应用深度学习的中文命名实体识别研究综述*

潘俊[1] 李萌配[1] 王贤明[2]
(1. 浙江科技学院理学院，杭州 310023；2. 温州理工学院数据科学与人工智能学院，温州 325035)

摘要：命名实体识别是自然语言处理领域的基础性工作，旨在从非结构化文本中识别出具有特定意义的实体并分类，在多种自然语言处理任务中发挥重要作用。由于中文命名实体没有明显的边界标记，且存在歧义和嵌套等问题，其识别过程比英语等其他语言要更为复杂。近年来，深度学习技术发展迅速，在中文命名实体识别中得到广泛应用，并已成为主流方法。系统梳理中文命名实体识别中深度学习技术的研究进展，重点从文本表示、特征编码、预测解码3个方面，对比分析代表性工作的关联性和关键技术，讨论研究中存在的问题、现有解决方案和未来的研究方向。

关键词：中文命名实体识别；深度学习；自然语言处理；编码解码框架
中图分类号：TP311；G35　DOI：10.3772/j.issn.1673-2286.2023.05.001
引文格式：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(5)：1-9.

命名实体是指文本中具有特定意义的实体，如人名、地名、机构名等。作为自然语言处理（Natural Language Processing，NLP）中的基础性工作，命名实体识别（Named Entity Recognition，NER）一直广受研究者的关注。

NER经历了从基于规则到基于机器学习，再到基于深度学习的发展过程。基于规则的方法依靠人工构建的规则模板来匹配命名实体，简单易行，但存在构建成本高、覆盖范围小、可移植性差等缺点。基于机器学习的方法通过数学模型从有标注数据中学习文本特征和模型参数，泛化性和扩展性较好，但效果仍依赖于有效特征的选取和数据质量等因素。深度学习则强调从数据中自动提取不同抽象层次的特征并进行非线性变换，更符合人类的分层认知过程，近年来NLP领域取得了较大进展，循环神经网络（RNN）、卷积神经网络（CNN）、图神经网络（GNN）等结构被应用于NER，有效推动了NER技术的进步。

不同语种的语言常有各自独特的语言结构和特征，除英语外，一些研究对特定语种的NER进行了研究[1]。中文NER更具挑战性，困难主要表现在：①边界模糊，没有空格分隔符；②构成复杂，存在大量缩（简）写、实体嵌套、多义词等；③形态特殊，缺少前后缀、词元、词干、专名大写等单词形态学特征。针对这些问题，一些应用新兴深度学习技术的方法诞生，极大推动了中文NER的进步。为此，本文将重点从文本表示、特征编码、预测解码3个方面对中文NER深度学习研究进行全面梳理和分析，并总结该领域的最新进展和未来方向，以期为进一步研究提供有益参考。

1 应用深度学习的中文NER处理过程

应用深度学习的中文NER处理过程（见图1）重点关注两个问题：①如何有效编码句子；②如何准确生成

收稿日期：2023-03-16
*本研究得到浙江省公益技术应用研究计划项目“多源异构数据融合的农业知识服务关键技术及应用”（LGN21F020003）、浙江省高校重大人文社科攻关计划项目“江南士人群体社会关系网络与地域文化演进研究”（2023QN088）资助。

句子标签序列。具体而言，包括以下3个环节。

（1）文本表示。将输入文本以字或词为基本单位转化为向量表示，可结合形态学、语言学等特征以获得更丰富的语义表示。

（2）特征编码。利用神经网络模型对文本表示逐层变换与组合，提取上下文特征并编码，建立上下文依赖关系模型，并可通过注意力机制加强字词语义信息。

（3）预测解码。根据上下文编码，对命名实体的标签进行解码和预测，输出标签序列。

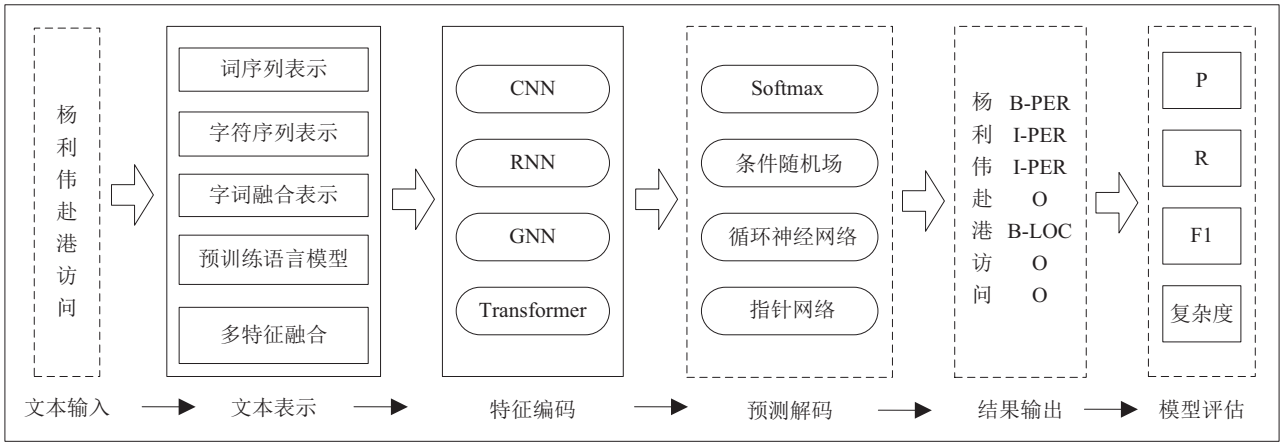

图1 应用深度学习的中文NER处理过程

1.1 文本表示

基于深度学习的文本表示把字、词、短语等粒度的语言单元，以及形态学、语言学、语义学等特征纳入统一的语义空间学习，将语言单元映射为低维稠密的实数向量，是当前主流的表示方式。文本表示的方法主要有3种。①静态表示，在大规模语料上学习语言模型，得到与上下文无关的词元向量，可分为词序列表示、字符序列表示、字词融合表示等。词序列表示具有较强的语义表达能力，但存在分词错误传播等问题；字符序列表示不需要考虑分词和词库规模，但缺少词汇语义和边界信息；字词融合表示在字表示中引入了词汇信息，但增加了词典构建和训练成本。②动态表示，通过预训练语言模型获得与上下文相关的词元向量，与静态表示相比，可以解决一词多义问题。动态表示分为双向语言模型和遮蔽语言模型：前者从两个方向抽取文本特征进行拼接；后者通过遮蔽文本的部分内容来学习共现信息，特征提取能力更强。③多特征融合表示，融合词性、部首、笔画、字形、输入法、依存句法等额外特征，增强了文本表示。不同类型的文本表示各具特点，选择什么样的上下文及其特征组合作为输入，通常要根据具体的目标而定。

1.1.1 基于词序列的文本表示

基于词序列的文本表示先对文本分词，再将离散符号表示的词汇映射成低维的实数向量，作为特征编码层的输入。在中文NER中，基于词序列的文本表示有一定优势，因为词边界通常也是实体边界，且词汇比字符含有更多语义信息。黄晓辉等[2]构建了面向中文分词与NER联合学习的序列标注模型，实体识别效果有较大提升。但词序列表示对词典的依赖性较高，对未登录词（Out of Vocabulary，OOV）问题较敏感。OOV是指出现在数据集中但不在词典中的未知词，有可能是一些重要实体。如果标注数据不充足或词典不全面，OOV的表示就难以被学习，导致识别错误。此外，在分词与NER多任务联合学习时要共享参数，分词错误或引入错误的先验知识也会降低准确率。

1.1.2 基于字符序列的文本表示

字符序列表示能够解决OOV问题，是中文NER的常用表示方法[3-5]。Liu等[3]发现，在没有全局特征的情况下，基于字符的方法对人名和地名等实体的识别更准确。Dong等[4]采用基于字符表示的双向LSTM-CRF模型，并融入汉字部首特征。苏丰龙等[5]采用基于字符表

示的残差门控卷积模型，并融入字符所属词的位置向量。然而，字符级别的表示不能表达一些有特殊含义的词语，也没有利用词边界信息，可能会丢失重要信息。

1.1.3 基于字词融合序列的文本表示

近年来中文NER研究尝试在字符表示中引入词典信息。相关工作主要关注网络结构和模型速度两个方面。Zhang等[6]在LSTM网络中引入动态的词神经网络单元，将以当前字符结尾的词汇分别输入Lattice LSTM模型以融合词汇信息。但受限于构造方式，Lattice LSTM模型只能给词尾字符添加词汇，后续工作对此作了改进。Gui等[7]通过CNN对N-gram词汇的卷积操作，捕获句中潜在的词汇信息，并利用反思机制将高层信息重新输入底层，缓解了字符与多个词汇匹配时的冲突问题。Sui等[8]通过3种构图方式多角度引入词汇信息，并融合3种构图的向量表示。Gui等[9]利用基于词典信息的图网络捕捉局部信息，通过全局中继节点融合全局信息，缓解了词边界模糊和歧义问题。Xue等[10]提出多孔词格模型，在捕获长程依赖关系的同时，提升局部建模能力。Li等[11]设计了一种位置编码来融合Lattice结构，可直接为字符与匹配词之间的交互建模。

上述工作均通过编码层融合词汇信息，不易迁移。为此，Liu等[12]基于Lattice LSTM模型，用固定长度的编码来表示词尾字符的词汇信息。陈淳等[13]分别输入字和词的嵌入表示，通过融合双向注意力机制的高速网络抽取有效字的组合信息。除词典外，还可以融合实体词典（gazetteers）来提升NER表现[14]。这些工作均在表示层融合词汇信息，再接入具体的网络，具备可迁移性。

1.1.4 基于预训练语言模型的文本表示

静态向量难以解决一词多义问题，一种方案是由预训练语言模型动态生成向量。预训练语言模型可分为两类。①双向语言模型，以Elmo模型为代表，罗凌等[15]用Elmo模型学习与上下文相关且包含汉字结构的字向量，以识别电子病历命名实体。②遮蔽语言模型，以BERT模型为代表，即随机遮蔽输入序列的部分元素，通过恢复原始序列来学习语言。双向语言模型不能同时利用上文和下文的信息，而BERT模型可以通过双向编码来习得字符的向量表示，如吴俊等[16]针对专业术语实体表述多样、结构复杂的特点，用BERT模型将文本转化为字符序列表示，输入BiLSTM-CRF模型获得术语实体标签序列。

BERT模型的中文模型以字为粒度切分，仅能获得字符序列表示。RoBERTa模型是BERT模型的改进版本，采用全词掩码技术，可获得词序列向量表示，所以一些工作用RoBERTa模型来获得更多的词边界信息[17]。为了将词汇信息整合到BERT模型的底层，Liu等[18]提出了LEBERT模型，将词汇的类型标签信息深度融合到Transformer层，以更充分地提取和利用词汇信息。

1.1.5 基于多特征融合的文本表示

NER是一个重底层的任务，提升性能的另一种思路是融合额外的形态学、语言学等特征。

（1）拼音特征。汉字存在多音不同义现象，例如在“盖叫天赴铅山演出”这句话中，“盖”读“gě”，表示姓氏，“铅”读“yán”，表示地名，通过引入拼音特征可丰富词元的语义表示。Feng等[19]在Flat Lattice Transformer[11]的基础上加入拼音特征，并用CNN提取拼音局部特征，再用两个Transformer模块融合Flat Lattice嵌入和拼音嵌入以解决同音字问题，有效提升了NER性能。

（2）部首笔画特征。汉字偏旁部首、笔画字形等特征具有表意功能。Wu等[20]将汉字分解为部首结构，再通过CNN得到字符的部首特征嵌入。此外，笔画能够刻画更细粒度的汉字结构，罗凌等[15]、Dong等[4]在输入层融入笔画特征，取得了良好效果。

（3）字形特征。汉字字形具有语义信息[21]。字形和语境是有关联的，相邻字符的字形之间存在交互信息。Xuan等[22]将汉字视为字形编码，利用字符图谱序列卷积网络来编码字形，捕捉相邻字形图之间的信息。Meng等[21]提出了一种汉字字形向量表示方法，利用古代文字和现代文字的组合来丰富字符图像的象形信息，并设计了田字格卷积网络来获得字形向量。

（4）句法特征。引入句法成分、依存关系、词性（POS）标签等句法特征，有利于提升NER性能。不同种类的句法特征对句子的贡献度不同，需通过加权组合以有效利用句法信息。Nie等[23]将句子的POS标签、句法成分、依存关系等类型的特征输入键值记忆网络，得到包含句法信息及上下文的编码，再输入句法关注网络进行加权处理。Xu等[24]先将句法结构信息输入图卷积网络进行编码，再利用门控机制来整合依存树捕获的结构信息和线性序列捕获的上下文信息。

1.2 特征编码

特征编码阶段旨在捕获输入序列的上下文依赖关系，一般通过多层神经网络来实现，常见的网络结构有4种。①RNN，利用前一时刻的状态计算当前时刻的状态，结构简单且适合序列建模，尤其是双向RNN能有效捕获特定范围内的信息[25]，Lattice LSTM[6]、WC-LSTM[12]等模型均采用了双向LSTM的结构。但RNN本质上是一个马尔可夫决策过程，无法捕获句子间不同位置的信息，不支持并行计算。②CNN，具有并行处理的优势，但只能获取局部信息。为捕获更大范围的信息，研究者提出空洞卷积[26-27]、门控卷积[5,28]来扩大感受域。③GNN，适合为字、词和句子之间的复杂依赖关系建模，如CGN[8]、LGN[9]、多图结构[14]等模型，其缺点是结构较复杂，训练难度大。④Transformer，能直接获取全局信息，支持并行计算，如PLTE[10]、FLAT[11]、TENER[29]模型均使用了Transformer结构及变体，但自注意力机制存在训练成本高、对小数据集的拟合效果有限等问题。

1.2.1 基于RNN

RNN及其变体是一种记忆性的神经网络，在NER任务中，常用的是双向长短期记忆网络。Huang等[25]首次采用BiLSTM-CRF模型来解决NLP的基准序列标注问题。为了引入中文词典信息，Lattice LSTM模型对LSTM的网络结构作了修改，在Bi-LSTM网络中引入词汇格子神经单元，通过门控机制控制信息流，可有效融合字符匹配词汇的信息。但每个字符只能融合以其为结尾的词汇信息，造成信息损失，且字符对应的格子单元数量各异，不能实现batch并行化训练。为此，许多工作进一步优化了LSTM网络。例如，Liu等[12]改进了Lattice LSTM网络，将词汇信息整合到每个字符中，同时将词汇编码为定长向量，使其适合批处理训练。

1.2.2 基于CNN

CNN能逐层提取并组合局部特征。相比RNN，CNN以前馈方式处理输入序列，具有可并行处理的优势，但在提取长程依赖信息方面较为受限。

为了扩大感受域，Strubell等[26]提出了IDCNN模型，通过堆叠共享参数的空洞卷积层来提取更大范围的特征。此外，在卷积层中使用不同步长的滤波器，可提取不同粒度（如部首、偏旁、字符等）的局部连续特征[27]。另一种捕获长程依赖信息的方法是门控机制，即先用卷积操作学习词元的局部特征，再通过门控结构为词元建立关联，从而将局部上下文特征融合到全局上下文中[28]。为了进一步融合词典信息，Gui等[7]提出了LR-CNN模型，使用CNN对字符特征进行编码并提取bigram特征，并通过多层堆叠获取多种k-gram信息。

1.2.3 基于GNN

在中文NER中，GNN主要用于对字词建模。Ding等[14]将Lattice LSTM结构从链式转为图式，把字符、匹配的实体词的类型作为节点并连边，用GRU（Gated Recurrent Unit）模型更新权重。Sui等[8]提出基于协作图的神经网络来对字符和词汇的交互建模，设计了3种构图方式以获取全方位词汇信息，通过融合层合并特征。Gui等[9]也将字符和匹配词汇作为节点并连边，通过图结构融合局部信息，并增加全局节点以获取全局语义信息。以上模型主要基于RNN框架，Tang等[30]通过构建字词有向无环图并输入双向图卷积层来提取特征。Nie等[31]构建了基于多粒度特征的异构网络，使用图卷积对输入文本的词性、字、词语、依存句法、义原等特征进行建模，取得了较好的结果。

1.2.4 基于Transformer编码器

Transformer是一种基于自注意力机制和全连接层的编码-解码模型，具有强大的语义特征提取、长程特征捕获及并行计算能力。Yan等[29]提出了TENER模型，此模型采用自适应Transformer编码器，结合方向感知、距离感知和注意力分布，能同时捕获相对位置与绝对位置信息，可有效地对词与字符的交互建模。Xue等[10]基于Transformer的变体Star Transformer，加入位置关系表示，通过多孔机制改变格子感知注意力的分布，提高了局部建模能力，但该模型将格子间信息转化为多种相对关系，可能会丢失格子的位置信息。为此，Li等[11]提出一种扁平格子的结构，可依据编码词元之间“头”与“尾”的位置来捕获词信息。

1.3 预测解码

预测解码阶段旨在对特征提取层的编码进行解码，输出对应的标签序列。早期方法在使用多层感知机解码时，对输入序列的每个词元都独立选择概率最大的标签。这种方法简单高效，但并非最优，因为它假设词元的标签相互独立，而相邻标签之间往往存在一定联系。因此，NER的输出更适合用结构化的预测序列来表示，主流方法有：①CRF，可采用动态规划进行全局优化，但复杂度高，不适用于弱相关性标签数据；②RNN，可对所有历史标签建模，适用于标签数量较多的数据；③指针网络，解决了可变长度的输出词典问题，具有较强的嵌套识别能力，但其采用的贪心或束搜索优化策略存在复杂度较高等问题。

1.3.1 基于CRF的解码

NER任务的输出标签之间具有较强的依赖关系，例如I-PER与E-ORG两个标签不可能相邻出现。若仅对每个字符的输出标签进行独立预测，会降低识别精度。CRF模型能学习标签之间的约束关系，是常用的全局标签解码器之一，可独立应用于各类特征提取器[25]。然而，在实践中CRF使用Viterbi算法来输出得分最高的标签序列，训练时长会随不同实体类型数量的增加而急剧增长。一种解决方法是重排序（reranking）技术，即先用基础模型获得前k个候选序列，再用复杂模型对这k个候选序列重新评分，最后输出最高分值序列[32]。

1.3.2 基于RNN的解码

受限于马尔可夫假设，CRF仅依赖前一个标签进行预测。若放宽马尔可夫独立性假设，可以考虑在预测解码层引入RNN来建模历史标签。Cui等[33]提出了一个分层结构的标签注意力网络，通过分层注意力机制，逐步完善每个给定词汇的标签分布，以显式地利用标签信息并捕获潜在的长程标签依赖信息。Shen等[34]使用LSTM作为解码器，采用从左向右的贪心预测方法进行训练，标注精度和训练速度均优于CRF。Xue等[10]利用BiGRU-CRF对多孔格子Transformer编码器提取的特征进行解码，弥补了编码层在捕捉序列信息方面的不足。黄晓辉等[2]同样利用改进的RNN构建标签解码层，实现了对标签序列长程依赖的有效建模。

1.3.3 基于指针网络的解码

指针网络是一种能够生成可变大小输出序列的神经网络架构，旨在解决传统的序列到序列模型必须固定序列长度的局限性。基于指针网络的NER解码将待标注的序列作为输入，通过预测实体的起始索引与结尾索引来确定实体位置，然后预测实体类别。Zhai等[35]采用结合指针网络的模型对序列分块并进行标注。Li等[36]先使用单层GRU解码，然后通过指针网络来解决实体边界不确定的问题。指针网络在提取嵌套实体方面有比较大的优势，但是指针网络通常将多实体抽取转化为N个二分类问题，序列过长会导致收敛速度过慢。

2 应用深度学习的中文NER研究展望

应用深度学习的中文NER在算法精度、训练成本和运行效率等方面都取得了很好的效果，但该领域仍有许多问题需要研究，主要体现在以下方面。

2.1 小样本命名实体识别

深度学习NER模型需要大量标注数据，在NLP中语料虽易获取，但标注成本高，因此，中文NER的小样本学习十分重要。目前的解决思路主要有数据增强、迁移学习和元学习等。

（1）数据增强。目的是在不增加人工标注成本的情况下，通过添加噪声来扩大训练集规模。主要方法有3种。①EDA（Easy Data Augmentation）操作，通过同义词替换、随机噪声注入、混合交叉等操作添加噪声。宋希良等[37]采用新类型人名实体替换策略生成伪训练数据，提升了识别性能。②远程监督，利用外部词典或知识库，通过对无标注数据进行脚本标注获得训练数据。刘哲宁等[38]构建了存储实体类型与实体名的知识库，自动对特定领域语料进行伪标注。③生成模型，通过语言模型生成标注语料。Zhou等[39]同样引入标记序列线性化策略，利用句子上下文和实体标签进行基于实体的数据增强。3类方法各有特点：EDA要注意噪声的合理性，避免破坏命名实体的合法性；远程监督降低了标注成本，但依赖知识库，并可能引入大量噪声；生成模型无需外部资源，在具备大量无标注语料时较为适用。

（2）迁移学习。目的是利用源域的数据和知识来提升目标域命名实体识别的性能，可分为数据迁移[40]和

模型迁移[41]两类。数据迁移的思想是利用资源丰富语言的标注数据，通过信息对齐将源语言映射到低资源语言中，然后基于这些数据来训练模型。模型迁移旨在将源模型的部分参数或特征迁移至目标域模型，包括：①跨语言迁移，在源语言上训练模型，再通过目标语言对模型调优，或通过共享参数进行知识迁移；②跨领域迁移，利用高资源领域的标注数据训练模型并进行知识迁移；③跨任务迁移，利用词性标注、分词、关系抽取等相关任务的标注数据和知识来提升NER效果。

（3）元学习。目标是让模型获得元知识，以指导小样本新任务学习。Li等[42]提出了MetaNER模型，利用来自多个领域的数据获得稳健和通用的初始化表征，结合未知领域的少量标注数据实现领域适配。Wu等[43]利用少量目标语言测试数据，对训练后的源语言模型调优，并用改进的元学习算法获得更优的模型初始化参数，取得了较好结果。

2.2 嵌套命名实体识别

嵌套命名实体是指一个实体内部存在另一个实体或与之重叠的情景。其词元可能有多种实体标签，例如“[[杭州市][西湖区]行知小学]”是2层嵌套实体，“杭”与“西”都有B-LOC或B-ORG两种标签。Li等[44]提出了一个同时处理嵌套实体与扁平实体的框架，将NER任务视为机器阅读理解任务，从而将嵌套实体的不同标签转化为多个阅读理解问题的答案。

2.3 非正式文本命名实体识别

微博、微信等非正式文本已成为信息传递的主要载体，这些数据从词汇到文法都常有不规范之处。现有方法主要利用实体词典、维基百科等外部资源来提升性能，或使用在大型社交媒体上训练的文本表示作为输入，通过语义增强等方法来提高NER的性能。Nie等[45]用预训练词向量增强数据语义，并用语义增强模块进行编码，然后用GRU模块来融合语义特征。Chen等[46]先通过自注意力机制提取字符向量，并与词向量组合，解决了社交媒体文本的实体边界模糊问题。

2.4 深度学习其他技术的应用

随着深度学习研究的深入，研究人员陆续将生成对抗网络（Generative Adversarial Networks，GAN）、知识蒸馏、多任务联合学习等新成果应用于NER任务，出现一些新的研究趋势。

（1）生成对抗网络。GAN是一种包含生成模型与对抗模型的深度学习模型[36]，GAN在NER任务中主要有两个方面的应用：①对抗迁移，通过对抗训练来实现跨语言、跨领域或跨任务的知识迁移；②添加扰动，通过向训练样本添加扰动，提升模型的性能。

（2）知识蒸馏。知识蒸馏是一种教师-学生训练结构，在知识蒸馏过程中，长命名实体序列存在对应多种标签组合的情况，导致教师模型提炼知识的过程非常耗时。因此，Zhou等[47]用Viterbi算法计算教师模型中概率最大的标签序列，将其作为主要知识，再用交叉熵函数计算其他标签序列的概率。此外，BERT模型也可以作为教师模型来指导轻量级的NER学生模型。

（3）多任务联合学习。NER与分词、实体链接、事件检测等任务之间存在相关性，通过多任务联合学习能有效共享隐性知识，缓解误差传播问题。Martins等[48]将实体链接与实体识别进行联合训练，发现实体链接中的指称项表示提升了NER的效果。黄晓辉等[2]通过卷积RNN构建特征编码层，并设计了统一的中文分词和实体识别标签结构，以实现分词与实体识别的联合学习。余传明等[49]提出一种多任务深度学习的实体与事件联合抽取模型，利用实体识别与事件检测联合模型的共享信息，提高了两个任务的表现能力。

3 总结与讨论

本文对深度学习在中文NER中的研究进展作了详细梳理，重点从文本表示、特征编码和预测解码3个方面，对代表性工作的关键技术和联系进行了分析与对比，以期为相关领域的研究人员提供较为全面的参考。

传统机器学习在中文NER的数据标注、特征选择等方面存在较多困难，而深度学习为此提供了3个方面的增益。首先，深度学习将文本视作不同粒度的语言单元序列，可统一将文本表示为低维稠密的向量并作为特征编码层的输入。其次，使用CNN、GNN、RNN或其他神经网络，对输入向量进行非线性组合转换，可逐层提取抽象特征并编码。再次，通过线性结构或者非线性神经网络单元，可实现结构化的预测解码。

从文本表示来看，深度学习的目的是对离散字符序列进行转换，获得尽量丰富的语义信息，以提高后续

NER过程的性能。然而，由于中文表述的多样性和模糊性，低频词、多义词和OOV的表示仍需进一步研究。如何有效融合更多精细的语言学特征来建立多粒度文本的联合表示学习模型，如何引入已有的语义网、知识库等信息来建立文本表示与外部知识的联系，是需要进一步关注的问题。

从特征编码来看，深度神经网络在提取全局特征和上下文信息方面具有强大的能力，但其计算复杂度较高。虽然局部连接、权重共享、汇聚操作等设计有助于缓解这一问题，但长距离依赖信息的特征提取始终是一个关键的挑战。为此，研究人员相继提出LSTM、GRU、CNN、GNN以及Transformer编码网络等特征提取器及变体，试图为输入序列建立长距离依赖关系。在此过程中，注意力机制成为平衡模型复杂性和表达能力的重要方法，如何设计注意力机制并与新模型融合以提高NER模型的可解释性，是值得研究的课题。

从预测解码来看，CRF、RNN和指针网络等解码方法对输出依赖的学习能力仍然有限，如何更有效地利用序列标注的结构化输出标签信息是值得探索的问题。此外，除了命名实体，信息抽取还包括关系、事件、情感、观点等目标，因此建立统一的信息抽取模型，为多样化任务提供快速响应，是当前研究的重要方向。已有一些工作尝试抛弃序列标注的研究范式，转而采用生成式模型，试图对包括中文NER在内的多种信息抽取任务进行统一。然而，如何对不同的抽取目标进行统一编码以自适应地生成标签结构，如何对解码语义空间施加约束从而使预测输出精确可控是很大的挑战。

参考文献

[1] AJEES P A，MANJU K，IDICULA S M. An improved word representation for deep learning based NER in Indian languages [J]. Information，2019，10（6）：186.

[2] 黄晓辉，乔立升，余文涛，等. 中文分词与命名实体识别的联合学习 [J]. 国防科技大学学报，2021，43（1）：86-94.

[3] LIU Z X，ZHU C H，ZHAO T J. Chinese named entity recognition with a sequence labeling approach：based on characters，or based on words? [M] //HUANG D S，ZHANG X，GARCÍA C A R，et al. Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence. Lecture Notes in Computer Science. Berlin：Springer，2010，6216：634-640.

[4] DONG C H，ZHANG J J，ZONG C Q，et al. Character-based LSTM-CRF with radical-level features for Chinese named entity recognition [M] //LIN C Y，XUE N W，ZHAO D Y，et al. Natural Language Understanding and Intelligent Applications. Lecture Notes in Computer Science. Cham：Springer，2016，10102：239-250.

[5] 苏丰龙，孙承哲，景宁. 融合上下文的残差门卷积实体抽取 [J]. 北京大学学报（自然科学版），2022，58（1）：69-76.

[6] ZHANG Y E，YANG J E. Chinese NER using lattice LSTM [C] //Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018，1：1554-1564.

[7] GUI T，MA R T，ZHANG Q，et al. CNN-based Chinese NER with lexicon rethinking [C] //Proceedings of the 28th International Joint Conference on Artificial Intelligence. 2019：4982-4988.

[8] SUI D B，CHEN Y B，LIU K，et al. Leverage lexical knowledge for Chinese named entity recognition via collaborative graph network [C] //Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing（EMNLP-IJCNLP）. 2019：3830-3840.

[9] GUI T，ZOU Y C，ZHANG Q，et al. A lexicon-based graph neural network for Chinese NER [C] //Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing（EMNLP-IJCNLP）. 2019：1040-1050.

[10] XUE M G，YU B W，LIU T W，et al. Porous lattice transformer encoder for Chinese NER [C] //Proceedings of the 28th International Conference on Computational Linguistics. 2020：3831-3841.

[11] LI X N，YAN H，QIU X P，et al. FLAT：Chinese NER using flat-lattice transformer [C] //Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020：6836-6842.

[12] LIU W，XU T G，XU Q H，et al. An encoding strategy based word-character LSTM for Chinese NER [C] //Proceedings of the 17th Conference of the North American Chapter of the Association for Computational Linguistics. 2019：2379-2389.

[13] 陈淳，李明扬，孔芳. 基于两段高速网络的命名实体识别 [J]. 中文信息学报，2022，36（3）：64-72.

[14] DING R X，XIE P J，ZHANG X Y，et al. A neural multi-digraph model for Chinese NER with gazetteers [C] //

Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019：1462-1467.

[15] 罗凌，杨志豪，宋雅文，等. 基于笔画ELMo和多任务学习的中文电子病历命名实体识别研究 [J]. 计算机学报，2020，43（10）：1943-1957.

[16] 吴俊，程垚，郝瀚，等. 基于BERT嵌入BiLSTM-CRF模型的中文专业术语抽取研究 [J]. 情报学报，2020，39（4）：409-418.

[17] 张云秋，汪洋，李博诚. 基于RoBERTa-wwm动态融合模型的中文电子病历命名实体识别 [J]. 数据分析与知识发现，2022，6（Z1）：242-250.

[18] LIU W，FU X Y，ZHANG Y E，et al. Lexicon enhanced Chinese sequence labeling using BERT adapter [C] //Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 2021，1：5847-5858.

[19] FENG P，LI G L，WANG Y Y，et al. Chinese named entity recognition based on embedded pinyin information [C] //2022 2nd International Conference on Consumer Electronics and Computer Engineering（ICCECE）. 2022：719-722.

[20] WU S A，SONG X N，FENG Z H. MECT：multi-metadata embedding based cross-transformer for Chinese named entity recognition [C] //Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 2021，1：1529-1539.

[21] MENG Y X，WU W，WANG F，et al. Glyce：glyph-vectors for Chinese character representations [EB/OL]. [2022-08-20]. https://arxiv.org/abs/1901.10125.

[22] XUAN Z Y，BAO R，JIANG S Y. FGN：fusion glyph network for Chinese named entity recognition [M] //CHEN H J，LIU K，SUN Y Z，et al. Knowledge Graph and Semantic Computing：Knowledge Graph and Cognitive Intelligence. Communications in Computer and Information Science. 2021，1356：28-40.

[23] NIE Y Y，TIAN Y H，SONG Y，et al. Improving named entity recognition with attentive ensemble of syntactic information [C] //Findings of the Association for Computational Linguistics：EMNLP 2020. 2020：4231-4245.

[24] XU L，JIE Z M，LU W，et al. Better feature integration for named entity recognition [C] //Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics：Human Language Technologies. 2021：3457-3469.

[25] HUANG Z H，XU W，YU K. Bidirectional LSTM-CRF models for sequence tagging [EB/OL]. [2022-08-20]. https://arxiv.org/abs/1508.01991.

[26] 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.

[27] QIU J H，ZHOU Y M，WANG Q，et al. Chinese clinical named entity recognition using residual dilated convolutional neural network with conditional random field [J]. IEEE Transactions on Nanobioscience，2019，18（3）：306-315.

[28] CHEN H，LIN Z J，DING G G，et al. GRN：gated relation network to enhance convolutional neural network for named entity recognition [J]. Proceedings of the AAAI Conference on Artificial Intelligence，2019，33（1）：6236-6243.

[29] YAN H，DENG B C，LI X N，et al. TENER：adapting transformer encoder for named entity recognition [EB/OL]. [2022-06-20]. https：//arxiv.org/abs/1911.04474.

[30] TANG Z，WAN B Y，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.

[31] NIE Y，ZHANG Y L，PENG Y K，et al. Borrowing wisdom from world：modeling rich external knowledge for Chinese named entity recognition [J]. Neural Computing and Applications，2022，34（6）：4905-4922.

[32] GUO J，HAN Y X，KE Y Z. A neural-based re-ranking model for Chinese named entity recognition [J]. International Journal of Reasoning-Based Intelligent Systems，2019，11（3）：265-272.

[33] CUI L Y，ZHANG Y E. Hierarchically-refined label attention network for sequence labeling [C] //Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing（EMNLP-IJCNLP）. 2019：4115-4128.

[34] SHEN Y Y，YUN H，LIPTON Z，et al. Deep active learning for named entity recognition [C] //Proceedings of the 2nd Workshop on Representation Learning for NLP. 2017：252-256.

[35] ZHAI F F，POTDAR S，XIANG B，et al. Neural models for sequence chunking [C] //Proceedings of the 31st AAAI Conference on Artificial Intelligence. 2017：3365-3371.

[36] LI J，YE D H，SHANG S. Adversarial transfer for named entity boundary detection with pointer networks [C] //Proceedings of the 28th International Joint Conference on Artificial Intelligence. 2019：5053-5059.

[37] 宋希良，韩先培，孙乐. 面向新类型人名识别的数据增强方法 [J]. 中文信息学报，2019，33（6）：72-79.

[38] 刘哲宁，朱聪慧，郑德权，等. 面向特定标注数据稀缺领域的命名实体识别 [J]. 指挥信息系统与技术，2019，10（5）：14-18.

[39] ZHOU R，LI X，HE R D，et al. MELM：data augmentation with masked entity language modeling for low-resource NER [C] //Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022，1：2251-2262.

[40] NI J，DINU G，FLORIAN R. Weakly supervised cross-lingual named entity recognition via effective annotation and representation projection [EB/OL]. [2022-10-08]. https://arxiv.org/abs/1707.02483.

[41] LIU Z H，WINATA G I，FUNG P. Zero-resource cross-domain named entity recognition [C] //Proceedings of the 5th Workshop on Representation Learning for NLP. 2020：1-6.

[42] LI J，SHANG S，SHAO L. MetaNER：named entity recognition with meta-learning [C] //Proceedings of The Web Conference 2020. 2020：429-440.

[43] WU Q H，LIN Z J，WANG G X，et al. Enhanced meta-learning for cross-lingual named entity recognition with minimal resources [J]. Proceedings of the AAAI Conference on Artificial Intelligence，2020，34（5）：9274-9281.

[44] LI X Y，FENG J R，MENG Y X，et al. A unified MRC framework for named entity recognition [C] //Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020：5849-5859.

[45] NIE Y Y，TIAN Y H，WAN X A，et al. Named entity recognition for social media texts with semantic augmentation [C] //Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing（EMNLP）. 2020：1383-1391.

[46] CHEN C，LI M，KONG F. Lightweight named entity recognition for Weibo based on word and character [C] // Proceedings of the 19th Chinese National Conference on Computational Linguistics. 2020：402-413.

[47] ZHOU X A，ZHANG X A，TAO C Y，et al. Multi-grained knowledge distillation for named entity recognition [C] // Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics：Human Language Technologies. 2021：5704-5716.

[48] MARTINS P H，MARINHO Z，MARTINS A F T. Joint learning of named entity recognition and entity linking [C] // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics：Student Research Workshop. 2019：190-196.

[49] 余传明，林虹君，张贞港. 基于多任务深度学习的实体和事件联合抽取模型 [J]. 数据分析与知识发现，2022，6（Z1）：117-128.

作者简介

潘俊，男，博士，副教授，研究方向：信息管理、自然语言处理。
李萌配，男，硕士研究生，研究方向：自然语言处理。
王贤明，男，硕士，教授，通信作者，研究方向：文本挖掘、舆情分析，E-mail：xmwung@ustc.edu。

Review of Chinese Named Entity Recognition Based on Deep Learning

PAN Jun[1]　LI MengPei[1]　WANG XianMing[2]
(1. School of Science, Zhejiang University of Science and Technology, Hangzhou 310023, P. R. China;
2. School of Data Science and Artificial Intelligence, Wenzhou University of Technology, Wenzhou 325035, P. R. China)

Abstract: Named Entity Recognition(NER) is a fundamental task in Natural Language Processing(NLP) that aims to identify and clarify entities with specific meanings from unstructured text. It is an indispensable part of various downstream NLP fields. Chinese NER is more difficult than English and other languages because there are no obvious boundary markers, and the entities have problems such as ambiguity and nesting, which pose great challenges for existing methods. In recent years, deep learning technology has developed rapidly and has been widely applied in Chinese NER. We give a comprehensive study of recent advances in deep learning research of Chinese NER from perspectives of text representation, context encoding, and tag decoding, focusing on key techniques and relationship among these works. Furthermore, we summarize the main challenges and the latest advances of Chinese NER, and give a discussion on possible future work.

Keywords: Chinese Named Entity Recognition; Deep Learning; Natural Language Processing; Encoder Decoder Architecture

（责任编辑：王玮）