列信息。

#### 3.4.2 RNN

RNN相较于CNN在NLP领域更受欢迎,RNN将文本语言视作序列数据,可处理序列数据,有效解决了CNN的弊端。在实际应用中,RNN的变体LSTM使用最多,其单元结构在RNN基础上增加了遗忘门、更新门和输出门,缓解了RNN梯度消失和梯度爆炸的问题。Huang等<sup>[45]</sup>首次使用BiLSTM-CRF模型处理NLP领域的序列标注任务,Word2vec-BiLSTM-CRF模型在当时取得了英文NER的最佳效果,模型鲁棒性得到了验证,后续在NER领域得到广泛使用。Liu等<sup>[46]</sup>提出一种中文细粒度命名实体识别模型En2BiLSTM-CRF,采用残差连接思想将两层双向LSTM的结果进行点对点合并,有效提升了NER模型的语义抽取能力,在CLUENER2020数据集上F1值达到87.2%。Gregoric等<sup>[47]</sup>提出一种新的NER架构,在模型输入端使用多个独立的双向LSTM单元,并采用模型间正则化来促进其多样化,在英文CoNLL-03数据集的F1值约为91.48%。2018年,Zhang等<sup>[48]</sup>首次提出Lattice LSTM中文NER模型,该模型相较于基于字符和基于词语的中文NER方法既可有效避免中文分词错误,还能自动从上下文中匹配词语信息,有效提升了NER效果,但无法并行化训练且只适配于LSTM,因此迁移性较差。Ma等<sup>[49]</sup>基于文献[48]方法保留了LSTM的单元结构,在模型输入层中融合词典信息,编码层除了双向LSTM外还额外扩充了CNN、Transformer。综上,循环神经网络主要优势是能较好地抽取文本的序列信息,但存在无法并行化的问题。

#### 3.4.3 GCN

CNN和RNN在NLP领域都取得了不错的效果,但无法有效处理图结构数据,而图神经网络能解决这类问题,其中图卷积神经网络在NER任务中使用最为广泛。Kipf等<sup>[50]</sup>首次提出图卷积神经网络的概念。Marcheggiani等<sup>[51]</sup>在Kipf基础上实现了语义角色标注,首次将GCN应用于NLP领域。Cetoli等<sup>[52]</sup>使用GCN处理NER任务,在经典模型BiLSTM-CRF中间额外添加了GCN层,还结合句法依存关系构造依赖图,GCN将结点信息传递到最近的邻居结点,实验表明该算法的F1值得到有效提升,证实了GCN在解决NER问题中的有效性。Tang等<sup>[53]</sup>将词汇信息与GCN相结合提出WC-GCN模型,将单词和字符看为图中结点,可同时处理两个方向的有向无环图,实验结果优于LSTM模型。荆鑫等<sup>[54]</sup>将GCN与ELMo模型相结合提出DTGCN网络模型,主要分为外部模块和内部模块,外部模

块主要处理字和句子,内部模块主要通过GCN对实体和实体关系进行建模,在自建的核心语料集的实验发现,所提模型相较于BiLSTM-CRF的P值、R值、F1值分别提升20.62%、11.69%、16.53%。

#### 3.4.4 Transformer

基于Transformer的方法主要代表为BERT类的预训练语言模型。BERT模型通过无监督训练从文本数据中学习丰富的语言表示,只需在其输出上添加合适的标签分类层就可处理命名实体识别任务,并通过微调使其适应特定的NER任务。后续研究者们常利用BERT等预训练模型生成动态词向量,同时结合CNN等神经网络模型处理NER任务。Souza等<sup>[55]</sup>提出BERT-CRF模型,利用BERT预训练模型生成词向量,同时结合CRF训练得到的约束条件实现命名实体识别。Li等<sup>[56]</sup>将外部词典信息融入BERT模型提出LEBERT-BiLSTM-CRF模型,相较于BERT-BiLSTM-CRF模型而言P值、R值、F1值分别提升1.02%、0.84%、0.93%,证实了引入外部词典的有效性。Xiong等<sup>[57]</sup>为自动化识别中国政府官方文件,基于ALBERT训练了专门处理中国政府文件的GovAlbert模型,取得了优秀的识别效果,侧面也证明了通用的预训练模型在特定领域具有局限性。李妮等<sup>[58]</sup>针对中文命名实体识别提出一种基于BERT-IDCNN-CRF的模型,重复使用自膨胀卷积的模块达到共享参数目的,避免了CNN过拟合问题,在MSRA语料上F1值达到了94.41%。Yao等<sup>[59]</sup>将ALBERT模型与注意力机制相结合,在BiLSTM和CRF层添加多头注意力层识别对制造文本的细粒度命名实体。

在医学领域,基于预训练模型的命名实体识别研究很多。Wen等<sup>[60]</sup>使用BERT对中医文本进行实体识别,证明了预训练模型的有效性。Lee等<sup>[61]</sup>针对生物医学领域提出预训练模型BioBERT,除了BERT模型的语料库还额外添加了生物医学领域的文本库进行预训练,增强了模型在生物医学领域的适用性。Zhu等<sup>[62]</sup>针对中文电子病历的NER提出BSBC模型框架,将BERT和多个堆叠的BiLSTM相结合,最后经过CRF解码层得到标签序列,这是首次将堆叠BiLSTM用于医学NER任务。Wu等<sup>[63]</sup>基于RoBERTa预训练模型学习医学文本特征,利用双向LSTM提取偏旁部首等特征并拼接RoBERTa所生成的特征向量,最后经过CRF层进行序列解码,该模型在CCKS2017和CCKS2019数据集的F1值分别达到93.26%、82.87%。综上,本文针对文本编码层中的几种方法进行比较分析,如表5所示。

### 3.5 标签解码层

标签解码层作为NER深度学习模型的最后一层,将文本编码层所生成的上下文表示作为输入并生成最终的标签序列,常用于多层感知机、循环神经网络、条件随机场(Conditional Random Field,CRF)等。在深度学习盛行的时代背景之下CRF使用最广泛,在机器学习时代就在NER任务中取得了不错的效果,但现在更多作为NER深度学习

Table 5 Comparison of text encoding layer methods

表5 文本编码层方法比较

代表模型	优点	缺点
CNN	可以对数据并行化处理,良好地处理高维数据	不能完整地抽取序列信息
RNN	解决了CNN无法存储序列信息的问题	容易产生梯度消失和梯度爆炸的问题
LSTM	解决了RNN的长期依赖问题	没有完全解决梯度问题
GCN	图结构的数据模型可以更好地挖掘实体之间的联系	模型比较笨重并且可扩展性差
Transformer方法	对Transformer进行预训练,生成深层的语言表征	需要训练数据和高性能的GPU

模型的标签解码层。CRF作为一种生成图模型有效考虑了序列标签间的约束关系,在NER任务中取得了良好的效果。Yi等<sup>[64]</sup>通过深入研究网络安全实体提出RDF-CRF模型,在面对网络安全数据集较少的情况,将CRF和规则模版、中文词典相结合,有效提升了安全实体识别的准确率和召回率。Lee等<sup>[65]</sup>借助CRF实现了问答领域的细粒度命名实体识别,并使用最大熵对命名实体进行分类。Lin等<sup>[66]</sup>在双向LSTM层后接入CRF作为解码层,提出了一种可靠性感知名称标记模型,有效提升了词嵌入的质量。Ghaddar等<sup>[67]</sup>在双向LSTM后接入CRF作为解码层,在CoNLL03和OntoNotes5.0数据集上F1值分别达到91.73%、87.95%。Strubell等<sup>[68]</sup>在ID-CNN编码层后接入CRF作为标签解码层,在CoNLL03数据集上的F1值达到90.65%。喻金平等<sup>[69]</sup>提出基于RoBERTa-wwm-BiLSTM-CRF的实体识别模型,在自建的扶持政策数据集上F1值达到91.7%。

## 4 结语

本文对命名实体识别的背景、定义、实现工具、常用数据集、评估指标和命名实体识别方法进行了阐述和归纳,使相关学者了解到命名实体识别的基本信息和发展历程。具体将命名实体识别技术的发展分为以下3个阶段:①基于规则的命名实体识别方法;②基于统计机器学习的命名实体识别方法;③基于深度学习的命名实体识别方法。虽然命名实体识别方法日渐成熟,命名实体识别的研究工作取得了显著成效,但仍然存在许多不足之处,有待进一步完善,后续研究可从以下3个方面入手:

(1)细粒度命名实体识别研究。传统命名实体识别只识别人名、地名等实体类别,实体种类较少且划分较为宽泛。细粒度命名实体识别更符合真实世界的应用场景,良好的细粒度NER模型可帮助学者更准确地提取信息,也可提升搜索引擎准确性。Xu等<sup>[70]</sup>发布的CLUENER2020数据集包含10种不同类型的命名实体,该数据集为中文细粒度NER研究提供了重要支撑,但目前细粒度命名实体识别研究工作主要存在数据稀疏和歧义性问题。数据稀疏

问题主要指细粒度NER的数据标注工作更困难,需要更多细节和领域知识;歧义性问题主要指一些特定词汇含有多种不同含义,必须充分了解上下文信息才可解决歧义性问题。为此,可考虑采用数据增强技术扩充细粒度NER任务的标注数据,也可考虑结合知识图谱提供实体的上下文信息和相关领域知识。

(2)NER模型的迁移能力研究。无论是中文领域还是英文领域,现有NER模型都在部分数据集上取得了优异的表现,但常常因标签不匹配、场景不一致等问题在一个数据集表现优异的模型很难直接应用于其他领域,各领域之间的差异也导致难以实现资源共享<sup>[71]</sup>。为此,针对不同语言种类,可使用BERT等预训练模型的多语言版本帮助模型理解多种语言的共同特征;针对不同领域,选择合适的迁移学习策略也十分重要,由于不同方法适用于不同的迁移场景,具备良好迁移性的NER模型可在不同场景中重复使用,以有效提升模型的可用性和资源利用率,因此寻求更优的方法来提升模型的可迁移性是该领域的重要研究方向。

(3)多任务学习的NER研究。多任务学习是将NER任务和其他任务合并在一起进行模型训练,可在提升模型性能的同时减少开发成本,并在数据集稀疏的情况下多任务学习可缓解模型过拟合,提高模型泛化能力。命名实体识别的好坏将直接影响各项下游任务的表现,传统的管道模型很容易造成错误传递,而多任务学习可共享信息,以有效解决错误传递等问题<sup>[72]</sup>。例如,在问答模型的构造中通常将实体识别和关系识别两项任务同时进行,但在模型的训练过程中会出现数据不平衡的问题,可考虑为不同任务分配不同权重以减轻某些任务对模型性能的干扰。综上,利用多任务学习进一步研究NER也是未来的研究方向。

## 参考文献:

- [1] PAN Z G. Research on the recognition of Chinese named entity based on rules and statistics[J]. Information Science, 2012, 30(5): 708-712, 786. 潘正高. 基于规则和统计相结合的中文命名实体识别研究[J]. 情报科学, 2012, 30(5): 708-712, 786.
- [2] GRISHMAN R, SUNDHEIM B. Message understanding conference-6: a brief history [C]// Proceedings of the 16th Conference on Computational Linguistics, 1996: 466-471.
- [3] SAJU C J, SHAJA A S. A survey on efficient extraction of named entities from new domains using big data analytics [C]// 2017 Second International Conference on Recent Trends and Challenges in Computational Models, 2017: 170-175.
- [4] DAI X. Recognizing complex entity mentions: a review and future directions [C]// Proceedings of ACL 2018, Student Research Workshop, 2018: 37-44.
- [5] LI J, SUN A X, HAN J L, et al. A survey on deep learning for named entity recognition [J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(1): 50-70.
- [6] PAN J, LI M P, WANG X M. A review of Chinese named entity recognition using deep learning [J]. Digital Library Forum, 2023, 19(5): 1-9. 潘俊,李萌配,王贤明. 应用深度学习的中文命名实体识别研究综述

- [J]. 数字图书馆论坛, 2023, 19(5): 1-9.
- [7] QI P N, LIAO Y L, QIN B. A review of Chinese-named entity recognition based on deep learning[J]. Journal of Chinese Computer Systems, 2023, 44(9): 1857-1868.  
祁鹏年, 廖雨伦, 覃颀. 基于深度学习的中文命名实体识别研究综述[J]. 小型微型计算机系统, 2023, 44(9): 1857-1868.
- [8] GAO X, WANG S, ZHU J W, et al. Overview of named entity recognition tasks[J]. Computer Science, 2023, 50(S1): 26-33.  
高翔, 王石, 朱俊武, 等. 命名实体识别任务综述[J]. 计算机科学, 2023, 50(S1): 26-33.
- [9] RAU L F. Extracting company names from text[C]// Proceedings of the 7th IEEE Conference on Artificial Intelligence Application, 1991: 29-32.
- [10] YUAN J D, PAN M M, ZHANG T, et al. Electricity safety domain named entity recognition based on rules and dictionaries[J]. Electronic Technology Applications, 2022, 48(12): 22-27.  
袁金斗, 潘明明, 张腾, 等. 基于规则和词典的用电安全领域命名实体识别[J]. 电子技术应用, 2022, 48(12): 22-27.
- [11] SHAALAN K, RAZA H. NERA: named entity recognition for arabic[J]. Journal of the American Society for Information Science and Technology, 2009, 60(8): 1652-1663.
- [12] YAN P. Research on Chinese-named entity recognition based on the combination of rules and probability and statistics[J]. Computer and Digital Engineering, 2011, 39(9): 88-91.  
闫萍. 基于规则和概率统计相结合的中文命名实体识别研究[J]. 计算机与数字工程, 2011, 39(9): 88-91.
- [13] BAO Z S, SONG B Y, ZHANG W B, et al. Research on named entity recognition of ancient Chinese medicine books based on semi-supervised learning and rules[J]. Journal of Chinese Information Technology, 2022, 36(6): 90-100.  
包振山, 宋秉彦, 张文博, 等. 基于半监督学习和规则相结合的中医古籍命名实体识别研究[J]. 中文信息学报, 2022, 36(6): 90-100.
- [14] LIU H B, ZHANG D M, XIONG S F, et al. Named entity recognition of wheat pest naming fusing ALBERT with rules[J]. Computer Science and Exploration, 2023, 17(6): 1395-1404.  
刘合兵, 张德梦, 熊蜀峰, 等. 融合ALBERT与规则的小麦病虫害命名实体识别[J]. 计算机科学与探索, 2023, 17(6): 1395-1404.
- [15] QUIMBAYA A P, SIERRA A, GONZALEZ R, et al. Named entity recognition over electronic health records through a combined dictionary-based approach[J]. Procedia Computer Science, 2016, 100: 55-61.
- [16] BIKEL D M, SCHWARTZ R, WEISCHEDEL R M. An algorithm that learns what's in a name[J]. Machine Learning, 1999, 34: 211-231.
- [17] ZHOU G D, SU J. Named entity recognition using an HMM-based chunk tagger[C]// Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 2002: 473-480.
- [18] YU H K, ZHANG H P, LIU Q, et al. Chinese named entity recognition based on cascading hidden Markov model[J]. Journal on Communications, 2006(2): 87-94.  
俞鸿魁, 张华平, 刘群, 等. 基于层叠隐马尔可夫模型的中文命名实体识别[J]. 通信学报, 2006(2): 87-94.
- [19] ZHANG H P, LIU Q. Research on automatic recognition of Chinese names based on role labeling[J]. Chinese Journal of Computers, 2004(1): 85-91.  
张华平, 刘群. 基于角色标注的中国人名自动识别研究[J]. 计算机学报, 2004(1): 85-91.
- [20] BORTHWICK A, STERLING J, AGICHTEIN E, et al. NYU: description of the MENE named entity system as used in MUC-7[C]// Fairfax: Seventh Message Understanding Conference, 1998: 1-7.
- [21] BENDER O, OCH F J, NEY H. Maximum entropy models for named entity recognition[C]// Proceedings of the 7th Conference on Natural Language Learning, 2003: 148-151.
- [22] WANG J W. Chinese named entity recognition based on the maximum entropy model[D]. Nanjing: Nanjing University of Science and Technology, 2005.  
王江伟. 基于最大熵模型的中文命名实体识别[D]. 南京: 南京理工大学, 2005.
- [23] ZHANG Y J, XU Z T, XUE X Y. A Chinese named entity recognition model with maximum entropy of multiple features[J]. Computer Research and Development, 2008(6): 1004-1010.  
张玥杰, 徐智婷, 薛向阳. 融合多特征的最大熵汉语命名实体识别模型[J]. 计算机研究与发展, 2008(6): 1004-1010.
- [24] ISOZAKI H, KAZAWA H. Efficient support vector classifiers for named entity recognition[C]// Taipei: Proceedings of the 19th International Conference on Computational Linguistics, 2002.
- [25] ZHENG Y N, TIAN D G. Text classification research based on GloVe and SVM[J]. Software Guide, 2018, 17(6): 45-48, 52.  
郑亚南, 田大纲. 基于GloVe与SVM的文本分类研究[J]. 软件导刊, 2018, 17(6): 45-48, 52.
- [26] CHEN X, LIU H, CHEN Y Q. Identification of the name of the Chinese organization based on the support vector machine method[J]. Computer Application Research, 2008(2): 362-364, 367.  
陈霄, 刘慧, 陈玉泉. 基于支持向量机方法的中文组织机构名的识别[J]. 计算机应用研究, 2008(2): 362-364, 367.
- [27] LECUN Y, BENGIO Y, HINTON J. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [28] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training[EB/OL]. [https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf).
- [29] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171-4186.
- [30] BENGIO Y, DUCHARME R, VINCENT P, et al. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003, 3: 1137-1155.
- [31] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[DB/OL]. <https://arxiv.org/pdf/1301.3781.pdf>.
- [32] PENNINGTON J, SOCHER R, MANNING C. GloVe: global vectors for word representation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014: 1532-1543.
- [33] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018: 2227-2237.
- [34] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31th International Conference on Neural Information Processing Systems, 2017: 6000-6010.
- [35] BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners[DB/OL]. <https://arxiv.org/pdf/2005.14165v4.pdf>.
- [36] LIU Y, OTT M, GOYAL N, et al. RoBERTa: a robustly optimized BERT pretraining approach[DB/OL]. <https://arxiv.org/pdf/1907.11692v1.pdf>.
- [37] LAN Z, CHEN M, GOODMAN S, et al. ALBERT: a lite BERT for self-supervised learning of language representations[DB/OL]. <https://arxiv.org/pdf/1909.11942v6.pdf>.
- [38] JIAO X, YIN Y, SHANG L, et al. TinyBERT: distilling BERT for natural language understanding[DB/OL]. <https://arxiv.org/abs/1909.10351>.
- [39] JOSHI M, CHEN D, LIU Y, et al. SpanBERT: improving pre-training

- by representing and predicting spans[J]. Transactions of the Association for Computational Linguistics, 2020, 8: 64–77.
- [40] COLBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch[J]. Journal of Machine Learning Research, 2011, 12: 2493–2537.
- [41] YAO L, LIU H, LIU Y, et al. Biomedical named entity recognition based on deep neural network[J]. International Journal of Hybrid Information Technology, 2015, 8(8): 279–288.
- [42] ZHU Y, WANG G. CAN-NER: convolutional attention network for chinese named entity recognition[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 3384–3393.
- [43] GUI T, MA R, ZHANG Q, et al. CNN-Based Chinese NER with lexicon rethinking[C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence Main track, 2019: 4982–4988.
- [44] LIU Y P, LI D D. Chinese named entity recognition method based on BLSM-CN-CRF[J]. Journal of Harbin University of Science and Technology, 2020, 25(1): 115–120.  
刘宇鹏, 栗冬冬. 基于BLSTM-CNN-CRF的中文命名实体识别方法[J]. 哈尔滨理工大学学报, 2020, 25(1): 115–120.
- [45] HUANG Z, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging[DB/OL]. <https://arxiv.org/pdf/1508.01991v1.pdf>.
- [46] LIU J, XIA C, YAN H, et al. Innovative deep neural network modeling for fine-grained Chinese entity recognition[J]. Electronics, 2020, 9(6): 1001.
- [47] GREGORIC A, BACHRACH Y, COOPE S. Named entity recognition with parallel recurrent neural networks[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018: 69–74.
- [48] ZHANG Y, YANG J. Chinese NER using lattice LSTM[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018: 1554–1564.
- [49] MA R, PENG M, ZHANG Q, et al. Simplify the usage of lexicon in Chinese NER[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020: 5951–5960.
- [50] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[DB/OL]. <https://arxiv.org/pdf/1609.02907.pdf>.
- [51] MARCHEGGIANI D, TITOV I. Encoding sentences with graph convolutional networks for semantic role labeling[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017: 1506–1515.
- [52] CETOLI A, BRACAGLIA S, OHRNEY A, et al. Graph convolutional networks for named entity recognition[C]// Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories, 2017: 37–45.
- [53] TANG Z, WAN B, YANG L. Word-Character graph convolution network for Chinese named entity recognition[J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2020, 28: 1520–1532.
- [54] JING X, WANG H F, LIU Q F, et al. Named entity recognition in nuclear power field based on ELMo-GCN[J]. Journal of Beihang University, 2022, 48(12): 2556–2565.  
荆鑫, 王华峰, 刘潜峰, 等. 基于ELMo-GCN的核电领域命名实体识别[J]. 北京航空航天大学学报, 2022, 48(12): 2556–2565.
- [55] SOUZA F, NOGUEIRA R, LOTUFO R. Portuguese named entity recognition using BERT-CRF[DB/OL]. <https://arxiv.org/pdf/1909.10649.pdf>.
- [56] LI S, BAO Z, ZHAO S, et al. A LEBERT-based model for named entity recognition[C]// Proceedings of the 2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture, 2021: 980–983.
- [57] XIONG Z, KONG D, XIA Z, et al. Chinese government official document named entity recognition based on albert[C]// Proceedings of the 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics, 2021: 350–354.
- [58] LI N, GUAN H M, YANG P, et al. Chinese named entity recognition method based on BERT-IDCNN-CRF[J]. Journal of Shandong University (Science Edition), 2020, 55(1): 102–109.  
李妮, 关焕梅, 杨飘, 等. 基于BERT-IDCNN-CRF的中文命名实体识别方法[J]. 山东大学学报(理学版), 2020, 55(1): 102–109.
- [59] YAO L, HUANG H, WANG K, et al. Fine-grained mechanical chinese named entity recognition based on ALBERT-AttBiLSTM-CRF and transfer learning[J]. Symmetry, 2020, 12(2): 1986–2006.
- [60] WEN S, ZENG B, LIAO W. Named entity recognition for instructions of Chinese medicine based on pre-trained language model[C]// Proceedings of the 3rd International Conference on Natural Language Processing, 2021: 139–144.
- [61] LEE J, YOON W, KIM S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining[J]. Bioinformatics, 2020, 36(4): 1234–1240.
- [62] ZHU Z, LI J, ZHAO Q, et al. Medical named entity recognition of Chinese electronic medical records based on stacked Bidirectional Long Short-Term Memory[C]// Proceedings of the 2021 IEEE 45th Annual Computers, Software, and Applications Conference, 2021: 1930–1935.
- [63] WU Y, HUANG J, XU C, et al. Research on named entity recognition of electronic medical records based on RoBERTa and Radical-Level Feature[J]. Wireless Communications and Mobile Computing, 2021(10): 1–10.
- [64] YI F, JIANG B, WANG L, et al. Cybersecurity named entity recognition using multi-modal ensemble learning[J]. IEEE Access, 2020, 8: 63214–63222.
- [65] LEE C, HWANG Y, OH H, et al. Fine-grained named entity recognition using conditional random fields for question answering[C]// Proceedings of the Third Asia Conference on Information Retrieval Technology, 2006: 581–587.
- [66] LIN Y, LIU L, JI H, et al. Reliability-aware dynamic feature composition for name tagging[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 165–174.
- [67] GHADDAR A, LANGLAIS P. Robust lexical features for improved neural network named-entity recognition[C]// Proceedings of the 27th International Conference on Computational Linguistics, 2018: 1896–1907.
- [68] STRUBELL E, VERGA P, BELANGER D, et al. Fast and accurate entity recognition with iterated dilated convolutions[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017: 2670–2680.
- [69] YU J P, ZHU W F, LIAO L F. Research on text entity recognition of support policy based on RoBERTa-wwm-BiLSTM-CRF[J]. Computer Engineering and Science, 2023, 45(8): 1498–1507.  
喻金平, 朱伟锋, 廖列法. 基于RoBERTa-wwm-BiLSTM-CRF的扶持政策文本实体识别研究[J]. 计算机工程与科学, 2023, 45(8): 1498–1507.
- [70] XU L, TONG Y, DONG Q, et al. CLUENER2020: fine-grained named entity recognition dataset and benchmark for Chinese[DB/OL]. <https://arxiv.org/ftp/arxiv/papers/2001/2001.04351.pdf>.
- [71] AGARWAL O. Towards robust named entity recognition via temporal domain adaptation and entity context understanding[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2022: 12866–12867.
- [72] YU J K. Chinese electronic medical record naming entity recognition based on multi-task learning[J]. Software Guide, 2021, 20(11): 42–46.  
余俊康. 多任务学习的中文电子病历命名实体识别研究[J]. 软件导刊, 2021, 20(11): 42–46.