

# 基于大语言模型的命名实体识别<sup>①</sup>



叶名玮<sup>1</sup>, 汤嘉<sup>1</sup>, 郭燕<sup>1,2</sup>, 吴桂兴<sup>1,2</sup>

<sup>1</sup>(中国科学技术大学 软件学院, 合肥 230026)

<sup>2</sup>(中国科学技术大学 苏州高等研究院, 苏州 215123)

通信作者: 吴桂兴, E-mail: [gxwu@ustc.edu.cn](mailto:gxwu@ustc.edu.cn)

**摘要:** 虽然以 ChatGPT 为代表的自然语言生成 (NLG) 大语言模型在自然语言处理中的大多数任务中取得了良好的表现, 但其在序列识别任务, 如命名实体识别任务中的表现暂且不如基于 BERT 的深度学习模型. 针对这一点, 本文探究性的通过将现有的中文命名实体识别问题改造成机器阅读理解问题, 提出并设计了基于情境学习和模型微调的新方法, 使 NLG 语言模型在识别命名实体达到了更好的效果, 并且该方法不同于其他方法需要改变基层模型的预训练参数. 同时, 由于命名实体是模型生成的结果而不是对原始数据的分类, 不存在边界问题. 为了验证新框架在命名实体识别任务上的有效性, 我们在多个中文命名实体识别数据集上进行了实验. 其中, 在 Resume 和 Weibo 数据集上的  $F1$  分数分别达到了 96.04% 和 67.87%, 相较于 SOTA 模型分别提高了 0.4 和 2.7 个百分点, 从而验证了新框架能有效利用 NLG 语言模型在文本生成上的优势完成命名实体识别任务.

**关键词:** 命名实体识别; 模型微调; 机器阅读理解; 情境学习; 大语言模型

引用格式: 叶名玮, 汤嘉, 郭燕, 吴桂兴. 基于大语言模型的命名实体识别. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9586.html>

## Named Entity Recognition Based on Large Language Model

YE Ming-Wei<sup>1</sup>, TANG Jia<sup>1</sup>, GUO Yan<sup>1,2</sup>, WU Gui-Xing<sup>1,2</sup>

<sup>1</sup>(School of Software Engineering, University of Science and Technology of China, Hefei 230026, China)

<sup>2</sup>(Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215123, China)

**Abstract:** While natural language generation (NLG)-based large language models, represented by ChatGPT, perform well in various natural language processing tasks, their performance in sequence recognition tasks, such as named entity recognition, is somewhat inferior to that of bidirectional encoder representations from Transformer (BERT)-based deep learning models. To address this issue, this study first transforms the existing Chinese named entity recognition problem into a machine reading comprehension problem. A new name entity recognition method based on in-context learning and fine tuning is proposed, thereby enabling NLG-based language models to achieve good results in named entity recognition without changing base model pre-training parameters. Additionally, since named entities are generated by the model rather than classified from original data, there are no boundary issues. To verify the effectiveness of the new framework on named entity recognition tasks, experiments are conducted on some Chinese named entity recognition datasets. On the Resume and Weibo datasets, the  $F1$  scores reach 96.04% and 67.87% respectively, a gain of 0.4 and 2.7 percentage points over the state-of-the-art models, confirming that the new framework can effectively utilize the text generation advantages of NLG-based language models to complete named entity recognition tasks.

**Key words:** named entity recognition (NER); fine tuning of models; machine reading comprehension; in-context learning; large language model (LLM)

<sup>①</sup> 基金项目: 江苏省自然科学基金面上项目 (BK20161209)

收稿时间: 2024-02-03; 修改时间: 2024-02-29; 采用时间: 2024-03-19; csa 在线出版时间: 2024-07-03

如今,在高速发展的人工智能领域,抽取非结构化文本中所包含的信息是自然语言处理 (natural language processing, NLP) 领域重要的基础工作. 现阶段信息抽取领域主要有 3 个子任务, 分别是命名实体识别 (named entity recognition, NER)、关系抽取 (relation extraction) 和事件抽取 (event extraction). 其中命名实体识别一般被建模为序列标注模型, 旨在通过序列标注的方法, 对文本中的人名、地名、组织名等特定类型进行提取和识别, 而命名实体识别的效果会直接影响信息抽取的其他任务.

统计学和机器学习是早期命名实体识别任务主要使用的解决方法. 近年来, 深度学习中的神经网络模型在各 NLP 任务中取得了良好的效果, 因其不再像早期的方法那样依赖大量的手工特征工程以及专家知识, 被广泛应用于命名实体识别任务中. 其中的代表模型就是: 基于 Transformer<sup>[1]</sup> 的双向编码器表达 (bidirectional encoder representation from Transformers, BERT<sup>[2]</sup>) 预训练语言模型. 现阶段很多命名实体识别任务的当前最佳 (state-of-the-art model, SOTA) 模型都是基于 BERT 或其他预训练模型来提取语句中的特征向量, 然后搭建 LSTM 层 (long short-term memory)、CRF (conditional random field) 层或其他神经网络结构层. 这些模型无一例外都依赖于 BERT 等预训练模型的自然语言理解 (natural language understanding, NLU) 能力. 与此同时, 随着 ChatGPT<sup>[3]</sup> 在对话任务中的显现出令人惊叹的效果, 研究者不断发掘大语言模型 (large language model, LLM) 的潜力. 目前大语言模型的应用领域不断拓展, 已经在自然语言生成 (natural language generation, NLG) 领域, 如摘要提取、广告文案生成和交流对话等任务<sup>[4]</sup>中, 达到甚至超越了普通人的表现. 然而, 在以命名实体识别为代表的序列任务中, 以 NLG 为导向的大语言模型有其天然的劣势. 因为该类模型的预训练过程往往侧重于问答、续写等文本生成任务, 所以在完成序列预测任务时就会出现很多问题, 比如预测结果与句子的长度不匹配, 有奇怪的续写等问题.

近年来, 机器阅读理解 (machine reading comprehension, MRC) 任务<sup>[5]</sup> 由于其优秀的泛用性, 研究者们可以将各种 NLP 任务改造成 MRC 任务来预测结果, NER 任务也不例外. 因此, 受 MRC2NER<sup>[6]</sup> 的启发, 本文研究了基于大语言模型的命名实体识别方法. 该方法首先对传统的命名实体识别数据集进行改造, 将其变成

NLG 语言模型擅长的问答对话数据集; 然后将各个回答的答案实体返回到序列标记中; 最后通过冲突检测来解决扁平命名实体识别任务中出现嵌套实体的情况. 本文在多个中文命名实体识别数据集上进行了实验, 并在 Resume<sup>[7]</sup> 和 Weibo<sup>[8]</sup> 上的 F1 分数超过了 SOTA 模型, 从而验证了该方法的有效性. 总的来说, 本文的主要贡献如下: (1) 通过重新构建训练数据将序列标注任务变成机器阅读理解任务, 有效避免了 NLG 语言模型仅善于语言生成领域的缺陷. (2) 提出了基于 NLG 语言模型的情境学习和模型微调两种方法, 在实体识别任务上可以达到甚至超过 SOTA 的效果. (3) 本文提出的方法不需要改变预训练模型的初始权重, 可以有效适配其他任务.

本文首先描述命名实体识别和大语言模型的相关工作. 第 2 节给出了本文方法的总体流程并进行具体阐述. 第 3 节分别给出了该方法的实验结果并加以分析. 最后总结全文, 并展望未来的研究方向.

## 1 相关工作

### 1.1 命名实体识别

命名实体识别任务是自然语言处理领域许多任务的上游任务, 它的处理效果直接影响着其他领域的任务, 如信息抽取、知识图谱等. 在早期传统的命名实体识别研究中, Li 等<sup>[9]</sup>、Konkol 等<sup>[10]</sup> 在命名实体识别任务中加入了 CRF, 分别在印地语、捷克语命名实体识别任务并取得了不错的效果. Collobert 等<sup>[11]</sup> 将卷积神经网络 (convolutional neural network, CNN) 对隐藏向量进行特征提取, 并且构造了一种句子级别的似然函数来估算模型的效果, 取得了不错的效果. 继 Transformer 和 BERT 提出之后, 将预训练模型与其他神经网络相结合成为研究热点. 该方法可以大大减少模型的训练成本, 只需要在特定任务上进行模型微调取得超出当时 SOTA 的成绩. 而与 BERT 同时期出现的 GPT (generative pre-train) 系列模型, 由于其更善于文本生成方向的任务, 故很少有将其应用在命名实体识别任务上的研究. 然而, 随着 2023 年末 ChatGPT 的亮相, 以 GPT 为代表的大语言模型开始被广泛关注, 引起了新一轮的技术变革和研究热潮.

命名实体识别任务通过数据中实体是否嵌套可分为嵌套命名实体识别任务和扁平命名实体识别任务. 如图 1 所示, 在“昨日中国外交部严厉声明”这句话中,

如果我们的命名实体识别任务中需要同时识别出句子中出现的政治实体和组织实体,那么嵌套任务中会将“中国”一词分类到两个实体类别中,而扁平任务中仅仅根据最长分词原则,将“中国外交部”整个短语分类到组织类别中.目前嵌套命名实体识别任务<sup>[12]</sup>的中文数据集较少,主要集中在医学<sup>[13]</sup>、军事<sup>[14]</sup>等垂直领域,因此本文主要研究大语言模型在扁平中文命名实体识别任务中的实现与效果.

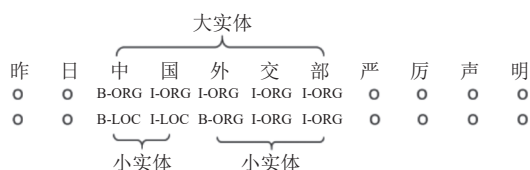


图1 嵌套命名实体和扁平命名实体

### 1.2 大语言模型

近年来,人工智能领域竞争的焦点之一就是通用人工智能 (artificial general intelligence, AGI), 其中取得的成绩斐然,已经在多个深度学习领域引起了革命性的变化.这些模型,如 GPT-3<sup>[15]</sup>、BERT 和 Transformer,已被广泛应用于各种任务,包括文本生成、翻译、摘要与问答系统等.研究者们<sup>[16,17]</sup>在近些年的研究进一步推动了这些模型在理解复杂语境和生成连贯文本方面的能力.这些研究展示了大语言模型在理解自然语言处理复杂查询和生成高质量回答方面的巨大潜力.

目前应用大型语言模型到下游任务的策略主要分为两种:微调 (fine-tuning) 和情境学习 (in-context learning, ICL). 微调策略涉及将预训练的模型用作特定任务的训练起点,然后在特定的下游数据集上进行额外的训练.这一策略已被 Raffel 等<sup>[18]</sup>和 Roberts 等<sup>[19]</sup>在 2018–2020 年间广泛探索,并取得了显著的成效.模型微调可以使模型能够更好地适应特定任务的需求,从而在各种特定应用中实现优化性能.大语言模型适配下游任务的另一种策略是情境学习,该方法由 Radford 等<sup>[20]</sup>于 2019 年首次提出.情境学习不需要额外的训练周期,它依赖于模型已有的知识和能力,通过在少数示例的指导下,引导模型生成特定任务的文本. Perez 等<sup>[21]</sup>、Lu 等<sup>[22]</sup>和 Rubin 等<sup>[23]</sup>在 2021 年的研究中指出,改进的提示和示例可以显著提升情境学习的效果.

## 2 方法介绍

NLG 语言模型完成命名实体识别任务之前,需要

对传统的 NER 数据集进行改造,将其适配成问答类型的数据集.进行数据改造后,本文提出了以下两种方案来完成命名实体识别任务:一种模型微调,通过训练下游任务的低阶适应模型 LoRA,本文提出的方法不需要改变预训练模型的参数;另一种是情境学习,即给予模型相似的任务案例,使用原模型直接完成任务.本文先通过微调方式,将得到的结果与其他深度学习模型训练后的结果进行比对;然后通过给予最常出现的几种类别实体不同的情境案例数量,研究情境学习在命名实体识别上的潜力.

### 2.1 数据集改造

扁平命名实体识别数据集常采用 BIO 标注方式,数据中的每个元素分别标注为“B-X”“I-X”或者“O”中的一个.“B-X”“I-X”分别表示此元素所在的数据片段属于 X 类型并且此元素在该实体的开头、中间位置,“O”表示该元素不属于任何类型.给定一个文本序列  $X = \{x_1, x_2, \dots, x_n\}$ , 其中  $n$  表示文本的长度,我们需要找出 X 中存在的所有命名实体,并将其分类成给定好的  $y$  类,  $y$  属于 Y. Y 是定义好的一系列可能的类别 (比如任务、地点与机构).

数据集结构如下:首先我们需要将标签风格的命名实体识别数据集转化为一组 (数据, 问题, 答案) 的三元组,对每一个类别,我们需要有一系列的问题  $Query = \{q_1, q_2, \dots, q_m\}$ ,  $m$  是类别的个数.标注的实体的列表是  $Answer = \{a_1, a_2, \dots, a_m\}$ , 每一个  $a_i$  代表一个回答.如果存在  $q_i$  类别,则  $a_i$  存储文中出现的实体;如果没有  $q_i$  类别的实体,则存储“无该类实体”.我们通过构建  $(text, q_i, a_i)$  三元组来提取源文本中出现的信息.如表 1 所示,在训练数据“京东方在合肥建立新厂房”中,我们的命名实体识别任务需要提取组织机构、地理位置和人物.因此,分别构建 3 个对话,以问答的方式将 NER 数据集中的原始结果分别分到 3 个回答中.如果某条数据不存在某一类别,则统一一个回答.如表 1 中在询问输入语句中是否含有人物实体类型时,由于“京东方在合肥建立新厂房”这句话中并未出现人物,所以回答为:“无该类实体”.

问题构建步骤中的  $q$  在 MRC 任务中是十分重要的,它等同于其他 NLP 任务中的提示 Prompt<sup>[24]</sup>.随着预训练模型+微调和预训练模型+Prompt 在 NLP 任务中的流行, Liu 等<sup>[24]</sup>发现对于同一个任务,使用不同的 Prompt 会导致不同的模型效果.因此某种程度上说,



Prompt 也是一种超参数. 我们构建了一个 Prompt 类, 以便于我们可以很便捷地更改 Prompt, 来调整得到最好的结果. 针对数据中需要识别的类型, 分别构建不同的 Prompt. 本文方法与 GPT-NER 的不同之处在于其输出整个测试语句, 使用特殊符号比如“@”“#”来标记识别出的实体. 然而在输入数据过长时, 输出的测试句往往会与输入的测试句存在着一定的偏差. 通过直接输出实体则可以有效解决这个问题.

表 1 以“京东方在合肥建立新厂房”为例创建三元组

实体类型	$q$	$a$
组织机构	文中提到了哪些组织机构? 比如大学、公司、俱乐部	京东方
地理位置	数据中提到了哪些地点? 如自然景点、行政区域、人造建筑	合肥
人物	数据中提到了哪些人物? 如人名, 称呼, 外号等	无该类别实体

## 2.2 模型微调方式

随着深度学习的模型越来越大, 性能越来越强, 将预训练模型在下游任务上进行微调就可以有效适配新任务. 然而, 如果更改了预训练模型的权重, 就会导致模型的泛用性变差. 因此, 我们基于大模型的低阶适应 (low-rank adaptation, LoRA)<sup>[25]</sup> 来进行有监督学习, LoRA 微调不会更改预训练模型的权重. 假设预训练模型  $P_{\Phi}(y|x)$  的初始权重为  $\Phi$ , 改造后的 NER 任务是一系列的对话:  $Z = \{(x_i, y_i)\}, i = 1, \dots, N$ , 其中  $x_i$  是 Prompt+训练数据,  $y_i$  是数据中的命名实体. 在全局微调过程中, 通过训练数据将预训练模型的初始权重  $\Phi_0$  升级到  $\Phi_0 + \Delta\Phi$  来最大化条件语言建模目标, 如式 (1) 所示:

$$\max_{\Phi} \sum_{(x,y) \in Z} \sum_{t=1}^{|y|} \log(P_{\Phi}(y_t|x, y < t)) \quad (1)$$

在全局微调中, 学习到的新参数增量  $\Delta\Phi$  和预训练模型的  $\Phi_0$  是一致的, 会大大提高我们的资源需求. 而使用低阶适应, 使用较少的参数  $\Theta$  也能达到同样的效果, 即  $\Delta\Phi = \Delta\Phi(\Theta)$ , 其中  $|\Theta| \ll |\Phi_0|$ . 优化目标就变成寻找  $\Theta$ , 如式 (2) 所示:

$$\max_{\Theta} \sum_{(x,y) \in Z} \sum_{t=1}^{|y|} \log(P_{\Phi_0 + \Delta\Phi(\Theta)}(y_t|x, y < t)) \quad (2)$$

设输入向量  $x$  为长度为  $d$  的向量, 在模型的微调过程中, 往往会通过下游任务的训练数据改变模型的权重  $W$ . 如式 (3) 所示,  $W_0x$  表示预训练模型的原始计

算结果, 通过  $\Delta Wx$  来适配新的任务. LoRA 通过更新两个低维度的矩阵, 以求达到与  $\Delta W$  相同的效果.

$$h = W_0x + \Delta Wx = W_0x + BAx \quad (3)$$

其中,  $B \in R^{d \times r}, A \in R^{r \times d}$ , 而  $r$  远小于  $d$ .

## 2.3 情境学习

情境学习方式需要找到与预测数据相似的训练数据, 并提供给模型进行参考, 这种模型的好处在于不会修改模型本身, 通过发掘模型的潜力来适配任务, 所以, 参考案例的数量和质量直接影响着后续的训练结果. 以 Resume 数据集中所构造的 Prompt 为例, 如下所示.

### 任务描述

你是一位命名实体识别的标注人员, 请你根据下面案例, 输出最后一条数据中出现的组织机构. 比如大学、公司、俱乐部等.

### 案例展示

数据 1: 就读于兰州商学院会计学专业, 中国社科院研究生院国际贸易专业硕士研究生

回答 1: 兰州商学院|中国社科院研究生院

数据 2: 无境外永久居留权, 本科学历.

回答 2: 无该类别实体

...

数据  $k$ : 12 月进入双成药业财务部工作, 历任主管会计、财务部副经理等职位.

回答  $k$ : 双成药业

### 测试句输入

测试数据: 曾任华中理工大学机械厂财务部财务经理.

### 期望回答

回答: 华中理工大学机械厂

如上例所示, 整个 Prompt 分成 3 个部分: 任务描述、案例展示和测试句输入. 任务描述中, 通过尝试不同的提示词来达到更好的效果. 案例的质量直接影响着结果的好坏, 我们基于 SimCSE 模型<sup>[26]</sup>来从训练数据中抽取相似的案例, 作为测试数据的相似情境, 供大语言模型参考. 使用 KNN 方法来选取最相近的  $K$  个案例. 测试句输入就是测试集中的某条语句, 通过对比期望回答和测试集答案来计算模型识别的准确率.

由于每个类别的预测过程不是同时的, 所以在处理扁平命名实体识别任务时, 会出现输出的实体产生冲突的情况. 如在这句话中“小明曾任石油公司市场部副经理”. 我们与模型对话后将其输出结果进行数据汇总时会发生冲突:

提问: ...输出数据中出现的组织机构?...

回答: 石油公司市场部

提问: ...输出数据中出现的职位?...

回答: 市场部副经理

在上述例子中,我们可以看见由于每个类别的实体并非是同时输出的,导致“市场部”同时出现在了组织类别和职位类别之中,因此,需要增添一个判断的部分来解决分歧的地方。

### 3 实验结果与分析

实验使用的环境是 Ubuntu 20.4 操作系统, GPU 为 NVIDIA GeForce RTX4090, 深度学习框架为 PyTorch, 使用的优化器为 Adam, 微调时最大输入长度设置为 256, Batch size 设置为 8, 初始学习率为 0.02. 实验所采用的语言模型为 ChatGLM<sup>[27,28]</sup>.

#### 3.1 数据集介绍

本文实验数据集为 MSRA、Resume 和 Weibo 这 3 个中文命名实体识别公开数据集. MSRA 数据集来自于第 4 届国际中文处理竞赛, 有人物、地点和组织 3 种实体类型. Resume 在 2018 年由 Zhang 等<sup>[29]</sup>构建, 分别标注了国籍、学历、姓名、籍贯、组织、专业、民族和职位 8 种实体类别. Weibo 中文数据集不仅标注了 4 种实体: 人物、地点、组织和地缘政治, 同时还将其分类成命名实体 (named entity) 和广义实体 (nominal mention) 两种类别. 数据集详情如表 2 所示.

表 2 实验数据集

语料库名称	数据集大小 (千字符)	实体类别数	实体数 (k)
MSRA	2 341	3	81
Resume	153	8	16
Weibo	103	4	3

#### 3.2 对比模型介绍

为了验证模型的有效性, 选取了以下对比模型:

(1) Lattice LSTM+CRF<sup>[29]</sup>: Zhang 等在 2018 年首次引入词典信息处理中文命名实体识别任务, 设计出了 Lattice LSTM 结构.

(2) LR-CNN+CRF<sup>[30]</sup>: Gui 等在 2019 年设计了 Rethinking 机制, 提出了 LR-CNN 模型.

(3) LGN+CRF<sup>[31]</sup>: Gui 等在 2019 年将 CNER 任务转化为图节点分类问题, 从而提出了 LGN (lexicon-based graph neural network) 模型.

(4) FLAT<sup>[32]</sup>: Li 等在 2020 年将 lattice 结构从一个有向无环图展平为 flat-lattice Transformer 结构, 从而提出了 FLAT (flat-lattice Transformer) 网络模型.

(5) NFLAT<sup>[33]</sup>: Wu 等在 2022 年将 FLAT 模型进行改造, 使提出的 NFLAT 模型具有词汇融合和上下文

特征编码解耦能力, 减少了不必要的“word-character”和“word-word”的 attention 计算.

本文采用精确率 (Precision), 召回率 (Recall) 和 F1 分数 (F1) 作为模型的评价指标, 具体公式如下:

$$Precision = \frac{correct}{pred\_entities\_num} \quad (4)$$

$$Recall = \frac{correct}{gold\_entities\_num} \quad (5)$$

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (6)$$

其中, correct、pred\_entities\_num 和 gold\_entities\_num 分别表示预测正确的实体数量、预测的实体总数量和样本中的实体总数量.

#### 3.3 模型的性能对比

为了验证本文提出的方法在命名实体识别任务上的有效性, 本节将通过实验结果与其他模型进行比较, 实验结果如表 3 所示. 其中 Weibo 数据集的 NE、NM 分别为 named entity 和 nominal mention 两种类型的 F1 分数. 在 MSRA 数据集中, 由于实体分布较为分散, 所以并未采用 ICL 对输出结果进行判别. 从表 3 可以看出, 相比其他深度学习模型, 我们的实验模型在 Resume 和 Weibo 两个数据集上均表现最优. 同时在 MSRA 数据集上已经接近 SOTA 模型 NFLAT 的成绩. 具体而言, 当没有使用情境学习时, 我们的模型已经在 Resume 和 Weibo 上的 F1 分数上接近甚至超过了其他模型的表现. 当我们引入情境学习进行最后的问答结果校正后, 可以发现: 在 Resume 数据集上, 我们的方法已经相较于 LSTM、CNN、GNN 和 Transformer 架构的 SOTA 模型在 F1 分数上分别提高了 1.58%、0.93%、0.67% 和 0.18%; 而在 Weibo 数据集中, F1 分数分别提高了 9.08%、7.95%、7.67% 和 4.45%. 实验表明本文提出的方法可以胜任命名实体识别这样的序列标注任务.

#### 3.4 情境学习的表现

为了研究在不改变模型权重的情形下, 仅通过情境学习来测试大语言模型在命名实体识别任务中的表现, 我们选取了 Resume 数据集中出现最多的 3 类实体 (职位、组织和学历) 进行实验. 人物实体由于具有特殊的信息结构, 几乎所有的模型都能很好地提取出人物实体, 因此排除在考察的范围内. 在表 4 中, Random 表示从训练数据中随机找出 K 个样本 (重复 5 次, 取均值), Find\_Sim 为基于 SimCSE 方法, 计算文本相似度

查找  $K$  个与测试数据相近的详细案例进行情境学习之后得到的结果. 我们将本文方法和 GPT-NER 进行效果比对. 实验结果如表 4 所示. 从表 4 可以看出, 随着情境学习给予的样本越来越多, 可以观察到 3 个类别的准确率不断上升. 当测试模型使用不同的实体提取策略时, 可以看出本文方法具有更好的表现, 这是由于 GPT-NER 方法中需要对整个测试语句进行重复输出, 并使用特殊符号标记目标实体, 这会增加了语言模型

的任务量, 使得任务效果变差; 而采用直接回答实体的策略则会接近大语言模型在预训练阶段的任务, 所以效果会更好. 同时, 在经过相似案例查找之后, 可以看出 3 个类别实体的  $F1$  分数数值在不断提高, 分别提高了 20.6%、13.7% 和 25.4%. 从表 4 的实验结果可以得知, 无论采用哪种策略提取实体、哪种策略查找相似案例, 使用更多的相似案例是可以有效提高实体识别的准确率的. 这与我们的认知相符.

表 3 不同模型在数据集上的表现 (%)

模型	MSRA			Resume			Weibo		
	Precision	Recall	F1	Precision	Recall	F1	NE	NM	F1
Lattice LSTM+CRF	93.57	92.79	93.18	94.81	94.11	94.46	53.04	62.25	58.79
LR-CNN+CRF	94.50	92.93	93.71	95.37	94.84	95.11	57.14	66.67	59.92
LGN+CRF	94.19	92.73	93.46	95.28	95.46	95.37	55.34	64.98	60.21
FLAT	—	—	94.12	—	—	<b>95.86</b>	—	—	<b>63.42</b>
NFLAT	94.92	94.19	<b>94.55</b>	95.63	95.52	95.58	59.10	63.76	61.94
Ours (without ICL)	94.74	94.03	<b>94.37</b>	96.30	95.33	<b>95.81</b>	60.43	67.29	<b>64.31</b>
Ours (with ICL)	—	—	—	<b>96.41</b>	<b>95.67</b>	<b>96.04</b>	<b>64.21</b>	<b>69.20</b>	<b>67.87</b>

表 4 情境学习的表现 (%)

K	方法	Random			Find_Sim		
		职位	组织	学历	职位	组织	学历
1 shot	GPT-NER	12.7	22.6	19.5	17.2	28.5	21.6
	Ours	21.4	32.3	27.9	37.1	44.2	38.4
2 shots	GPT-NER	14.8	27.9	24.7	22.4	38.2	27.8
	Ours	27.6	43.8	36.2	40.5	49.6	47.5
5 shots	GPT-NER	26.3	33.2	26.9	38.3	47.2	33.9
	Ours	33.7	47.2	40.3	53.7	56.4	55.4
10 shots	GPT-NER	30.6	37.4	30.5	47.2	53.6	47.6
	Ours	<b>40.2</b>	<b>53.6</b>	<b>44.9</b>	<b>60.8</b>	<b>67.3</b>	<b>70.3</b>

#### 4 结束语

本文描述了一种基于大语言模型实现中文命名实体识别的新方法. 通过基于一种机器阅读理解方法 MRC2NER, 我们解决了 NLG 语言模型在处理命名实体识别任务上存在的困难. 我们提出了模型微调和情境学习两种方法, 显著提高了大语言模型在命名实体识别上的性能. 实验结果表明, 本文提出的基于大语言模型并结合数据改造的方法, 在有监督学习环境中, 相较于其他的 NER 模型有着更好的表现. 此外, 情境学习在数据不足的情况下, 也能在命名实体识别任务上接近有监督学习的效果, 说明了大语言模型在处理此类任务时具有优秀的适应性和灵活性.

在未来面对实体类别增加、标注实体不断减少的挑战时, 我们将探讨粗细粒度实体识别和小样本实体识别问题. 粗细粒度实体识别涉及识别文本中不同层级的

实体, 这对 NER 系统提出了新的要求. 一种可能的应对方法是: 通过两级提问方法, 先后区分实体的粗粒度和细粒度, 从而提高系统的准确性和灵活性. 而小样本命名实体识别, 现在的研究方向主要在数据增强和类型实体聚类, 而 NLG 语言模型在这方面有很大的应用前景.

#### 参考文献

- 1 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 2 Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional Transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: ACL, 2019. 4171–4186.
- 3 OpenAI. Introducing ChatGPT. <http://openai.com/blog/chatgpt>. (2022-11-30).
- 4 赵朝阳, 朱贵波, 王金桥. ChatGPT 给语言大模型带来的启示和多模态大模型新的发展思路. 数据分析与知识发现, 2023, 7(3): 26–35.
- 5 丁美荣, 刘鸿业, 徐马一, 等. 面向机器阅读理解的多任务层次微调模型. 计算机系统应用, 2022, 31(3): 212–219. [doi: 10.15888/j.cnki.csa.008417]
- 6 Li XY, Feng JR, Meng YX, *et al.* A unified MRC framework for named entity recognition. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 5849–5859.
- 7 He HF, Sun X. F-score driven max margin neural network



- for named entity recognition in Chinese social media. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. Valencia: ACL, 2017. 713–718.
- 8 Peng NY, Dredze M. Named entity recognition for Chinese social media with jointly trained embeddings. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL, 2015. 548–554.
- 9 Li W, Mccallum A. Rapid development of Hindi named entity recognition using conditional random fields and feature induction. ACM Transactions on Asian Language Information Processing, 2003, 2(3): 290–294. [doi: [10.1145/979872.979879](https://doi.org/10.1145/979872.979879)]
- 10 Konkol M, Konopik M. CRF-based Czech named entity recognizer and consolidation of Czech NER research. Proceedings of the 16th International Conference on Text, Speech and Dialogue. Pilsen: Springer, 2013. 153–160.
- 11 Collobert R, Weston J. A unified architecture for natural language processing: Deep neural networks with multitask learning. Proceedings of the 25th International Conference on Machine Learning. Helsinki: ACM, 2008. 160–167.
- 12 张汝佳, 代璐, 王邦, 等. 基于深度学习的中文命名实体识别最新研究进展综述. 中文信息学报, 2022, 36(6): 20–35. [doi: [10.3969/j.issn.1003-0077.2022.06.002](https://doi.org/10.3969/j.issn.1003-0077.2022.06.002)]
- 13 王珅. 基于嵌套命名实体识别的肿瘤知识图谱构建 [硕士学位论文]. 北京: 北京交通大学, 2020.
- 14 关景文, 宋晓, 李晓庆, 等. 导弹领域文本嵌套命名实体识别方法研究. 系统仿真学报, 2023, 35(8): 1757–1767.
- 15 Brown T, Mann B, Ryder N, *et al.* Language models are few-shot learners. Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020, 33. 159.
- 16 Chowdhery A, Narang S, Devlin J, *et al.* Palm: Scaling language modeling with pathways. The Journal of Machine Learning Research, 2023, 24(1): 240.
- 17 Liu X, Zheng YN, Du ZX, *et al.* GPT understands, too. arXiv:2103.10385, 2021. [doi: [10.1016/j.aiopen.2023.08.012](https://doi.org/10.1016/j.aiopen.2023.08.012)]
- 18 Raffel C, Shazeer N, Roberts A, *et al.* Exploring the limits of transfer learning with a unified text-to-text Transformer. The Journal of Machine Learning Research, 2020, 21(1): 140.
- 19 Roberts A, Raffel C, Shazeer N. How much knowledge can you pack into the parameters of a language model? Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). ACL, 2020. 5418–5426.
- 20 Radford A, Wu J, Child R, *et al.* Language models are unsupervised multitask learners. OpenAI Blog, 2019, 1(8): 9.
- 21 Perez E, Kiela D, Cho K. True few-shot learning with language models. Proceedings of the 35th Conference on neural information processing systems. OpenReview.net, 2021. 11054–11070.
- 22 Lu Y, Bartolo M, Moore A, *et al.* Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Dublin: ACL, 2022. 8086–8098.
- 23 Rubin O, Herzig J, Berant J. Learning to retrieve prompts for in-context learning. Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Seattle: ACL, 2022. 2655–2671.
- 24 Liu PF, Yuan WZ, Fu JL, *et al.* Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 2023, 55(9): 195.
- 25 Hu EJ, Shen YL, Wallis P, *et al.* LoRA: Low-rank adaptation of large language models. Proceedings of the 10th International Conference on Learning Representations. OpenReview.net, 2021.
- 26 Gao TY, Yao XC, Chen DQ. SimCSE: Simple contrastive learning of sentence embeddings. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Punta Cana: Association for Computational Linguistics (ACL), 2021. 6894–6910.
- 27 Du ZX, Qian YJ, Liu X, *et al.* GLM: General language model pretraining with autoregressive blank infilling. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Dublin: ACL, 2022. 320–335.
- 28 Zeng AH, Liu X, Du ZX, *et al.* GLM-130B: An open bilingual pre-trained model. Proceedings of the 11th International Conference on Learning Representations. Kigali: OpenReview.net, 2022.
- 29 Zhang Y, Yang J. Chinese NER Using Lattice LSTM. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne: ACL, 2018. 1554–1564.
- 30 Gui T, Ma RT, Zhang Q, *et al.* CNN-based Chinese NER with lexicon rethinking. Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao: ijcai.org, 2019. 4982–4988. [doi: [10.24963/ijcai.2019/692](https://doi.org/10.24963/ijcai.2019/692)]
- 31 Gui T, Zou YC, Zhang Q, *et al.* A lexicon-based graph neural network for Chinese NER. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: ACL, 2019. 1040–1050.
- 32 Li XN, Yan H, Qiu XP, *et al.* FLAT: Chinese NER using flat-lattice Transformer. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 6836–6842.
- 33 Wu S, Song XN, Feng ZH, *et al.* NFLAT: Non-flat-lattice Transformer for Chinese named entity recognition. arXiv: 2205.05832, 2022.

(校对责编: 孙君艳)