

## 多层结构化语义知识增强的跨领域命名实体识别

张文韩<sup>1</sup> 刘小明<sup>1,4</sup> 杨 关<sup>1,2</sup> 刘 杰<sup>3,4</sup>

<sup>1</sup>(中原工学院计算机学院 郑州 450007)

<sup>2</sup>(河南省网络舆情监测与智能分析重点实验室(中原工学院) 郑州 450007)

<sup>3</sup>(北方工业大学信息学院 北京 100144)

<sup>4</sup>(国家语委中国语言智能研究中心(首都师范大学) 北京 102206)

(2020107234@zut.edu.cn)

## Cross-Domain Named Entity Recognition of Multi-Level Structured Semantic Knowledge Enhancement

Zhang Wenhan<sup>1</sup>, Liu Xiaoming<sup>1,4</sup>, Yang Guan<sup>1,2</sup>, and Liu Jie<sup>3,4</sup>

<sup>1</sup>(School of Computer Science, Zhongyuan University of Technology, Zhengzhou 450007)

<sup>2</sup>(Henan provincial Key Laboratory on Public Opinion Intelligent Analysis (Zhongyuan University of Technology), Zhengzhou 450007)

<sup>3</sup>(School of Information Science, North China University of Technology, Beijing 100144)

<sup>4</sup>(China Language Intelligence Research Center of the National Language Commission (Capital Normal University), Beijing 102206)

**Abstract** Cross-domain named entity recognition aims to alleviate the problem of insufficient annotation data in the target domain. Most existing methods, which exploit the feature representation or model parameter sharing to achieve cross-domain transfer of entity recognition capabilities and can only partially utilize structured knowledge entailed in text sequences. To address this, we propose a multi-level structured semantic knowledge enhanced cross-domain named entity recognition MSKE-CDNER, which could facilitate the transfer of entity recognition capabilities by aligning the structured knowledge representations embedded in the source and target domains from multiple levels. First, MSKE-CDNER uses the structural feature representation layer to achieve structured semantic knowledge representations of texts from different fields' structured alignment. And then, these structured semantic representations are aligned at the corresponding layers by a latent alignment module to obtain cross-domain invariant knowledge. Finally, this cross-domain consistent structured knowledge is fused with domain-specific knowledge to enhance the generalization capability of the model. Experiments on five datasets and a specific cross-domain named entity recognition dataset have shown that the average performance of MSKE-CDNER improved by 0.43% and 1.47% compared with the current models. All of these indicate that exploiting text sequences' structured semantic knowledge representation could effectively enhance entity recognition in the target domain.

**Key words** cross-domain named entity recognition; cross-domain transfer; structured alignment; structured knowledge; domain invariant knowledge

**摘 要** 跨域命名实体识别旨在缓解目标领域标注数据不足的问题。现有方法通常利用特征表示或者模型参数的共享来实现实体识别能力的跨领域迁移,但对文本序列中结构化知识的充分利用仍有所欠缺。基

收稿日期: 2022-05-21; 修回日期: 2023-02-20

基金项目: 国家重点研发计划项目(2020AAA0109700); 国家自然科学基金项目(62076167); 河南省高等学校重点科研项目(23A520022)

This work was supported by the National Key Research and Development Program of China (2020AAA0109700), the National Natural Science Foundation of China (62076167), and the Key Scientific Research Projects of Colleges and Universities in Henan Province (23A520022).

通信作者: 刘小明 (ming616@zut.edu.cn)

于此,提出了基于多层结构化语义知识增强的跨领域命名实体识别(multi-level structured semantic knowledge enhanced cross-domain named entity recognition, MSKE-CDNER)模型,即通过在多个层级实现对源领域和目标领域文本各自蕴含的结构化表示的对齐来促进实体识别能力跨领域迁移。首先,MSKE-CDNER利用结构特征表示层从不同领域中获取文本的结构化语义知识表示;然后,将获得的结构化语义知识表示通过潜层对齐模块在对应的层级进行结构化对齐,获取结构化的跨领域不变知识,从而提高模型对文本结构化知识的利用;此外,将域不变知识与特定域知识融合,进一步增强模型的泛化能力;最后,分别在5个英文数据集和特定的跨域命名实体识别数据集上进行实验。结果显示,对比当前跨域模型,MSKE-CDNER的平均性能提高了0.43%和1.47%,表明利用特征表示中的结构化知识可以有效提高目标领域的实体识别能力。

**关键词** 跨域命名实体识别;跨领域迁移;结构化对齐;结构化知识;域不变知识

**中图法分类号** TP391.1

命名实体识别是文本处理和信息抽取中的一个基本任务,其作用是识别文本中具有特殊意义的实体,是知识图谱、实体关系抽取、人机对话系统等的前置任务<sup>[1-2]</sup>。由于特定目标领域标注数据难以获取,领域之间存在数据分布和标注类型的差异,源领域上训练好的模型难以直接被应用于目标领域。所以,能够将模型应用于数据资源匮乏的跨域命名实体识别(cross-domain named-entity recognition, CD-NER)引起了广泛关注<sup>[3]</sup>。

CD-NER任务主要面临2个问题:1)数据资源匮乏导致模型无法充分学习隐藏的特征表示;2)由于不同领域中相同实体代表不同的含义,实体歧义导致模型迁移时出现负迁移。为了解决领域数据资源匮乏导致的模型效果不佳问题,Jia等人<sup>[4]</sup>设计了一个利用语言模型的参数生成网络作为跨域之间的领域自适应来执行跨域之间的知识迁移,同时通过学习相关领域的知识来增强模型的泛化能力;Wang等人<sup>[5]</sup>提出了一个跨领域的标签感知双迁移学习框架,通过少量标记数据将训练好的模型应用于特定目标领域。为解决实体歧义性问题,Jia等人<sup>[6]</sup>设计了一个多细胞组成的长短期记忆(multi-cell compositional long short-term memory, multi-cell LSTM)网络,对不同实体类型分别建模处理,在实体层面进行跨域知识的迁移。另有相关学者使用预训练与微调相结合、参数分享<sup>[7-8]</sup>等方法实现跨域迁移,也取得了显著效果。但是,文献[4-8]的方法通常利用语义特征之间的域不变知识来实现跨领域知识的迁移,对语义特征中的结构化知识未能很好地利用。

关于结构化知识,其在很多领域已有广泛研究<sup>[9]</sup>。例如,化合物和分子建模<sup>[10]</sup>、大脑连通性<sup>[11]</sup>、社交网络<sup>[12]</sup>中都包含了结构化知识。另外,Yang等人<sup>[13]</sup>提出迁移学习的一个关键问题是如何识别并利用不同

领域中共性的结构化知识来提升迁移的效果。如Alvarez-Melis等人<sup>[14]</sup>将结构化信息融入最优传输的框架,且框架利用特征表示中的结构化信息,在USPS和MNIST数字分类任务上取得了明显的提升。Chen等人<sup>[15]</sup>认为好的模型不仅需要考虑域之间的节点关系(特征信息),还应该考虑域内实体特征间的关系(结构化信息)。

本文认为文本表示中的结构化信息可以促进跨域任务中相关知识的迁移,同时提出将结构化语义知识应用于跨域命名实体识别任务,设计一种基于多层结构化语义知识增强的跨领域命名实体识别(multi-level structured semantic knowledge enhanced cross-domain named entity recognition, MSKE-CDNER)模型。具体而言,在MSKE-CDNER架构中,构建多层结构化对齐机制,在不同层中将具有相似结构的文本特征进行对齐,充分获取语义特征的结构化信息。对齐机制如图1所示,模型在嵌入层对文本的语义和句法特征进行对齐,获取和迁移结构化的语义表示;在隐藏层对上下文特征进行结构化的对齐,获取和迁移领域之间结构化的域不变知识,进一步提高目标领域实体识别能力。

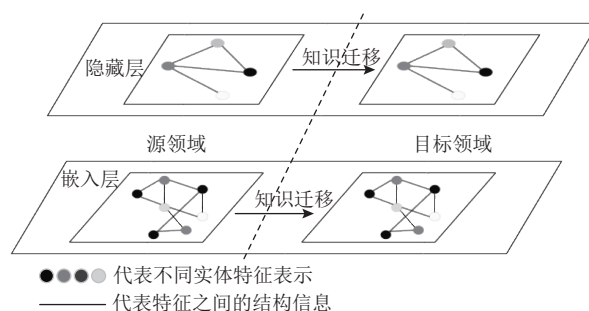


Fig. 1 Multi-level structural alignment mechanism for feature representation

图1 特征表示的多层结构化对齐机制

综上所述,本文的主要贡献有3点:

1) 提出一种利用文本序列中蕴含的结构化知识增强跨领域实体识别模型迁移能力的方法,利用结构化知识跨领域不变的稳定性促进跨领域知识的迁移.

2) 设计一个基于多层结构化语义知识增强的跨领域命名实体识别模型.对于多层结构化对齐机制,模型在嵌入层设计了结构特征表示模块,使用图最优传输(graph optimal transport, GOT)<sup>[15]</sup>获取实体特征的结构化知识,充分挖掘文本原语义信息;在隐藏层设计了潜层对齐模块,利用对齐度量准则获取结构化域不变知识,促进实体识别能力向目标领域迁移.

3) 通过在5个英文数据集和特定的跨域NER数据集上进行实验,表明利用特征表示中的结构化语义信息可以促进跨域命名实体识别中知识的迁移,验证了所提模型的有效性.

## 1 相关工作

本节将从跨域命名实体识别、领域自适应和结构化知识3个方面介绍相关工作.

### 1.1 跨域命名实体识别

跨域命名实体识别旨在利用从标注数据丰富的源领域中学习到的知识来提高目标领域的实体识别能力,因其可以缓解深度学习中数据依赖和训练数据不足的问题,引起了众多学者的关注.例如,张晗等人<sup>[16]</sup>通过引入生成式对抗网络和文档层面的全局变量,来解决领域内标注数据匮乏和实体歧义性问题. Jia 等人<sup>[4]</sup>设计了一个新的参数生成网络,将从命名实体识别或者语言模型任务中获得的参数分解成一个由元参数与任务嵌入向量和一个域嵌入向量组成的集合,通过域向量和任务向量之间的相似性来学习各个域和任务之间的相关性,进而在域之间进行知识的迁移. Li 等人<sup>[17]</sup>结合元学习和对抗学习的策略,获取序列标注中稳健、通用和可转移的特征,并用于领域自适应.该策略的主要优势在于它能够通过域中少量的数据快速适应新的领域. Chen 等人<sup>[18]</sup>将数据特征表示从高资源转移到低资源,从而达到利用高资源数据特征的目的,在低资源环境下取得了显著的提升.但是,文献[16–18]的方法仅考虑了序列的特征信息,对于特征信息中的结构化信息未能很好地利用.

### 1.2 领域自适应

领域自适应是迁移学习中的一个热点,其中学习域不变知识是无监督场景中的一个主流方法.由于实现预期质量和数量的标注数据消耗过高,所以领域自适应对于许多NLP任务非常重要.例如 Hao 等人<sup>[19]</sup>提出了一个半监督的解耦框架,通过使用3种互信息正则化项,分别最大化域特定潜变量和域不变潜变量与原始嵌入之间的互信息、最小化域特定潜变量与域不变潜变量之间的互信息,将特定域信息和域不变信息进行分解. Nozza 等人<sup>[20]</sup>提出了一种适应词嵌入的方法,通过增加用于源领域到目标领域特征空间适应的语义信息,来提高模型的领域自适应能力. 李鑫等人<sup>[21]</sup>提出一个更稳定的训练风险函数来分离出更优的真实特征,提高模型的泛化能力. Gu 等人<sup>[22]</sup>通过引入公共的编码器以及解码器、特定的编码器以及解码器来对领域内公共信息和特定领域信息进行显式建模,以便在域内训练时利用域外数据增强模型的泛化能力. Dong 等人<sup>[23]</sup>针对多模态中存在的语义差异和领域差异提出一个多级对齐网络,通过多级对齐网络减少相关差异、提高领域自适应能力. Zhang 等人<sup>[24]</sup>提出一种利用词级和篇章级域不变特征的领域自适应提取方法,并以渐进式的预训练模型提高域不变特征的提取能力,从而实现知识迁移.但是,该方法缺少对文本结构化知识的提取和利用.本文认为,跨领域迁移时,不仅要利用词汇、语句和篇章等各级的域不变特征,还应该进一步利用各层级的结构化域不变知识.因此,本文针对特征之间的结构差异和领域之间的差异设计了2个对齐模块,通过利用各层级的域不变知识来增强模型的领域自适应能力.

### 1.3 结构化知识

结构化知识因其具有稳定的跨领域不变性,可用于促进跨域知识的迁移<sup>[13]</sup>.例如, Swarup 等人<sup>[25]</sup>使用一个多层、稀疏连接的神经网络来学习任务间的结构表示,促进跨域知识的迁移. Lee 等人<sup>[26]</sup>利用表示学习识别图的结构化知识,然后迁移源领域中的结构化知识,构造一个不需要收集和重新训练数据的目标领域模型. Yang 等人<sup>[13]</sup>基于3种递进形式的实验,将复杂的结构化知识从源领域迁移到目标领域,充分证实了结构化知识可以促进跨域知识的迁移. Zheng 等人<sup>[27]</sup>提出一种标签结构迁移的跨域NER新方法,在源标签空间和目标标签空间中构造图,将跨域问题表示为图匹配问题,从而解决标签不



匹配问题。但是,该方法仅利用了标签中的结构化知识,对于特征表示中的结构化知识未能很好地利用。Courty 等人<sup>[28]</sup>指出,在进行领域自适应时,使用结构化信息约束可以将相同结构的特征匹配到目标空间中的紧凑区域,而不是不相交的区域。Chen 等人<sup>[15]</sup>提出了一个使用最优传输(optimal transport, OT)的跨域对齐框架图最优传输,在框架内同时使用 2 种不同类型的 OT 距离,将跨域对齐问题转化为图匹配问题,实现更好的跨域对齐。以视觉问答(visual question answering, VQA),任务为例,将 GOT 应用于跨模态对齐,可以学习到图像和问题之间的潜在对齐关系,更好地理解图像和问题的上下文。在 VQA 中使用 GOT 将跨模态对齐转化为图匹配问题,图匹配作为一个有效的正则化项来促进 VQA 的语义理解。图像和文本属于跨域问题中不同的领域,所以,在跨域任务中使用图最优传输可以进一步促进跨域知识的迁移。因此,在本文中引入 GOT 来获取并利用特征表示中的结构化知识,促进跨域知识的迁移,进而提高目标领域实体识别能力。

## 2 问题定义

跨领域序列标注:分别从源领域和目标领域中获取对应的数据集  $D_{\text{Src}} = (\mathbf{x}^{(i)}, \mathbf{y}^{(i)})_{i=1}^{N_{\text{Src}}}$  和  $D_{\text{Tgt}} = (\mathbf{x}^{(j)}, \mathbf{y}^{(j)})_{j=1}^{N_{\text{Tgt}}}$ , 其中  $N_{\text{Src}}$  和  $N_{\text{Tgt}}$  分别为数据集  $D_{\text{Src}}$  和  $D_{\text{Tgt}}$  的样本数量。对于任意的数据样本  $(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}) \in D_{\text{Src}}$ , 定义符号序列  $\mathbf{x}^{(i)} = (x_1, x_2, \dots, x_t)$  作为输入, 标签序列  $\mathbf{y}^{(i)} = (y_1, y_2, \dots, y_t)$  作为对应的输出, 其中下标  $t$  为数据样本的长度。

针对 CD-NER 任务,通常先在源领域  $D_{\text{Src}}$  上训练出一个实体识别模型  $M_{\text{Src}}$ , 任务的目标是在目标领域数据集  $D_{\text{Tgt}}$  上对  $M_{\text{Src}}$  进行训练, 得到一个目标领域上的实体识别模型  $M_{\text{Tgt}}$ 。模型  $M_{\text{Tgt}}$  能在给定输入序列和模型参数的情况下最大化标签序列中的条件概率分布, 即

$$P(\mathbf{y}|\mathbf{x}; \theta) = \prod_{t=1}^T P(y_t|x_t, y_{1:t-1}; \theta), \quad (1)$$

$$\begin{aligned} \mathbf{Y}^{\text{Src}} &= \arg \max_{\mathbf{y}^{\text{Src}}} P(\mathbf{y}^{\text{Src}}|\mathbf{x}^{\text{Src}}; \theta) \\ \Rightarrow \mathbf{Y}^{\text{Tgt}} &= \arg \max_{\mathbf{y}^{\text{Tgt}}} P(\mathbf{y}^{\text{Tgt}}|\mathbf{x}^{\text{Tgt}}; \theta), \end{aligned} \quad (2)$$

其中  $\mathbf{Y}^{\text{Src}}$ ,  $\mathbf{Y}^{\text{Tgt}}$  为经过最大化标签之后的输出序列,  $\mathbf{y}^{\text{Src}}$ ,  $\mathbf{x}^{\text{Src}}$  代表源领域的输出序列与输入序列,  $\mathbf{y}^{\text{Tgt}}$ ,  $\mathbf{x}^{\text{Tgt}}$  代表目标领域的输出序列和输入序列。

## 3 MSKE-CDNER 模型

本文提出了 MSKE-CDNER, 在文本表示的词嵌入层利用 GOT 获取特征表示中的结构化语义信息, 并在潜层空间中利用结构化对齐度量准则获取结构化域不变知识, 在多个层级通过结构化对齐促进跨域知识的迁移, 进而提高模型对结构化信息的利用。

### 3.1 模型架构

如图 2 所示, MSKE-CDNER 架构分为 4 层: 第 1 层为输入层, 由源领域、目标领域的数据集组成; 第 2 层为结构特征表示层, 包括特征表示层和结构对齐层, 用于将词向量进行编码并进行结构化对齐, 获取文本的结构化语义表示; 第 3 层为模型的主体部分, 由编码器、潜层结构化对齐等组成, 用于将文本的结构化语义表示对齐获取结构化域不变知识; 第 4 层为解码层, 对源领域和目标领域采取不同条件随机场来获取最优的结果。

### 3.2 多层结构化对齐机制

为了更好地识别、迁移特征表示中的结构化域不变知识, MSKE-CDNER 利用多层结构化对齐机制, 分别在嵌入层和隐藏层对特征表示进行结构化对齐。

#### 3.2.1 结构特征表示层

特征表示结构化对齐如图 3 所示。在对齐时, 不仅对相似的实体特征进行对齐(黑色实线对齐部分为节点对齐, 其中节点指的是句子中的实体), 而且对相似的特征关系也进行对齐(灰色虚线对齐部分为边对齐, 其中边指的是句子中实体和实体之间的相关信息), 使得模型在获取实体特征信息时能够获取到实体特征之间的结构化信息, 从而学习到更能代表原语义信息的特征表示。

模型在嵌入层使用 GOT<sup>[15]</sup> 作为一种结构化对齐方法, 获取特征表示中的结构化知识, 如图 4 所示。图最优传输针对跨域对齐结合了 2 种形式的最优传输距离, 一种用于节点(实体)匹配的沃瑟斯坦距离(Wasserstein distance, WD), 一种用于边(结构)匹配的 GWD(Gromov-WD)<sup>[29]</sup>, 使用 2 种结合的最优传输框架将跨域迁移转换为从一个域分布到另一个域分布的转移嵌入表示, 实现自我标准化对齐, 进而提高特征信息的可解释性。首先, 模型分别将源领域数据和目标领域数据送入嵌入层, 获取到对应的源领域语义特征  $\mathbf{e}_i^{\text{Src}}$  和目标领域语义特征  $\mathbf{e}_i^{\text{Tgt}}$ , 在结构对齐模块中, 对  $\mathbf{e}_i^{\text{Src}}$  和  $\mathbf{e}_i^{\text{Tgt}}$  进行规格化处理得到  $\mathbf{e}^{\text{Src}}$  和  $\mathbf{e}^{\text{Tgt}}$ , 计

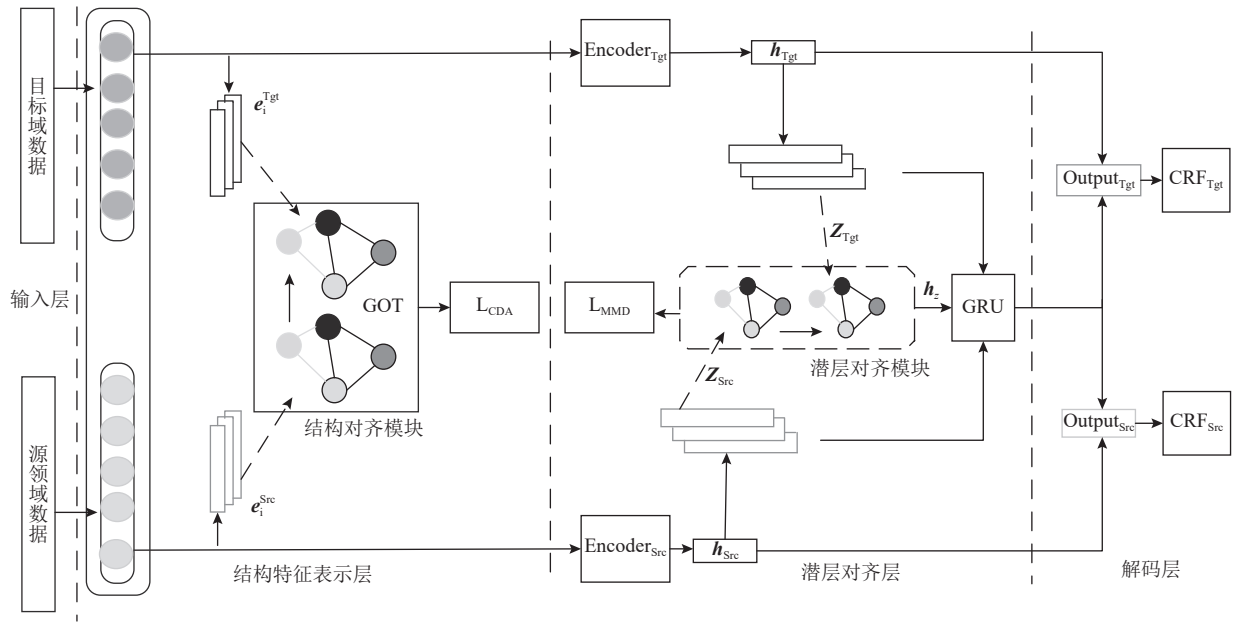


Fig. 2 Overall framework of MSKE-CDNER

图2 MSKE-CDNER 整体框架

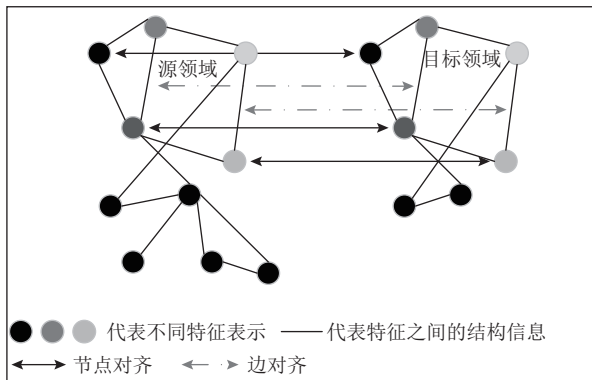


Fig. 3 Features represent structural alignment

图3 特征表示结构化对齐

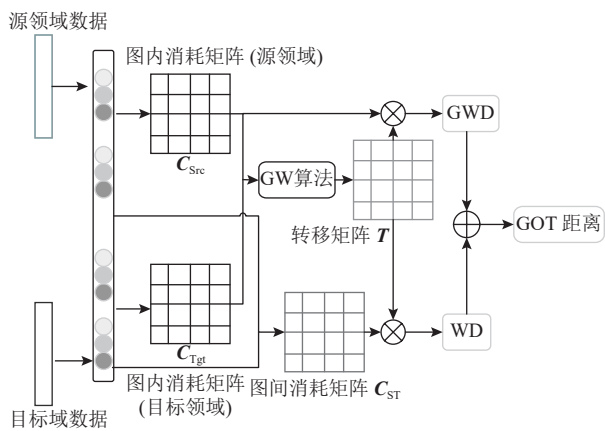


Fig. 4 Structural alignment layers

图4 结构对齐层

算规格化后的向量和原始特征向量的余弦相似度, 得到相似度矩阵  $C_{Src}$  和  $C_{Tgt}$ . 相似度矩阵包含实体和实

体之间的关联信息, 因此, 将其看作文本的结构化表示. 之后构建对应的图结构  $\mathcal{G}_x(\mathbf{V}_x, \mathbf{E}_x)$ , 其中节点  $i \in \mathbf{V}_x$  代表一个特征向量  $\mathbf{x}_i$ , 添加图中节点之间的余弦相似性(相似度矩阵)作为边  $\mathbf{E}_x$ . 为获取源领域和目标领域之间的相似性, 将源领域特征和目标领域特征进行余弦相似性计算, 得到跨域相似度矩阵  $C_{ST}$ . 具体计算为:

$$\mathbf{e}^{Src} = \frac{\mathbf{e}_i^{Src}}{\max(\|\mathbf{e}_i^{Src}\|_2, \epsilon)}, \quad (3)$$

$$\mathbf{e}^{Tgt} = \frac{\mathbf{e}_i^{Tgt}}{\max(\|\mathbf{e}_i^{Tgt}\|_2, \epsilon)}, \quad (4)$$

$$C_{Src} = 1 - \mathbf{e}^{Src} (\mathbf{e}^{Src})^T, \quad (5)$$

$$C_{Tgt} = 1 - \mathbf{e}^{Tgt} (\mathbf{e}^{Tgt})^T, \quad (6)$$

$$C_{ST} = 1 - \mathbf{e}^{Src} (\mathbf{e}^{Tgt})^T, \quad (7)$$

其中  $\epsilon$  设置为  $1E-12$ ,  $\max(\|\cdot\|_2, \epsilon)$  代表向量的 2 范数. 然后, WD 通过测量节点之间的距离, 用于对语义特征进行节点对齐; GWD 通过测量图中边之间的距离, 用于边节点对齐. 对于 WD 的计算有:

$$D_{WD}(\mu, \nu) = \inf_{\gamma \in \Pi(\mu, \nu)} \mathbb{E}_{(x, y) \sim \gamma} [c(x, y)] =$$

$$\min_{T \in \Pi(\mu, \nu)} \sum_{i=1}^n \sum_{j=1}^m T_{ij} \cdot c(x_i, y_j), \quad (8)$$

其中  $\mu$  和  $\nu$  为 2 个离散分布, 公式化表示为  $\mu = \sum_{i=1}^n u_i \delta_{x_i}$ ,

$\nu = \sum_{j=1}^m \nu_j \delta_{y_j}$ ,  $\delta_x$  为在  $x$  上的狄拉克函数.  $\prod(\mu, \nu)$  表示所有带有边缘  $\mu(x)$ ,  $\nu(y)$  的联合分布  $\gamma(x, y)$ ,  $u$  和  $\nu$  分别代表  $n$  维和  $m$  维权重向量;  $x, y$  分别表示传入的源领域和目标领域的语义特征;  $c(x_i, y_j)$  是评测  $x_i$  到  $y_j$  的消耗函数, 在此处选用余弦函数. 对于 GWD 的具体计算公式为:

$$D_{\text{GWD}}(\mu, \nu) = \inf_{\gamma \in \Pi(\mu, \nu)} E_{(x, y) \sim \gamma, (x', y') \sim \gamma} [L(x, y, x', y')] = \min_{\hat{T} \in \Pi(\mu, \nu)} \sum_{i, i', j, j'} \hat{T}_{ij} \hat{T}_{i'j'} L(x_i, y_j, x_{i'}, y_{j'}), \quad (9)$$

其中  $(x_i, x_{i'})$  和  $(y_j, y_{j'})$  分别代表源领域和目标领域中不同节点之间的边结构,  $L(\cdot)$  是评估不同域之间 2 对节点  $(x_i, x_{i'})$  和  $(y_j, y_{j'})$  结构相似性的消耗函数, 例如  $L(x_i, y_j, x_{i'}, y_{j'}) = \|c_1(x_i, x_{i'}) - c_2(y_j, y_{j'})\|$ ,  $c_1$  和  $c_2$  是在同一个图中评测节点相似性的函数, 在此处选取余弦函数, 矩阵  $\hat{T}$  为学习得到. 综上所述, 将图最优传输中计算得到的 2 种距离作为结构化对齐的损失函数  $L_{\text{CDA}}$ , 具体计算为

$$L_{\text{CDA}} = \lambda_1 \times D_{\text{WD}} + \lambda_2 \times D_{\text{GWD}}. \quad (10)$$

### 3.2.2 潜层对齐层

MSKE-CDNER 编码器使用双向门控循环单元 (gate recurrent unit, GRU) 神经网络提取文本序列中上下文特征表示, 同时编码器也适用于 Transformer Encoder 等通用方法. 为了在实验对比中更加公平地证明多层结构化对齐方法与现有方法对比的有效性, 在本文模型中未采用 Transformer Encoder 作为模型编码器. 为获取文本序列中的结构化域不变知识, 将获取的特征表示在潜层空间中利用 WD 进行度量对齐. 即把编码器获取的上下文特征  $h_{\text{Src}}$ ,  $h_{\text{Tgt}}$  映射到潜层空间, 在潜层空间中利用 WD 减少源领域、目标领域特征分布和标准正态分布之间的差异, 促进不同领域中共性知识的对齐, 从而获取域之间的不变知识. 特别地, 为了方便计算, 使用最大均值差异 (maximum mean discrepancy, MMD) 来近似 WD. 以源域为例, 具体计算公式为:

$$L_{\text{MMD}} = \left\| \int k(z_{\text{Src}}; \cdot) d p(z_{\text{Src}}) - \int k(z_{\text{Src}}; \cdot) d q(z_{\text{Src}}) \right\|_{H_k}, \quad (11)$$

其中  $z_{\text{Src}}$  是通过将  $h_{\text{Src}}$  映射到潜层空间中获得的,  $p(z_{\text{Src}})$  为先验假设的标准正态分布  $p(z_{\text{Src}}) = N(0, 1)$ ,  $q(z_{\text{Src}})$  表示为  $q(z_{\text{Src}}) = N(\mu_{z_{\text{Src}}}, \sigma_{z_{\text{Src}}}^2)$ ,  $\mu_{z_{\text{Src}}} = f(W_{\mu} h_n + b_{\mu})$ ,  $\sigma_{z_{\text{Src}}}^2 = f(W_{\sigma} h_n + b_{\sigma}^2)$ .  $[W_{\mu}; b_{\mu}]$ ,  $[W_{\sigma}; b_{\sigma}]$  分别表示  $\mu_{z_{\text{Src}}}$  和  $\sigma_{z_{\text{Src}}}^2$  可训练的参数.

在潜层空间中, 模型通过重构后的潜变量  $h_z$  获取

含有域不变知识的上下文语义特征表示. 即  $q(z_{\text{Src}})$  在潜层空间中经过重新参数化获取到新的潜变量  $z$ , 将  $z$  通过映射得到新的隐变量  $h_z$ , 然后将获取到的隐变量  $h_z$  和编码器的基本输出再次送入编码器 GRU, 重新学习获取序列的上下文特征表示, 从而获取含有结构化信息的域不变知识. 具体计算公式为:

$$h_{z_{\text{Src}}} = f(W_{z_{\text{Src}}} q(z_{\text{Src}}) + b_{z_{\text{Src}}}), \quad (12)$$

其中  $[W_{z_{\text{Src}}}; b_{z_{\text{Src}}}]$  为  $h_{z_{\text{Src}}}$  可训练的参数,  $b_{z_{\text{Src}}}$  为偏差.

### 3.3 解码层

模型的解码层针对源领域和目标领域采用不同标准的条件随机场 (conditional random field, CRF)<sup>[30]</sup>. CRF 是一个序列标注算法, 通过增强标签之间的约束, 获取序列全局最优解码<sup>[31]</sup>. 采用标准条件随机场中的负对数似然损失作为损失函数, 具体计算有:

$$L_{\text{ner}} = -\frac{1}{|D_{\text{ner}}|} \sum_{t=1}^N \lg(P(y_t | x_t)). \quad (13)$$

不同领域的实体类型不同, 但是边界信息是一样的, 可以适应于任何领域. 所以, 在实体识别任务中增加边界预测任务来学习共享知识中的边界信息. 对于实体边界预测的辅助任务使用交叉熵作为损失函数, 将其表示为  $L_{\text{bio}}$ , 具体计算有:

$$L_{\text{bio}} = -\sum_{i=1}^n P(x'_i) \lg Q(x_i). \quad (14)$$

综上所述, 模型相关损失函数可以定义为

$$L_{\text{ner}} = L_{\text{ner}}^{\text{Src}} + L_{\text{ner}}^{\text{Tgt}}, \quad (15)$$

$$L_{\text{bio}} = L_{\text{bio}}^{\text{Src}} + L_{\text{bio}}^{\text{Tgt}}. \quad (16)$$

### 3.4 优化目标

MSKE-CDNER 的整体损失函数表示为

$$Loss = L_{\text{ner}} + \alpha \times L_{\text{bio}} + \beta \times L_{\text{MMD}} + \gamma \times L_{\text{CDA}}, \quad (17)$$

其中  $\alpha, \beta, \gamma$  分别代表相关任务的权重.

### 3.5 优化算法

**算法 1.** MSKE-CDNER 的优化算法.

输入: 源领域和目标领域数据集  $S_{\text{ner}}, T_{\text{ner}}$ ;

输出: 适应于目标领域的模型  $M_{\text{Tgt}}$ .

① while 训练步骤没有结束

② for  $D$  in  $S_{\text{ner}}, T_{\text{ner}}$  do

③  $X, Y \leftarrow D$ ;

④ for  $x_i, y_j$  in  $X, Y$  do

/\*计算域内节点相似性\*/

⑤  $[C_{\text{Src}}]_{ij} = \cos(x_i, x_j)$ ; /\*式(3)(5)\*/

⑥  $[C_{\text{Tgt}}]_{ij} = \cos(y_i, y_j)$ ; /\*式(4)(6)\*/

/\*计算域外相似性\*/

⑦  $[C_{\text{ST}}]_{ij} = \cos(x_i, y_j)$ ; /\*式(7)\*/

```

⑧      计算  $L_{CDA}$ ; /*式(8)~(10)*/
⑨      end for
⑩      for  $x_t, y_t$  in  $X, Y$ 
⑪           $\{h_t, c_t\} = f_{\text{encoder}}(e_{1:N})$ ;
⑫      end for
⑬       $\{h_z, c_z\} \leftarrow \{q_z\} \leftarrow \{h_n, c_n\}$ 
⑭       $X' = (x'_1, x'_2, \dots, x'_n) \leftarrow \{h_z, c_z\}$ 
⑮      if  $D$  is  $S_{\text{ner}}$  then
⑯          计算  $L_{\text{ner}}^{\text{Src}}$ ; /*式(13)*/
⑰      else if  $D$  is  $T_{\text{ner}}$  then
⑱          计算  $L_{\text{ner}}^{\text{Tgt}}$ ; /*式(13)*/
⑲      end if
⑳      计算  $L_{\text{bio}}$ ; /*式(14)*/
㉑      计算  $L_{\text{MMD}}$ ; /*式(11)*/
㉒      计算  $L_{\text{ner}}, L_{\text{bio}}$ ; /*式(15)(16)*/
㉓       $L \leftarrow L_{\text{ner}} + \alpha \times L_{\text{bio}} + \beta \times L_{\text{MMD}} + \gamma \times L_{\text{CDA}}$ ;
㉔      end for
㉕      根据  $L$  更新网络参数  $\theta$ ;
㉖      end while

```

## 4 实 验

为了验证本文方法 MSKE-CDNER 对于 CD-NER 的有效性,在 5 个英文数据集和专门的跨域数据集上进行实验.分别从消融实验、显著性检测、参数分析、细粒度分析 4 个方面进行实验分析并展示实验效果.

### 4.1 实验数据

5 个英文数据集分别为 CoNLL-2003(Conll03), Twitter(T), Broad Twitter(BT), BioNLP13PC(PC), BioNLP13CG(CG).其中 CoNLL-2003, Twitter, Broad Twitter 数据集是相似的领域,实体类型大致类似,都包含人名(person, PER)、地名(location, LOC)、组织(organization, ORG),其中 CoNLL-2003 比 Twitter 多含有其他(miscellaneous, MISC)实体. BioNLP13PC 数据集和 BioNLP13CG 数据集属于医疗和生物领域,实体类型主要包含简单化学(simple chemical, CHEM)、细胞成分(cellular component, CC)、基因和基因产物(gene and gene product, GGP), BioNLP13CG 中还包括了物种(species, SPE)和细胞(cell, CELL),具体的数据集统计信息如表 1 所示.跨域数据集为 Liu 等人<sup>[1]</sup>提出的专门跨域 NER 数据集 CrossNER,其包含 5 个领域,分别为政治(politics)、自然科学(natural science)、音乐(music)、文学(literature)和人工智能

**Table 1 English Dataset Statistics**

**表 1 英文数据集统计信息**

数据集	类型	训练集数目	验证集数目	测试集数目
Conll03 <sup>[32]</sup>	句子	15 100	3 500	3 700
	实体	23 500	5 900	5 700
	单词	219 600	55 000	50 300
T <sup>[33]</sup>	句子	4 300	1 300	1 500
	实体	7 500	2 500	2 500
	单词	68 700	22 900	23 100
BT <sup>[34]</sup>	句子	6 300	1 000	2 000
	实体	8 800	1 700	4 300
	单词	106 300	16 000	37 400
PC <sup>[35]</sup>	句子	2 500	900	1 700
	实体	7 900	2 700	5 300
	单词	71 700	24 500	47 900
CG	句子	3 000	1 000	1 900
	实体	10 800	3 600	6 900
	单词	86 500	28 600	54 700

注: Conll03 (CoNLL-2003), T (Twitter), BT (Broad Twitter), PC (BioNLP13PC), CG (BioNLP13CG).

(artificial intelligence, AI),每个领域中含有特定的实体类型,具体的数据集统计信息如表 2 所示.

根据数据集中实体类型的不同以及相关领域的差异,可分为 2 组实验.第 1 组:从 5 个英文数据集中选取与 Jia 等人<sup>[6]</sup>相同的实验分组,例如,当 Twitter, Broad Twitter 作为目标领域数据集时,选用 CoNLL-2003 作为源领域数据.当 BioNLP13PC,作为目标领域数据集时,选用 CoNLL-2003 作为源领域数据集,从不同领域和相似领域之间进行实验来验证 MSKE-CDNER 在不同领域差异之间迁移的效果.第 2 组:选用与 CrossNER 中相同的实验分组将 5 个不同的特定

**Table 2 CrossNER Dataset Statistic**

**表 2 CrossNER 数据集统计信息**

领域	类型	训练集数目	验证集数目	测试集数目
政治	句子	200	500	600
	实体	1 300	3 400	4 200
自然科学	句子	200	400	500
	实体	1 000	2 500	3 000
音乐	句子	100	300	400
	实体	600	2 600	3 300
文学	句子	100	400	400
	实体	500	2 100	2 200
人工智能	句子	100	300	400
	实体	500	1 500	1 800



领域数据作为目标领域数据集, CoNLL-2003 作为源领域数据。

#### 4.2 实验设置

对于 5 个公开的英文数据集, 参照 NCRF++<sup>[36]</sup> 中模型的参数进行初始化设置。随着不同组实验中源领域和目标领域数据集的改变, 模型的参数也随之改变。例如, 源领域为 BioNLP13PC、目标领域为 BioNLP13CG 时, 优化器选用 SGD 算法, 学习率设置为 0.005, 学习率衰退设置为 0.01, 批次大小设置为 10, 隐状态维度为 250, 潜变量维度为 200, 为防止过拟合将 dropout 设置为 0.5。在实验中采用与 Jia 等人<sup>[6]</sup> 相同的初始化词向量和字符向量的方法。Twitter, Broad Twitter 作为目标领域数据集时, 使用 Glove 100-dim<sup>[37]</sup> 进行初始化获取词向量(word vector)的特征表示; 当 BioNLP13PC 和 BioNLP13CG 作为目标领域时, 词向量选用 PubMed 200-dim<sup>[38]</sup> 进行初始化。字符向量(char vector)采用随机初始化的形式, 通过卷积神经网络来提取字符特征表示, 最后将获取到的单词特征表示和字符特征表示进行拼接得到最终特征表示。

基于 CrossNER 数据集, 参照 5 个英文数据集上参数进行初始化设置, 经过参数调整发现, 在大多数领域中使用相同参数时, MSKE-CDNER 性能已较优, 证明模型鲁棒性较强。同时, 在某些域中模型参数随着目标域数据集的改变而微调时, 模型性能进一步提升。例如, 在 Music 域中优化器选用 SGD 算法, 学习率设置为 0.003, 学习率衰退设置为 0.03, 批次大小为 32, 隐状态维度为 250, 潜变量维度为 200, 为防止过拟合将 dropout 设置为 0.5。在实验中均使用 Glove 100-dim<sup>[37]</sup> 进行初始化获取词向量(word vector)的特征表示, Bert 优化获取到的词向量特征表示。字符向量采用随机初始化的形式, 通过卷积神经网络来提取字符特征表示, 将获取到的单词特征表示和字符特征表示进行拼接作为最终特征表示。

特别地, 在 5 个英文数据集的实验训练过程中, 当获取到目标领域的结束信号时标志着一个批次结束。源领域数据的读取操作不会因为批次的结束而进行重置, 它会继续加载数据, 直至加载到源领域数据的结束符号时再进行重置。在 CrossNER 数据集上进行实验时, 一个批次的结束替换为同时读取到源领域和目标领域数据中的结束符号。

#### 4.3 评测指标

本文采用与文献[4, 6]等一致的评测指标, 该指标认为只有当实体的类型与边界都识别正确时才认定预测准确。采用准确率(precision,  $P$ )、召回率(recall,

$R$ )和  $F1$  值计算最终得分。具体计算方式为:

$$P = \frac{TP}{TP + FP}, \quad (18)$$

$$R = \frac{TP}{TP + FN}, \quad (19)$$

$$F1 = 2 \times \frac{P \times R}{P + R}, \quad (20)$$

其中  $TP$  代表识别正确的实体个数,  $FP$  代表识别错误的实体个数,  $FN$  代表未识别出的实体个数。

#### 4.4 对比模型

为了验证 MSKE-CDNER 在跨域 NER 上的效果, 在不同的数据集上与相关模型进行对比实验。

1) BILSTM-CRF. BILSTM-CRF<sup>[39]</sup> 结合双向 LSTM 和条件随机场进行命名实体识别, 将源领域数据和目标领域数据结合, 共同训练模型。

2) Coach. Liu 等人<sup>[40]</sup> 提出了一个用 NER 域适配的框架 Coach, 它将任务分为 2 个阶段, 首先检测出实体, 然后对实体进行分类来解决特定领域数据稀缺问题。

3) MULTI-TASK+PGN. Jia 等人<sup>[4]</sup> 在源领域和目标领域中集成语言模型任务来执行跨域知识的迁移, 从而解决模型无法在无监督环境下进行训练的问题。

4) MULTI-TASK+GRAD. Zhou 等人<sup>[41]</sup> 提出一种新的传输方法, 通过对抗传输网络来进行高资源和低资源下特征的融合, 同时引入广义资源对抗判别器来提高模型的泛化能力。

5) MULTI-CELL-LSTM. Jia 等人<sup>[6]</sup> 基于 Bert 表示提出了一个多细胞 LSTM 结构, 针对不同实体类型分别建模, 在实体层面进行跨域知识的迁移, 解决实体在不同领域中含义不同的问题。

#### 4.5 实验结果

在 5 个英文数据集和 CrossNER 数据集上将 MSKE-CDNER 和其他相关方法进行实验对比, 结果如表 3 和表 4 所示。整体来看, MSKE-CDNER 在不同数据集上都取得了不错的结果。

如表 3 所示, MULTI-TASK(LSTM)相比于单任务的 BILSTM 在 Conll03→T 上  $F1$  值提高了 2.37%, 在 Conll03→BT 上  $F1$  值提高了 0.86%, 在 PC→CG 上  $F1$  值提高了 1.82%, 说明多任务架构可以提高 CD-NER 中实体识别能力。MULTI-TASK+PGN、MULTI-TASK+GRAD 和 MULTI-CELL-LSTM 相比于 MULTI-TASK(LSTM)在 PC→CG 上  $F1$  值分别提升了 0.11%, 0.57%, 0.95%, 说明充分学习、利用特征间的域不变知识可以缓解因数据资源缺乏导致的模型效果不佳问题。MSKE-CDNER 相比于当前热门模型 MULTI-CELL-



Table 3 F1 Experimental Results of the English Datasets

表3 英文数据集 F1 实验结果

%

方法	数据集				
	Conll03→T	Conll03→BT	Conll03→PC	Conll03→CG	PC→CG
BILSTM	77.18	72.98			79.24
MULTI-TASK (LSTM) <sup>[30]</sup>	79.55	73.84	83.09	77.73	81.06
MULTI-TASK+PGN <sup>[4]</sup>	80.07	73.70	85.54	79.86	81.17
MULTI-TASK+GRAD <sup>[41]</sup>	79.72	74.12			81.63
MULTI-CELL-LSTM <sup>[6]</sup>	80.86	74.83	86.26	80.74	82.01
MSKE-CDNER (本文)	<b>80.98</b>	<b>74.45</b>	<b>86.61</b>	<b>81.90</b>	<b>82.93</b>
MSKE-CDNER*	<b>81.32±0.07</b>	<b>74.51±0.04</b>	<b>86.76±0.13</b>	<b>81.86±0.14</b>	<b>82.85±0.31</b>

注: MSKE-CDNER\*数据表示实验平均值±标准偏差, 加粗数字表示 MSKE-CDNER 优于对比方法; Conll03 (CoNLL-2003), T (Twitter), BT (Broad Twitter), PC (BioNLP13PC), CG (BioNLP13CG).

Table 4 F1 Experiments Results of CrossNER Dataset

表4 CrossNER 数据集 F1 实验结果

%

方法	数据集					平均值
	Politics	Science	Music	Litera	AI	
BILSTM-CRF	56.60	49.97	44.79	43.03	43.56	47.59
Coach <sup>[40]</sup>	61.50	52.09	51.66	48.35	45.15	51.75
MULTI-TASK+PGN <sup>[4]</sup>	68.44	64.31	63.56	59.59	53.70	61.92
MULTI-CELL-LSTM <sup>[6]</sup>	70.56	66.42	70.52	66.96	58.28	66.55
MSKE-CDNER (本文)	<b>71.25</b>	<b>67.02</b>	<b>73.07</b>	<b>67.87</b>	<b>60.89</b>	<b>68.02</b>
MSKE-CDNER*	<b>71.15±0.06</b>	<b>66.84±0.16</b>	<b>72.62±0.42</b>	<b>67.51±0.27</b>	<b>60.58±0.47</b>	<b>67.74</b>

注: 加粗数字表示 MSKE-CDNER 优于对比方法; MSKE-CDNER\* 数据表示实验平均值±标准偏差.

LSTM 在 PC→CG 上 F1 值提高了 0.92%, 说明结构化语义知识可以促进跨域知识的迁移, 缓解实体歧义性的问题. 其中, 在 Broad Twitter 域中模型效果不佳, 考虑是因为 Broad Twitter 属于 Twitter 中的新闻领域, Conll03 也属于新闻领域, 领域之间的差异性较小, 在进行跨域对齐时图匹配对领域迁移之间的约束性较差, 导致迁移效果不佳. 而在 Conll03→PC 这组实验中, BioNLP13PC 属于医疗领域, Conll03 数据和 BioNLP13PC 这两者之间的数据差异性较大, 在进行跨域迁移时图匹配对其约束性较好. 实验对比表明: 领域之间差异性越大, 模型迁移效果越好, 这也更能说明语义特征中的结构化信息能够促进跨域知识的迁移, 领域差异越大, 结构化知识的约束作用越强, 迁移效果越佳. 然而, 现有研究方法中缺少对该类信息的挖掘和利用, MSKE-CDNER 的多层次结构迁移方法, 可以利用结构化信息实现了对模型跨领域迁移能力的增强.

在特定的 CD-NER 数据集 CrossNER 上进行验证, 结果如表 4 所示. MULTI-CELL-LSTM 相比 BILSTM-CRF 在 5 个不同领域中的 F1 值均有所提升, F1 平均

值提升了 18.96%. 由于 BILSTM-CRF 为单任务模型, 不能很好地利用源域中跨域不变的知识, 而 MULTI-CELL-LSTM 以多任务架构为基础搭建网络, 能够充分地利用源域中跨域不变的知识, 所以构建 MSKE-CDNER 时采用多任务学习范式为基础框架. 相比 MULTI-CELL-LSTM, MSKE-CDNER 在 5 个不同领域中 F1 值均有明显提升. 其中在 Politics 中 F1 值提高了 0.69%, 在 Science 中 F1 值提高了 0.60%, 在 Music 中 F1 值提高了 2.55%, 在 Litera 中 F1 值提高了 0.91%, 在 AI 中 F1 值提高了 2.61%, F1 平均值提高了 1.47%. 由于 MULTI-CELL-LSTM 仅考虑实体层面的特征信息, 忽略了特征信息中的结构化知识, 而 MSKE-CDNER 能够有效地利用特征信息中的结构化知识, 从而取得更优的性能.

为了更好地检验模型的可靠性, 在不同数据集上采用与 Baziotis 等人<sup>[42]</sup>一样的方式, 重复 3 次实验, 记录实验结果的平均值和标准差, 结果如表 3 和表 4 中 MSKE-CDNER\* 行所示. 对比现有的跨域方法, MSKE-CDNER 在 2 个不同的数据集下整体实验结果优于对比方法, 说明学习、利用特征的结构化知识可

以促进跨域知识的迁移。

4.6 实验分析

在本节中,选取 PC→CG 这组实验从消融实验、显著性检验、参数分析以及细粒度 4 个方面对实验进行分析。PC, CG 数据集属于医疗领域,领域中已标注的实体资源较少,选取这组实验进行分析更能说明 MSKE-CDNER 在数据资源匮乏领域下实体识别的有效性。

4.6.1 消融实验

为了验证多级结构化对齐机制的有效性,在 PC→CG 这组实验中进行消融实验,得到的对比结果如表 5 所示,可以看出机制中 3 个模块对实体识别性能提升均有所帮助, $\Delta$ 代表消融不同方法后  $F1$  值的对比差值。

Table 5 Ablation Study on PC→CG Dataset

表 5 在 PC→CG 数据集上的消融实验

方法	$P$	$R$	$F1$	$\Delta$
MSKE-CDNER (本文)	83.80	82.08	82.93	
$-L_{\text{bio}}$	83.74	81.15	82.56	-0.37
$-L_{\text{mmd}}$	83.79	81.68	82.72	-0.21
$-L_{\text{CDA}}$	83.25	81.87	82.42	-0.51

表 5 中  $-L_{\text{bio}}$  代表消去边界检测模块时得到的结果,  $P$  下降了 0.06%,  $R$  下降了 0.93%,  $F1$  值下降了 0.37%, 其中  $R$  在 3 个消融实验中下降最多, 说明学习共性边界信息能帮助模型正确地识别实体类型, 提升模型的性能, 加入此模块可以改善跨域命名实体识别的效果。同样地,  $-L_{\text{mmd}}$  表示消去潜层对齐模块的实验结果,  $P$  下降了 0.01%,  $R$  下降了 0.4%,  $F1$  值下降了 0.21%, 表明迁移源领域和目标领域中的共性知识可以促进实体识别能力跨领域迁移。 $-L_{\text{CDA}}$  代表消去结构化对齐模块的结果,  $P$  下降了 0.55%,  $R$  下降了 0.21%,  $F1$  值下降了 0.51%, 其中  $F1$  值在 3 个消融实验中下降最多。因为结构化对齐模块在获取特征表示的同时获取到了结构化知识, 在进行迁移时结构化信息因其跨域稳定性可以显著地提升目标领域的实体识别性能。

4.6.2 显著性检验

在 PC→CG 实验中进行显著性检验, 结果如表 6 所示。其中表中  $P$  值为  $\text{Prob}>F$ , 当  $P \leq 0.05$  时, 说明 PC 和 CG 之间有显著性差异。表 6 中  $P = 0.0277$ , 表明 MSKE-CDNER 有显著性差异;  $F$  为检验的统计量;  $P$  为用于检验的  $P$  值。

Table 6 ANOVA of MSKE-CDNER on PC→CG Dataset

表 6 MSKE-CDNER 在 PC→CG 数据集上的方差分析

来源	平方和	自由度	均方差	$F$ 值	$P$ 值
组	1.008 6	1	1.008 6	11.44	0.027 7
误差	0.352 53	4	0.088 13		
合计	1.361 13	5			

4.6.3 参数分析

为探讨结构化表示模块中参数  $\lambda_1, \lambda_2$  对实验的影响, 设置不同的参数值进行多轮试验, PC→CG 这组实验的参数调优如图 5 所示。在结构化表示模块中,  $Loss$  值由 WD 的距离和 GWD 的距离共同组成, 其中  $\lambda_1$  代表 WD 的权重,  $\lambda_2$  代表 GWD 的权重, 我们对其占比进行参数调优, 对于  $\lambda_1$  和  $\lambda_2$ , 取值分别设置为 0.1, 1, 10, 100。从图 5 中可以明显地看出, 当  $\lambda_1$  值固定时, 随着  $\lambda_2$  值的增加, 结构化信息的比值在整个特征表示中的比值也在增加。此时, 可以看出随着结构化信息权重的增加, 模型识别能力也随之提高, 模型性能越来越好, 说明结构化信息可以促进跨域知识的迁移, 在  $\lambda_2=100$  时取得最优结果。经过参数调优, 最终选取  $\lambda_1=0.1, \lambda_2=100$  作为 PC→CG 这组实验中的最优参数。同理, 选取  $\lambda_1=100, \lambda_2=1$  作为 Conll03→PC 这组实验的最优参数。

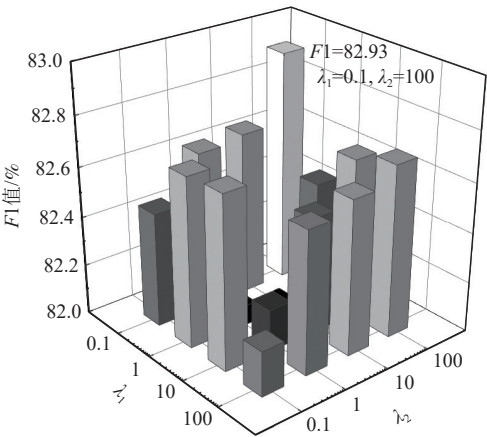


Fig. 5 Parameter analysis in the graph optimal transmission

图 5 图最优传输中的参数分析

4.6.4 细粒度分析

表 7 统计了 MSKE-CDNER 在 PC→CG 这组实验中的细粒度实验结果, 表 7 中记录了相关实体的准确率、召回率与  $F1$  值。由于相关数据集中的实体种类较多, 为便于讨论, 随机选取实体类型进行示例说明。与当前的跨域模型相比所有模型的  $F1$  值均有所提升, 总体  $F1$  值提升了近 0.92%, 证实了 MSKE-CDNER 的

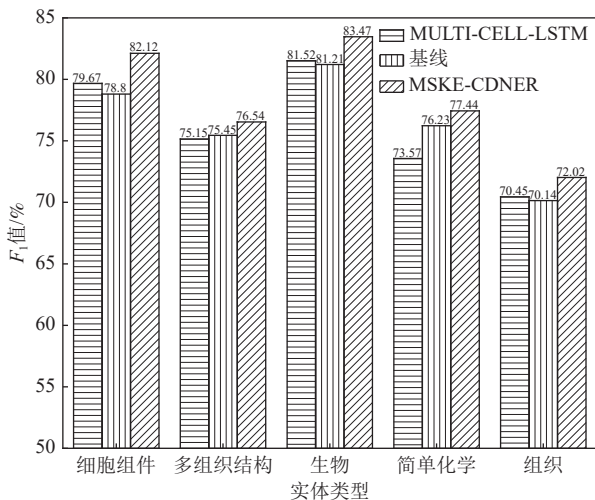
Table 7 Fine-Grained Analysis on PC→CG Dataset

表7 PC→CG数据集上的细粒度分析

实体类型	<i>P</i>	<i>R</i>	<i>F1</i>
细胞组件	82.58	81.67	82.12
多组织结构	77.18	75.91	76.54
生物	87.34	79.92	83.47
简单化学	81.05	74.14	77.44
组织	65.20	80.43	72.02

有效性。

为了能够清晰地对比出 MSKE-CDNER 在实体类型级别取得更优的结果,在不同类型的实体中将其与原模型进行细粒度分析对比,结果如图6所示。在相同的实体类型下,MSKE-CDNER 结果明显优于其他2种方法,这得益于实体内部的结构化知识在进行跨域迁移时有效缓解了不同域中实体歧义性的问题。

Fig. 6 Comparison of fine-grained *F1*-values among models图6 模型间细粒度*F1*值比较

## 5 总 结

本文提出了一种基于多层结构化语义知识增强的跨领域命名实体识别模型 MSKE-CDNER, 设计了多级结构化对齐机制, 利用文本序列的结构化语义信息, 分别将嵌入层获取的特征表示和隐藏层获取的上下文特征表示以结构化的形式对齐, 通过获取、迁移不同层次的结构化知识, 促进模型实体识别能力的跨领域迁移。在5个英文数据集以及 CrossNER 数据集上进行实验, 并与当前跨域方法相比, 结果表明, MSKE-CDNER 在跨域任务中取得了较好的结果,

表明学习和利用结构化知识能够更好地促进跨域知识的迁移。在将来的工作中, 我们会对域不变知识和特定域知识进行更优的解耦, 来获取更优的特征表示。

**作者贡献声明:** 张文韩调研文献、设计实验、撰写和修订论文; 刘小明提出研究思路、模型框架、内容规划、指导意见和修订论文; 杨关负责实验指导、修订论文; 刘杰提出指导意见、审阅和修订论文。

## 参 考 文 献

- [1] Liu Zihan, Xu Yan, Yu Tiezheng, et al. CrossNER: Evaluating cross-domain named entity recognition[C] //Proc of the 21st AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 13452–13460
- [2] Zhang Xuesong, Guo Ruiqiang, Huang Degen. Named entity recognition based on dependency[J]. Journal of Chinese Information Processing, 2021, 35(6): 63–73 (in Chinese)  
(张雪松, 郭瑞强, 黄德根. 基于依存关系的命名实体识别[J]. 中文信息学报, 2021, 35(6): 63–73)
- [3] Deng Yiyi, Wu Changxing, Wei Yongfeng, et al. A survey on named entity recognition based on deep learning[J]. Journal of Chinese Information Processing, 2021, 35(9): 30–45 (in Chinese)  
(邓依依, 邬昌兴, 魏永丰, 等. 基于深度学习的命名实体识别综述[J]. 中文信息学报, 2021, 35(9): 30–45)
- [4] Jia Chen, Liang Xiaobo, Zhang Yue. Cross-domain NER using cross-domain language modeling[C] //Proc of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2019: 2464–2474
- [5] Wang Zhenghui, Qu Yanru, Chen Liheng, et al. Label-Aware double transfer learning for cross-specialty medical named entity recognition[C] //Proc of the 2018 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2018: 1–15
- [6] Jia Chen, Zhang Yue. Multi-cell compositional LSTM for NER domain adaptation[C] //Proc of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2020: 5906–5917
- [7] Bill L Y, Lu Wei. Neural adaptation layers for cross-domain named entity recognition[C] //Proc of the 2018 Conf on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2018: 2012–2022
- [8] Rodriguez J D, Caldwell A, Liu A. Transfer learning for entity recognition of novel classes[C] //Proc of the 27th Int Conf on Computational Linguistics. Stroudsburg, PA: ACL, 2018: 1974–1985
- [9] Titouan V, Courty N, Tavenard R, et al. Optimal transport for structured data with application on graphs[C] //Proc of the 36th Int Conf on Machine Learning. New York: PMLR, 2019: 6275–6284
- [10] Kriege N M, Fey M, Fisseler D, et al. Recognizing cuneiform signs using graph based methods[C] //Proc of the Int Workshop on Cost-Sensitive Learning. New York: PMLR, 2018: 31–44

- [11] Ktena S I, Parisot S, Ferrante E, et al. Distance metric learning using graph convolutional networks: Application to functional brain networks[C] //Proc of Int Conf on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2017: 469–477
- [12] Yanardag P, Vishwanathan S. Deep graph kernels[C] //Proc of the 21st ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2015: 1365–1374
- [13] Yang Qiang, Zheng Vincent W, Li Bin, et al. Transfer learning by reusing structured knowledge[J]. *AI Magazine*, 2011, 32(2): 95–106
- [14] Alvarez-Melis D, Jaakkola T, Jegelka S. Structured optimal transport[C] //Proc of the 21st Int Conf on Artificial Intelligence and Statistics. New York: PMLR, 2018: 1771–1780
- [15] Chen Liqun, Gan Zhe, Cheng Yu, et al. Graph optimal transport for cross-domain alignment[C] //Proc of the 37th Int Conf on Machine Learning. New York: PMLR, 2020: 1542–1553
- [16] Zhang Han, Guo Yuanbo, Li Tao. Domain named entity recognition combining GAN and BiLSTM-Attention-CRF[J]. *Journal of Computer Research and Development*, 2019, 56(9): 1851–1858 (in Chinese)  
(张晗, 郭渊博, 李涛. 结合GAN与BiLSTM-Attention-CRF的领域命名实体识别[J]. *计算机研究与发展*, 2019, 56(9): 1851–1858)
- [17] Li Jing, Shang Shuo, Shao Ling. Metaner: Named entity recognition with meta-learning[C] //Proc of the Web Conf 2020. New York: ACM, 2020: 429–440
- [18] Chen Shuguang, Aguilar G, Neves L. Data augmentation for cross-domain named entity recognition[C] //Proc of the 2021 Conf on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2021: 5346–5356
- [19] Hao Zhifeng, Lv D, Li Zijian, et al. Semi-supervised disentangled framework for transferable named entity recognition[J]. *Neural Networks*, 2021, 135: 127–138
- [20] Nozza D, Manchanda P, Fersini E, et al. Learning to adapt with word embeddings: Domain adaptation of named entity recognition systems[J/OL]. *Information Processing & Management*, 2021[2023-01-16]. <https://www.sciencedirect.com/science/article/abs/pii/S0306457321000455>
- [21] Li Xin, Li Zheming, Wei Juhui, et al. Cross-domain adaptive learning model based on feature separation[J]. *Journal of Computer Research and Development*, 2022, 59(1): 105–117 (in Chinese)  
(李鑫, 李哲民, 魏居辉, 等. 基于特征分离的跨域自适应学习模型[J]. *计算机研究与发展*, 2022, 59(1): 105–117)
- [22] Gu Shuhao, Feng Yang, Liu Qun. Improving domain adaptation translation with domain invariant and specific information[C] //Proc of the 2019 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2019: 3081–3091
- [23] Dong Jianfeng, Long Zhongzi, Mao Xiaofeng, et al. Multi-level alignment network for domain adaptive cross-modal retrieval[J]. *Neurocomputing*, 2021, 440: 207–219
- [24] Zhang Tao, Xia Congying, Yu P S, et al. PDALN: Progressive domain adaptation over a pre-trained model for low-resource cross-domain named entity recognition[C] //Proc of the 2021 Conf on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2021: 5441–5451
- [25] Swarup S, Ray S R. Cross-domain knowledge transfer using structured representations[C] //Proc of the 21st National Conf on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conf. Menlo Park, CA: AAAI, 2006: 506–511
- [26] Lee J, Kim H, Lee J, et al. Transfer learning for deep learning on graph-structured data[C] //Proc of the 31st AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2017: 2154–2160
- [27] Zheng Junhao, Chen Haibin, Ma Qianlin. Cross-domain named entity recognition via graph matching[C] //Proc of Findings of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2022: 2670–2680
- [28] Courty N, Flamary R, Tuia D, et al. Optimal transport for domain adaptation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(9): 1853–1865
- [29] Peyré G, Cuturi M, Solomon J. Gromov-Wasserstein averaging of kernel and distance matrices[C] //Proc of the 33rd Int Conf on Machine Learning. New York: PMLR, 2016: 2664–2672
- [30] Ma X, Hovy E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF[C] //Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2016: 1064–1074
- [31] Wei Zhenkai, Cheng Meng, Zhou Xiabing, et al. Convolutional interactive attention mechanism for aspect extraction[J]. *Journal of Computer Research and Development*, 2020, 57(11): 2456–2466 (in Chinese)  
(尉桢楷, 程梦, 周夏冰, 等. 基于类卷积交互注意力机制的属性抽取研究[J]. *计算机研究与发展*, 2020, 57(11): 2456–2466)
- [32] Sang E F T K, De Meulder F. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition[C] //Proc of the 7th Conf on Natural Language Learning at HLT-NAACL 2003. Stroudsburg, PA: ACL, 2003: 142–147
- [33] Lu Di, Neves L, Carvalho V, et al. Visual attention model for name tagging in multimodal social media[C] //Proc of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2018: 1990–1999
- [34] Derczynski L, Bontcheva K, Roberts I. Broad Twitter corpus: A diverse named entity recognition resource[C] //Proc of the 26th Int Conf on Computational Linguistics: Technical Papers. Stroudsburg, PA: ACL, 2016: 1169–1179
- [35] Nédellec C, Bossy R, Kim J, et al. Overview of BioNLP shared task 2013[C] //Proc of the BioNLP Shared Task 2013 Workshop. Stroudsburg, PA: ACL, 2013: 1–7
- [36] Yang Jie, Zhang Yue. NCRF++: An open-source neural sequence labeling toolkit[C] //Proc of Association for Computational Linguistics 2018 System Demonstrations. Stroudsburg, PA: ACL, 2018: 74–79
- [37] Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C] //Proc of the 2014 Conf on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2014: 1532–1543
- [38] Chiu B, Crichton G, Korhonen A, et al. How to train good word embeddings for biomedical NLP[C] //Proc of the 15th Workshop on



- Biomedical Natural Language. Stroudsburg, PA: ACL, 2016: 166–174
- [39] Lample G, Ballesteros M, Subramanian S, et al. Neural architectures for named entity recognition[C] //Proc of the 2016 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2016: 260–270
- [40] Liu Zihan, Winata G I, Xu Peng, et al. Coach: A coarse-to-fine approach for cross-domain slot filling[C] //Proc of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2020: 19–25
- [41] Zhou Tianyi, Zhang Hao, Jin Di, et al. Dual adversarial neural transfer for low-resource named entity recognition[C] //Proc of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2019: 3461–3471
- [42] Baziotis C, Haddow B, Birch A. Language model prior for low-resource neural machine translation[C] //Proc of the 2020 Conf on Empirical Methods in Natural Language Processing(EMNLP). Stroudsburg, PA: ACL, 2020: 7622–7634



**Zhang Wenhan**, born in 1999. Master candidate. His main research interests include natural language processing and machine learning.

张文韩, 1999年生. 硕士研究生. 主要研究方向为自然语言处理和机器学习.



**Liu Xiaoming**, born in 1979. PhD, lecturer, master supervisor. Member of CCF. His main research interests include natural language processing, Chinese information processing, and machine learning.

刘小明, 1979年生. 博士, 讲师, 硕士生导师. CCF 会员. 主要研究方向为自然语言处理、中文信息处理、机器学习.



**Yang Guan**, born in 1974. PhD, associate professor, master supervisor. His main research interests include image processing and machine learning.

杨 关, 1974年生. 博士, 副教授, 硕士生导师. 主要研究方向为图像处理、机器学习.



**Liu Jie**, born in 1973. PhD, professor, PhD supervisor. His main research interests include natural language processing, knowledge engineering, and knowledge graph.

刘 杰, 1973年生. 博士, 教授, 博士生导师. 主要研究方向为自然语言处理、知识工程、知识图谱.