

# 小样本命名实体识别综述

**摘要：**命名实体识别是自然语言处理领域中的重要任务之一，其在信息抽取、问答系统和机器翻译等应用中具有广泛的应用。然而，由于缺乏人工标记数据，低资源和小样本的命名实体识别成为研究中的重要挑战。本文从传统命名实体识别方法开始，介绍了基于规则和字典的方法以及基于统计机器学习的方法。然后，重点介绍了基于深度学习的命名实体识别方法，并讨论了低资源和小样本命名实体识别的相关方法。最后，对目前的研究进行总结，并展望了未来研究的方向和挑战。

**关键词：**命名实体识别；深度学习；小样本

**Abstract:** Named Entity Recognition is an important task in the field of Natural Language Processing, and it has wide applications in information extraction, question answering systems, and machine translation, among others. However, due to the lack of annotated data, low-resource and small-sample NER has become a significant challenge in research. This paper provides an overview of research on NER in the context of insufficient labeled data. It starts by introducing traditional NER methods, including rule-based and dictionary-based approaches, as well as statistical machine learning methods. Then, it focuses on deep learning-based NER methods and discusses relevant approaches for low-resource and small-sample NER. Finally, it summarizes the current research and presents future research directions and challenges.

**Keywords:** Named Entity Recognition, Deep Learning, Few-Shot

# 引言

命名实体识别（Named Entity Recognition, NER）是自然语言处理的一项基本任务[1]。主要是将非结构化文本中的人名、地名、机构名和具有特定意义的实体抽取出来并加以归类，进而组织成半结构化或结构化的信息，再利用其他技术对文本实现分析和理解目的，这对于文本的结构化起着至关重要的作用。命名实体识别技术在信息抽取、信息检索、问答系统等多种自然语言处理技术领域有着广泛的应用。

命名实体识别研究历史最早可以追溯到 1991 年，Rau[2]在第七届 IEEE 人工智能应用会议上发表了“抽取和识别公司名称”的有关研究文章，陆续出现一些有关名词识别的研究。1996 年，“命名实体（Named Entity, NE）”一词首次用于第六届信息理解会议（MUC-6）[3]，会议将命名实体评测作为信息抽取的一个子任务。随后出现了一系列信息抽取的国际评测会议，诸如 CONLL、IEER-99、ACE 等，这些评测会议对命名实体识别的发展有极大的推动作用。

命名实体识别的主要技术主要有三类：基于规则和字典的方法、基于统计机器学习的方法、基于深度学习的方法。目前较为流行的是基于深度学习的方法，本文主要介绍基于深度学习的各类方法，尤其是针对标注数据不充分情况下的解决方案。

## 1 研究背景

### 1.1 命名实体识别

命名实体识别（named entity recognition, NER）是自然语言处理（natural language processing, NLP）的一项非常重要的基础任务，旨在自动检测文本中的命名实体并将其分类为预定义实体类型，例如人名、地名、组织机构名等，NER 是人机对话系统、机器翻译、文本摘要等的前置任务。命名实体可能有多种表现形式——全名、部分名称、别名和缩写——这使得实体链接成为自然语言处理中一项非常具有挑战性的任务。

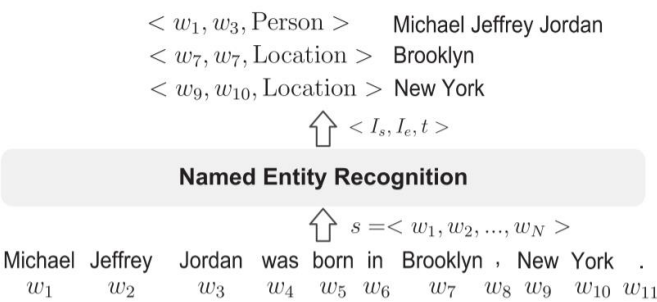


图 1-1 命名实体识别的目标

图 1-1 展示了 NER 任务的目标，模型的输入为一个句子  $\mathbf{x} = \{w_1, w_2, \dots, w_N\}$ ，其中  $w_i$  代表句中的第  $i$  个单词。NER 任务的目标就是给出命名实体的起始和终止边界，并给出该命名实体的类别，如图中的 Michael Jeffrey Jordan 就是一个命名实体，它的起止位置为  $[w_1, w_3]$ ，实体类型为 Person。

NER 任务的基本解决方法可以抽象为序列分类任务，并且通常使用 BIO（开头、内部、其他）或 BIOES（开头、内部、结束、其他）等标注预测方法。在这里，我们将重点介绍 BIO 标注方法。BIO 标注方法是给句子中的每个单词分配一个标签，该标签由两部分

组成。首先，标签表示单词在实体中的位置。标签中的 **B** 表示词是实体的第一个词，**I** 表示词是实体的中间词，**O** 表示词不是实体的一部分。其次，标签表示词对应的实体类型。例如，在上述针 **NER** 任务的示例中，我们需要预测词是属于人名 (**Person**) 还是地点 (**Location**) 等实体类型。因此，最终，每个词都会被标注为 **BIO** 加上实体类型的形式，这使得 **NER** 任务成为了一个文本序列分类任务。

## 1.2 低资源命名实体识别

尽管早期以及深度 **NER**[4-29]方法已经取得了非常高的识别精度，但是训练模型需要大规模标注数据，模型性能与标注数据量成正比，在训练语料匮乏的特定领域和小语种上，性能差强人意。由于训练集和测试集要求独立同分布，因此导致将已有的模型运用到其他领域或者语言上性能差强人意。

因此低资源情形下的命名实体识别任务亟待解决。其中低资源可以体现在两个方面：一种是同领域的，训练集和测试集的数据为同领域，但是标注数据较少；一种是跨领域的，训练集和测试集的数据属于不同领域，且标注数据较少。更极端的情况下，标注数据只有几个样本，这种情况被称为小样本 **NER**。

## 1.3 小样本命名实体识别

小样本命名实体识别任务是指在某个领域或场景下，只有极少量带有标注的样本数据可用于训练的命名实体识别 (**NER**) 任务。通常情况下，小样本指的是标注样本数量远远少于需要进行实体识别的目标领域或场景中的实体数目。在这种任务中，由于标注样本的稀缺性，传统的机器学习方法往往难以取得令人满意的结果。因此，针对小样本 **NER** 任务，需要采用特定的技术和方法，例如迁移学习、自适应学习、生成对抗网络 (**GAN**) 等，以充分利用有限的样本数据并提高 **NER** 模型的性能和泛化能力。小样本 **NER** 任务的关键挑战在于如何在缺乏大规模标注样本的情况下，有效地识别和推断出目标领域或场景中的命名实体，以满足实际应用的需求。

# 2 传统命名实体识别方法

## 2.1 基于规则和字典的方法

基于规则和字典的方法是命名实体识别中最早使用的方法。基于规则的方法要构造大量的规则集，规则集的构建大多采用语言学专家手工构造的规则模板。基于词典的方法需要建立命名实体词典，命名实体识别的过程就是在字典或专业领域知识库中查找的过程。早期的命名实体识别任务大多采用基于规则和词典的方法实现。

规则可以基于特定领域的地名词典[4]和句法-词汇模式[5]来设计。Kim[6]提出使用 Brill 规则推理方法来处理语音输入。该系统基于 Brill 的部分语音标记器自动生成规则。在生物医学领域，Hanisch 等人[7]提出了 ProMiner，它利用预处理的同义词词典来识别生物医学文本中的蛋白质提及和潜在基因。Quimbaya 等人[8]提出了一种基于词典的电子健康记录 **NER** 方法。其他一些著名的基于规则的 **NER** 系统包括 LaSIE-II[9]、NetOwl[10]、Facile[11]、

SAR[12]、FASTUS[13]和 LTG[14]系统，这些系统主要是基于手工制作的语义和句法规则来识别实体。

基于手工规则的方法虽然在信息详尽的小数据集上可达到较高的准确率，但面对大量的数据集以及特定领域，这种方式变得不再可取，旧的规则不适用于新命名实体识别词汇，规则之间常有冲突，有较大的局限性；新的规则库和词典的建立需要花费大量的时间及人力，并且这些系统难以转移到其他领域。因此这种方法逐渐被后来的基于统计的机器学习方法所替代。

## 2.2 基于统计机器学习的方法

随着许多基于机器学习的发展，机器学习方法被广泛用于 NER 任务。基于统计的机器学习方法是利用人工标注的语料进行有监督训练，NER 任务被抽象成一个多类分类或序列标注任务，给定带标注的数据样本，精心设计能代表训练样本的特征，然后利用训练好的机器学习模型实现标签的预测。

特征工程在有监督的 NER 系统中至关重要。特征向量表示是对文本的抽象，其中一个词由一个或多个布尔值、数字或名义值表示[15]。词级特征、列表查询特征（例如，维基百科地名和 DBpedia 地名）[16]，以及文档和语料库特征[17]已被广泛用于各种监督下的 NER 系统中。基于这些特征，许多机器学习算法被应用于有监督 NER，包括隐马尔可夫模型(HMM)[18]、决策树[19]、最大熵模型[20]、支持向量机(SVM)[21]和条件随机域(CRF)[22]。

Bikel 等人[23]提出了第一个基于 HMM 的 NER 系统，名为 *IdentiFinder*，用于识别和分类名称、日期、时间表达和数字量。此外，Szarvas 等人[24]通过使用 C4.5 决策树和 AdaBoostM1 学习算法开发了一个多语言 NER 系统，它的优势是通过不同的特征子集来训练几个独立的决策树分类器，然后通过投票法作出最终决定。

McNamee 和 Mayfield[25]使用 1,000 个语言相关特征和 258 个正字法和标点符号来训练 SVM 分类器，在预测实体标签时，SVM 不考虑 "相邻" 的单词，CRFs 考虑了上下文。Krishnan 和 Manning[26]提出了一种基于两个耦合的 CRF 分类器的两阶段方法。第二个 CRF 利用了从第一个 CRF 的输出中得到的潜在表示。我们注意到基于 CRF 的 NER 已经被广泛地应用于各个领域的文本，包括生物医学文本[27]、推特[28]和化学文本[29]。

基于统计机器学习的方法与之前的方法相比，效果上有了明显的提高，但是也需要具有专业领域知识的人进行大量人工标注，人工和时间的成本很高。

## 3 基于深度学习的命名实体识别方法

随着深度学习的兴起，使用深度学习方法解决命名实体识别问题成为了研究热点。该类方法的优势在于神经网络模型可以自动学习句子特征，无需复杂的特征工程。

解决 NER 问题最经典的深度学习模型结构是百度[30]于 2015 年提出的 LSTM-CRF 架构。传统的 LSTM 模型本身可以完成序列标注任务，通过对输入句子进行编码，可以得到每个单词在各个分类上的打分结果。然而，仅使用 LSTM 存在一个问题，即无法学习到输出标签之间的依赖关系。例如，在输出标签中，B-person（表示人名的开头）通常紧随其后的是 I-person（表示人名的中间部分），这种规律在 LSTM 中难以捕捉。因此，通常在 LSTM 之后添加 CRF 层。CRF（条件随机场）指的是在给定隐变量序列和观测序列的情况下，每个观测值仅与当前时刻的隐变量以及前一时刻的观测值相关联。CRF 的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。在 LSTM-CRF 中，CRF

模型对 NER 标签之间的转移关系进行建模，弥补了 LSTM 在这方面的不足。CRF 的核心作用是建模标签之间的依赖关系。LSTM-CRF 模型的整体结构如下图所示。

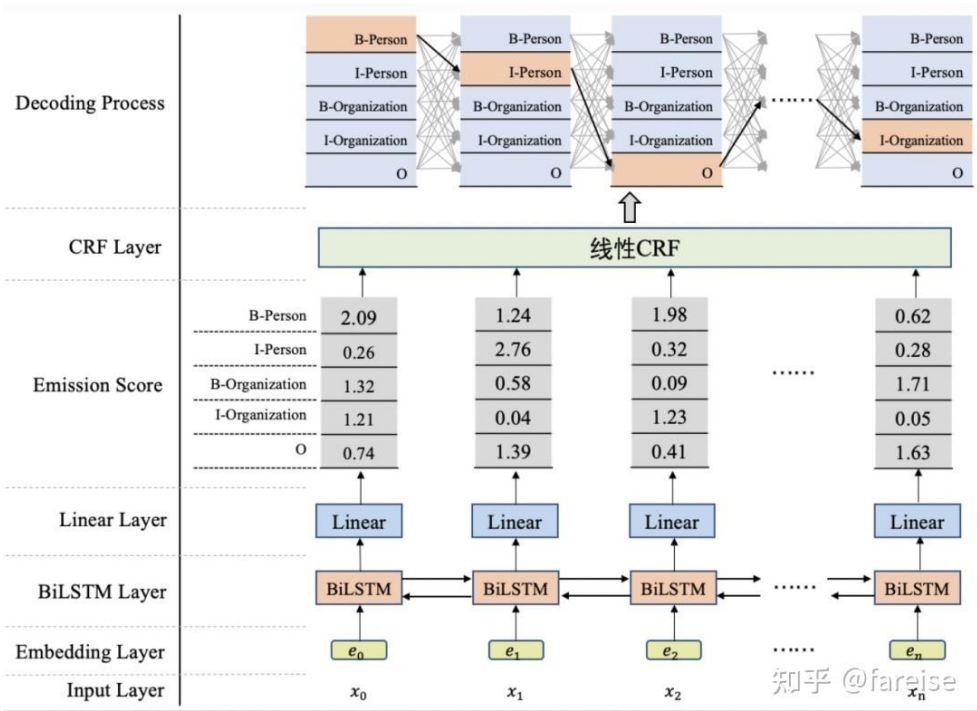


图 3-1 LSTM-CRF 模型结构图

后续很多工作都是在 LSTM-CRF 模型结构的基础上进行改进，例如将 LSTM 编码器替换为 Bert，或者将 CRF 替换成 Softmax。编码器部分，Souza 等[31]采用了 Bert-CRF 的模型结构解决 NER 任务。对于 CRF 部分，Cui 等[32]提出采用 label embedding 结合多层 attention 机制学习各个位置 label 之间的关系，相比 CRF 模型，这种方法可以建模更复杂的 label 之间关系。

## 4 低资源的命名实体识别方法

尽管近年来基于深度学习的 NER 迅速发展，但它们仍然面临着数据不足的挑战。

基于神经网络的模型可以从数据中自动学习的特征，对新实体有很好的概括性。深度学习模型通常对训练数据有很高的要求，这些模型的性能在很大程度上依赖于标注数据的数量和质量。然而，当语料库很小的时候，特别是在低资源的情况下[33]，这些方法的性能会明显下降，因为隐藏的特征表示不能被充分学习。同时，在许多领域，用于标注的注释预算远远低于未标注数据的总数。例如，在医学领域，有大量的未标注的医疗记录、临床报告和生物医学文献。对这些数据进行标注是相当繁重的，因为它需要扎实的医学知识背景，而这只能由医学专家来完成。这种标记数据不足的情况在许多现实世界的应用中普遍存在。为了使深度 NER 模型有更广泛的用途，减少其对标记数据的依赖性，近年来研究人员尝试了各种无监督、半监督或远监督的方法。

Wang 等人[34]使用间接监督方法，将概率逻辑和深度学习相结合，对整个数据集进行联合推理，从而摆脱了对标记样本的依赖。Chen 等人[35]提出了一种基于局部可加性的半监督 NER 数据增强(LADA)方法，该方法通过对彼此接近的序列插值来生成虚拟标记样本。

基于远监督的方法近年来也被应用在命名实体识别，其目的是通过搜索词典或知识库中的提及信息来减少昂贵的人力资源。Yang 等[36]利用基于字典的匹配自动生成训练样本，解



决了 NER 中标注不完全的问题。Shang 等[37]采用了远监督的方法，利用知识库等外部资源对特定类型的实体进行自动标注。然而，这些远监督方法的有效性通常依赖于领域词典或知识库的可用性和质量，而这些词典或知识库在许多领域可能是不可用的。此外，词汇外(OOV)词和提及链接中的词义消歧问题也限制了远程监督在实际 NER 应用中的表现。

## 5 小样本命名实体识别方法

当有足够的标记数据时，基于深度学习的方法【】(Huang et al. , 2015; Ma and Hovy, 2016; Lample et al. , 2016; Chiu and Nichols, 2016)可以获得相当不错的性能。但在实际应用中，很多情况下希望能够识别在训练中没见过的新类别。然而，为这些新类型收集额外的标记数据会耗费大量的时间和人力成本。能够根据目标域中少量的标记样本来识别新的实体类型的方法被称为小样本命名实体识别(Fritzler 等, 2019; Yang 和 Katiyar, 2020)，小样本命名实体识别在近年来引起了很大的研究兴趣。

小样本命名实体识别提出了一些独特的挑战，使其比其他 Few-shot 学习任务困难得多。当在某个特定领域中拥有大量 NER 任务的标注样本，但在目标领域中只有少量样本可用时，迁移学习技术可以用于提升 NER 效果。然而在 NER 中，不涉及任何已定义实体集的令牌被标记为 Outside (O)，因此，在训练实体集中标记为 O 的令牌可能对应于测试集中的有效目标实体。迁移学习的方法是在源领域的大量标注样本数据上进行预训练，然后在目标领域上进行微调。然而，在 NER 问题中，不同的场景需要预测不同的实体类型，这导致无法直接进行迁移。其次，作为序列标记任务，NER 需要根据一致性上下文以及标签内部的依赖关系分配标签[38]，这也给小样本命名实体识别任务带来挑战。

为了解决小样本学习下的 NER 任务，学术界提出了一些相应的方法，本文重点介绍基于样例匹配的方法和基于 prompt 的方法。

Ziyadi 等[39]提出了一种基于样例匹配的 NER 解决方法。其主要思路是利用一些具有标注样本的示例，从中识别出新数据中相关的实体。例如，在下面的示例中，右侧是需要进行实体识别的句子，左侧是一些具有标注的样本。通过利用左侧的标注信息，可以识别右侧句子中的实体，需要注意的是，左侧和右侧的实体名称是不同的：

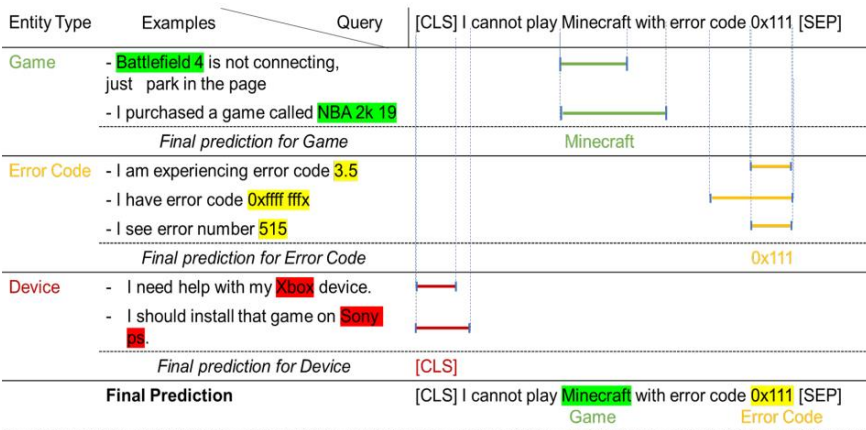


图 5-1 一种基于样例的 NER 方法的例子

Ziyadi 提出了一种两阶段的方法，用于解决命名实体识别任务。模型的第一部分，利用样例数据找到句子中最有可能是实体的 span 的起始位置和结束位置，然后将对应的实体标签分配给该 span。具体的方法是，在每个查询集上构建相应的支持集。在支持集中，为实体的前后添加特殊字符<e>和<\e>。通过对查询和支持集分别使用 Bert 进行编码，然后计算支持集中特殊字符<e>和<\e>以及查询中每个标记之间的匹配程度，从而确定查询中可能是实

体的 span 的起始和结束位置。对于每个查询，会有多个支持集句子可供参考。使用注意力机制来对每个支持集与查询之间的匹配进行加权，得到每个支持集与查询中起始和结束位置的相似度匹配分数。模型的第二部分是在第一步识别出的 entity span 的基础上，进一步识别这个 entity span 具体对应哪个 entity。这一步的具体做法和第一步类似，利用 support 样例计算 query 中每个 token 是各个 entity 类型的起始和终止位置的概率。

随着提示学习的兴起，Cui 等[40]采用提示学习[41]的思路解决小样本学习下的 NER 任务。传统的 NER 任务通常被视为序列标注任务，但这篇文章将其作为 cloze prompt 任务来处理。首先，需要手动定义一个正样本模板和一个负样本模板。对于给定的句子，如果某个词组是实体，那么对应的模板是" $x_{i:j}$  is a  $\langle y_k \rangle$ "; 如果某个词组不是实体，那么对应的模板是" $x_{i:j}$  is not an entity"。例如，对于输入文本"ACL will be held in Bangkok"，需要构建多组模板文本，对应每个词组是否为某个实体，比如"Bangkok is a location entity"。在训练阶段，根据标签构造所有实体模板对应的样本和非实体模板对应的样本。在训练过程中，原始文本被输入到经过预训练的 BART 编码器中，得到原文的编码表示。同时，在解码器中，预测根据模板生成的多组文本。例如，对于"ACL will be held in Bangkok"，在解码器阶段需要生成目标为"Bangkok is a location entity"的文本。这样，解码器学习到一种能力，根据原始文本输入，对使用模板构建的文本进行评分。如果"Bangkok is a location entity"的评分很高，说明该文本是正确的，那么就可以提取出"Bangkok"作为一个地名对应的实体。下图展示了模型的结构和模板示例。

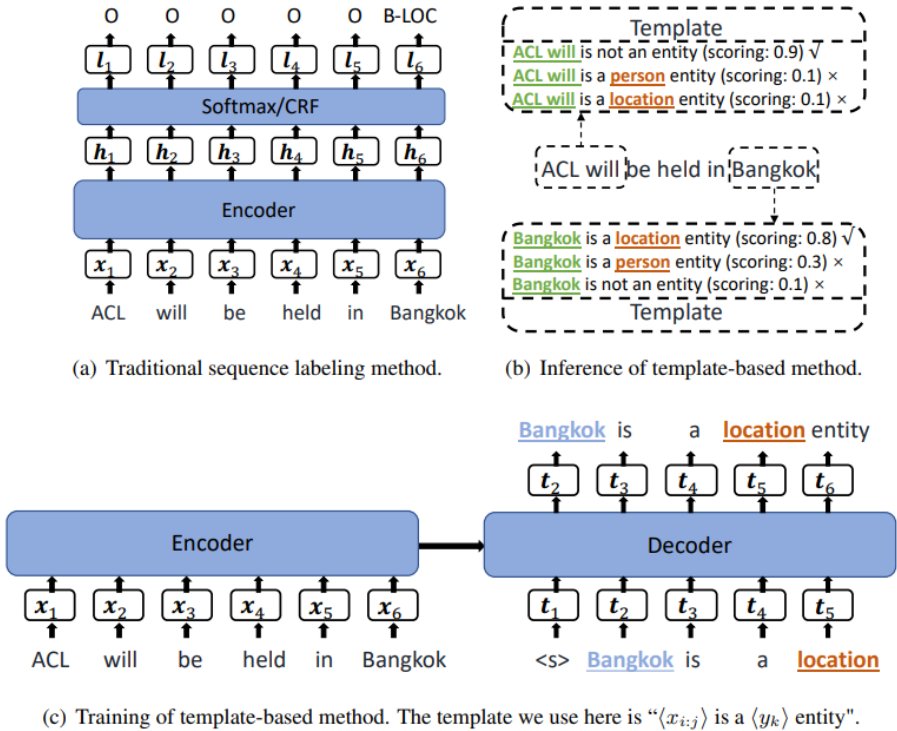


图 5-2 基于提示学习的 NER 任务

## 6 总结与展望

命名实体识别作为自然语言处理领域的重要任务，一直受到广泛关注和研究。在未来，我们可以期待以下几个方面的发展和改进：从模型性能上考虑，未来可以探索更复杂的模型架构、引入更多的上下文信息、利用多任务学习和迁移学习等技术来改善 NER 的准确性和

泛化能力；从可解释性考虑，可以从外部知识角度探索如何使 NER 模型更具解释性，以便用户和研究人员能够理解和解释模型的决策过程，并对结果进行调整和改进。

小样本命名实体识别是一个具有挑战性的任务，在未来的研究中，我们可以期待以下方面的发展和改进：数据增强方面，为了克服小样本数据的限制，进一步探索例如基于生成模型的数据合成方法、基于预训练语言模型的数据扩充等数据增强技术；领域适应方面：进一步开发更有效的迁移学习技术，将从大规模数据集中学到的知识和模型参数迁移到小样本任务中。

## 参考文献

- [1] LEI Jianbo, TANG Buzhou, LU Xueqin, et al. A comprehensive study of named entity recognition in Chinese clinical text[J]. Journal of the American Medical Informatics Association, 2014, 21(5): 808-814.
- [2] RAU L F. Extracting company names from text[C]//IEEE Conference on Artificial Intelligence Applications, 1991.
- [3] GRISHMAN R, SUNDHEIM B. Message understanding conference 6: a brief history[C]//Proceedings of the 16th Conference on Computational Linguistics-Volume 1, 1996.
- [4] O. Etzioni et al. ,“Unsupervised named-entity extraction from the web: An experimental study,”Artif.Intell.,vol.165,no.1,pp.91–134,2005.
- [5] S. Zhang and N. Elhadad, “Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts,” J. Biomed. Informat.,vol.46,no.6,pp.1088–1098,2013.
- [6] J.-H. Kim and P. C. Woodland, “A rule-based named entity recognition system for speech input, ”in Proc.6th Int. Conf. Spoken Lang.Process.,2000, pp.528–531.
- [7] D.Hanisch, K.Fundel, H.-T.Mevissen, R.Zimmer, and J. Fluck, “Prominer: Rule-based protein and gene entity recognition,” BMC Bioinf.,vol.6,no.1,2005,Art.no.S14.
- [8] A.P.Quimbaya et al., “Named entity recognition over electronic health records through a combined dictionary-based approach,” Procedia Comput.Sci.,vol.100,pp.55–61,2016.
- [9] K. Humphreys et al., “University of Sheffield: Description of the laSIE-II system as used for MUC-7,” in Proc.7th Message Understanding Conf.,1998, pp.1–20.
- [10] G. Krupka and K.IsoQuest, “Description of the nerOWL extractor system as used for MUC-7,” in Proc.7th Message Understanding Conf.,2005, pp.21–28.
- [11] W.J. Black, F. Rinaldi, and D.Mowatt, “FACILE: Description of the NE system used for MUC-7,” in Proc.7th Message Understanding Conf.,1998, pp.1–10.
- [12] C.Aone,L.Halverson,T.Hampton,and M.Ramos-Santacruz,“SRA:Description of the IE2 system used for MUC-7,”in Proc.7th Message Understanding Conf.,1998,pp.1–14.
- [13] D.E.Appelt et al., “SRI international FASTUS system: MUC-6 test results and analysis,” in Proc.6th Message Understanding Conf.,1995, pp.237–248.
- [14] A.Mikheev, M.Moens, and C. Grover, “Named entity recognition without gazetteers,” in Proc.9th Conf.Eur.Chapter Assoc.Comput.Linguistics,1999, pp. 1–8.
- [15] D. Nadeau and S.Sekine, “A survey of named entity recognition and classification,” LingvisticæInvestigationes, vol. 30, no.1, pp.3–26,2007.
- [16] A.Mikheev, “A knowledge-free method for capitalized word disambiguation, ”in Proc.37th Annu.Meeting Assoc.Comput.Linguistics,1999, pp.159–166.
- [17] Y.Ravin and N.Wacholder, “Extracting Names From Natural-Language Text,” IBM Research,



pp. 1–40, 1997.

- [18] S.R.Eddy, “Hidden markov models,” *Curr. Opinion Structural Biol.*, vol. 6, no. 3, pp. 361–365, 1996.
- [19] J.R. Quinlan, “Induction of decision trees,” *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
- [20] J.N. Kapur, *Maximum-Entropy Models in Science and Engineering*. Hoboken, NJ, USA: Wiley, 1989.
- [21] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, “Support vector machines,” *IEEE Intell. Syst. Their Appl.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 1998.
- [22] J.D. Lafferty, A. McCallum, and F.C.N. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” in *Proc. 18th Int. Conf. Mach. Learn.*, 2001, pp. 282–289.
- [23] D.M. Bikel, S. Miller, R. Schwartz, and R. Weischedel, “Nymble: A high-performance learning name-finder,” in *Proc. 5th Conf. Appl. Natural Lang. Process.*, 1997, pp. 194–201.
- [24] G. Szarvas, R. Farkas, and A. Kocsor, “A multilingual named entity recognition system using boosting and C4.5 decision tree learning algorithms,” in *Proc. Int. Conf. Discov. Sci.*, 2006, pp. 267–278.
- [25] P. McNamee and J. Mayfield, “Entity extraction without language-specific resources,” in *Proc. 6th Conf. Natural Lang. Learn.*, 2002, pp. 1–4.
- [26] V. Krishnan and C.D. Manning, “An effective two-stage model for exploiting non-local dependencies in named entity recognition,” in *Proc. 21st Annu. Meeting Assoc. Comput. Linguistics*, 2006, pp. 1121–1128.
- [27] B. Settles, “Biomedical named entity recognition using conditional random fields and rich feature sets,” in *Proc. Annu. Meeting Assoc. Comput. Linguistics*, 2004, pp. 104–107.
- [28] X. Liu, S. Zhang, F. Wei, and M. Zhou, “Recognizing named entities in tweets,” in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics*, 2011, pp. 359–367.
- [29] T. Rocktäschel, M. Weidlich, and U. Leser, “ChemSpot: A hybrid system for chemical named entity recognition,” *Bioinformatics*, vol. 28, no. 12, pp. 1633–1640, 2012.
- [30] Huang, Zhiheng et al. “Bidirectional LSTM-CRF Models for Sequence Tagging.” *ArXiv* abs/1508.01991 (2015): n. pag.
- [31] Souza, Fábio, Rodrigo Nogueira, and Roberto Lotufo. “Portuguese named entity recognition using BERT-CRF.” *arXiv preprint arXiv:1909.10649* (2019).
- [32] Cui, Leyang, and Yue Zhang. “Hierarchically-refined label attention network for sequence labeling.” *arXiv preprint arXiv:1908.08676* (2019).
- [33] Boliang Zhang, Xiaoman Pan, Tianlu Wang, Ashish Vaswani, Heng Ji, Kevin Knight, and Daniel Marcu. Name tagging for low-resource incident languages based on expectation-driven learning. In *NAACL HLT*, pp. 249–259, 2016.
- [34] Hai Wang and Hoifung Poon. 2018. Deep probabilistic logic: A unifying framework for indirect supervision. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 1891–1902.
- [35] Jiaao Chen, Zhenghui Wang, Ran Tian, Zichao Yang, and Diyi Yang. 2020. Local additivity based data augmentation for semi-supervised NER. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. 1241–1251.
- [36] Yaosheng Yang, Wenliang Chen, Zhenghua Li, Zhengqiu He, and Min Zhang. 2018. Distantly supervised NER with partial annotation learning and reinforcement learning. In *Proceedings of the 27th International Conference on Computational Linguistics*. 2159–2169.

- [37] Jingbo Shang,Liyuan Liu,Xiang Ren,Xiaotao Gu,Teng Ren,and Jiawei Han.2018.Learning named entity tagger using domain-specific dictionary.In Proceed ings of the 2018 Conference on Empirical Methods in Natural Language Processing.2054–2064.
- [38] Yi Yang and Arzoo Katiyar. 2020. Simple and effective few-shot named entity recognition with structured nearest neighbor learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6365–6375.
- [39] Ziyadi, Morteza, et al. "Example-based named entity recognition." *arXiv preprint arXiv:2008.10570* (2020).
- [40] Cui, Leyang, et al. "Template-based named entity recognition using BART." *arXiv preprint arXiv:2106.01760* (2021).
- [41] Liu, Pengfei et al. "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing." *ACM Computing Surveys* 55 (2021): 1 - 35.