

基于细粒度原型网络的小样本命名实体识别方法^{*}

戚荣志¹, 周俊宇¹, 李水艳², 毛莺池¹

¹(河海大学 计算机与信息学院, 江苏 南京 211106)

²(河海大学 理学院, 江苏 南京 211106)

通信作者: 李水艳, E-mail: lsy@hhu.edu.cn



摘要: 原型网络直接应用于小样本命名实体识别 (few-shot named entity recognition, FEW-NER) 时存在以下问题: 非实体之间不具有较强的语义关系, 对实体和非实体都采用相同的方式构造原型将会造成非实体原型不能准确表示非实体的语义特征; 仅使用平均实体向量表示作为原型的计算方式将难以捕捉语义特征相差较大的同类实体. 针对上述问题, 提出基于细粒度原型网络的小样本命名实体识别 (FEW-NER based on fine-grained prototypical networks, FNFP) 方法, 有助于提高小样本命名实体识别的标注效果. 首先, 为不同的查询集样本构造不同的非实体原型, 捕捉句子中关键的非实体语义特征, 得到更为细粒度的原型, 提升模型对非实体的识别效果; 然后, 设计一个不一致性度量模块以衡量同类实体之间的一致性, 对实体与非实体采用不同的度量函数, 从而减小同类样本之间的特征表示, 提升原型的特征表示能力; 最后, 引入维特比解码器捕捉标签转换关系, 优化最终的标注序列. 实验结果表明, 采用基于细粒度原型网络的小样本命名实体识别方法, 在大规模小样本命名实体识别数据集 FEW-NERD 上, 较基线方法获得提升; 同时在跨领域数据集上验证所提方法在不同领域场景下的泛化能力.

关键词: 小样本命名实体识别; 细粒度原型网络; 小样本学习; 特征表示

中图法分类号: TP18

中文引用格式: 戚荣志, 周俊宇, 李水艳, 毛莺池. 基于细粒度原型网络的小样本命名实体识别方法. 软件学报, 2024, 35(10): 4751-4765. <http://www.jos.org.cn/1000-9825/6979.htm>

英文引用格式: Qi RZ, Zhou JY, Li SY, Mao YC. Few-shot Named Entity Recognition Based on Fine-grained Prototypical Network. Ruan Jian Xue Bao/Journal of Software, 2024, 35(10): 4751-4765 (in Chinese). <http://www.jos.org.cn/1000-9825/6979.htm>

Few-shot Named Entity Recognition Based on Fine-grained Prototypical Network

QI Rong-Zhi¹, ZHOU Jun-Yu¹, LI Shui-Yan², MAO Ying-Chi¹

¹(College of Computer and Information, Hohai University, Nanjing 211106, China)

²(College of Science, Hohai University, Nanjing 211106, China)

Abstract: When prototypical networks are directly applied to few-shot named entity recognition (FEW-NER), there are the following problems: Non-entities do not have strong semantic relationships with each other, and using the same way to construct the prototype for both entities and non-entities will make non-entity prototypes fail to accurately represent the semantic characteristics of non-entities; using only the average entity vector as the computing method of the prototype will make it difficult to capture similar entities with different semantic features. To address these problems, this study proposes a FEW-NER based on fine-grained prototypical networks (FNFP) to improve the annotation effect of FEW-NER. Firstly, different non-entity prototypes are constructed for different query sets to capture the key semantic features of non-entities in sentences and obtain finer-grained prototypes to improve the recognition effect of non-entities. Then, an inconsistent metric module is designed to measure the inconsistency between similar entities, and different metric functions are applied to entities and non-entities, so as to reduce the feature representation between similar samples and improve the feature

* 基金项目: 国家重点研发计划 (2022YFC3005401)

收稿时间: 2023-01-16; 修改时间: 2023-04-02; 采用时间: 2023-06-03; jos 在线出版时间: 2023-09-06

CNKI 网络首发时间: 2023-09-08

representation of the prototype. Finally, a Viterbi decoder is introduced to capture the label transformation relationship and optimize the final annotation sequence. The experimental results show that the performance of the proposed method is improved compared with that of the large-scale FEW-NER dataset, namely FEW-NERD; and the generalization ability of this method in different domain scenarios is verified on the cross-domain dataset.

Key words: few-shot named entity recognition (FEW-NER); fine-grained prototypical network; few-shot learning; feature representation

命名实体识别 (named entity recognition, NER) 是信息抽取中一项关键任务,旨在从预定义类型的文本中识别出命名实体,例如人名、地名和组织名等^[1]. 作为一项核心的语言理解任务,广泛应用于新闻、医疗和社交等领域. 早期的命名实体识别任务主要采用基于规则和字典的方法^[2,3], 该类方法具有显著的识别效果,但需要消耗巨大的人力以制定识别规则,且具有较小的通用性,仅适合在特定领域中使用. 在传统机器学习兴起后,命名实体识别被表示为序列标注任务. 在标注的过程中,考虑上下文标签转化分布以输出最优的标注序列,主要方法包括:支持向量机 (SVM)^[4]、隐马尔可夫模型 (HMM)^[5]和条件随机场 (CRF)^[6]. 随着深度神经网络的不断发展,端到端训练方式的神经网络方法具有推断潜在特征与学习标签序列的优势,大量采用神经网络的方法在完全监督的命名实体识别任务^[7-9]中取得突破性进展,其分类效果依赖于大量已标注的样本. 目前,较多大规模标注数据集可以高质量地解决命名实体识别任务. 然而,在现实的特定领域,获取领域中大量标注样本不太可能,即使对原始样本进行大规模标注,也不可避免某些类别只包含少量样本. 因此,低资源环境下如何有效地进行命名实体识别引起了广泛关注. 小样本学习能够有效解决因样本不足而导致模型最终性能不佳的问题,该方法借鉴人类依靠少量数据就能做到快速学习的特点,使得深度学习能够更加接近人类的思维,因此,大量小样本命名实体识别的方法^[10-13]得到快速发展. 小样本命名实体识别的例子如图 1 所示,在见过支持集中少数句子后,对查询集的句子进行标注,通过发现查询集中单词“president”和支持集中单词“minister”语义相似,于是将支持集样本的类型分配给与之相似的查询集样本.

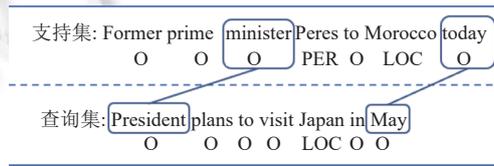


图 1 一个小样本命名实体识别的例子

原型网络是小样本命名实体识别的主要方法^[13-15],然而,小样本命名实体识别相对于其他小样本学习任务具有独特的挑战,现有的方法直接将原型网络应用于命名实体识别存在着一定问题,不能够发挥出其最大性能. 命名实体识别作为一项序列标注任务,实体的类别与上下文实体具有一定的关系,对标签依赖关系进行建模至关重要. 大部分非实体之间不具有较强的语义关系,训练集中被标记为非实体的单词,在测试集中可能表示为具体的实体,原型网络通过平均向量表示的方式构造实体原型与非实体原型,不能准确地表示非实体的语义特征;当支持集中同类实体的表示彼此相距甚远时,所得原型将很难捕捉到实体的共同特征.

针对上述问题,本文提出了一种基于细粒度原型网络的小样本命名实体识别方法,可以将给定的样本特征有效地映射到原型空间中. 首先,细粒度原型网络对实体与非实体采用不同的方式构造原型,为不同查询集样本构造不同的非实体原型,解决因非实体之间不具有较强的语义关系导致所构非实体原型不准确的问题,提升模型对非实体类型的识别效果;其次,不一致度量模块衡量支持集同类实体特征表示的不一致性,进而通过反向传播更新参数以使得同类型实体的空间表示能够相互靠近,有益于实体原型更好地表达实体特征;最后,引入维特比解码器捕捉标签依赖关系,结合上下文标签转换分布以优化最终的标注序列. 本文在最近提出的大规模小样本命名实体识别数据集上进行充分的实验验证,实验结果表明,所提方法几乎在所有小样本基准设置中取得了最先进的结果,说明所提方法在小样本命名实体识别中的有效性.

本文工作的主要贡献如下.

(1) 针对小样本命名实体识别的特点提出细粒度原型网络,充分考虑实体与非实体的差异性,对实体与非实体

采用不同的原型构造方式.

(2) 设计不一致性度量模块衡量同类实体的不一致性, 并根据实体与非实体的特点对其采用不同的度量函数, 使得构造的原型能够准确地表达样本的语义特征.

(3) 考虑命名实体识别这一序列标注任务中上下文所具有关系的特点, 在细粒度原型网络的输出端引入维特比解码器, 捕捉标签转换关系, 对模型分类结果进行优化以得到最优的标注序列.

本文第 1 节介绍小样本命名实体识别的相关工作. 第 2 节介绍所需要的基础知识, 包括小样本学习和原型网络. 第 3 节介绍本文提出的基于细粒度原型网络的小样本命名实体识别方法. 第 4 节通过实验与分析, 验证本文所提方法的有效性. 第 5 节对全文进行总结.

1 相关工作

现有命名实体识别模型大多数都是基于深度神经网络的完全监督方法, 其标注效果取得了突破性的进展, 这些方法依赖于大规模已标注的数据集. 然而, 对于现实特定环境而言, 获取大量标注数据不太可能. 为解决低资源环境下从少数样本中学习新知识以实现序列标注任务, 命名实体识别任务开始应用小样本学习 (few-shot learning) 的方法. 小样本学习旨在依靠每个类别的少量标注样本, 从而识别看不见的新类. 这种方法甚至可用于零样本学习 (zero-shot learning), 目标类可以在训练阶段中不可见, 因而, 模型不需要重新训练就能识别新类别.

早期 Bengio 等人^[16]的小样本学习研究主要集中在应用迁移学习方法来微调预训练模型, 这些模型从包含大量样本的公共类中提取潜在信息. Koch 等人^[17]在小样本学习的基础上提出度量学习, 通过计算样本之间的空间距离以衡量其相似度. Snell 等人^[18]提出原型网络应用于图像分类任务并取得先进的效果, 原型网络将查询集和支持集样本独立地编码到向量空间中, 原型代表支持集样本向量表示的平均值, 从而根据查询集样本向量表示与原型之间的距离进行分类. 随后原型网络也引入自然语言处理领域, Sun 等人^[19]将其成功应用于文本分类, Han 等人^[20]提出小样本关系抽取数据集并使用原型网络进行分类.

目前, 越来越多的研究者开始致力于小样本命名实体识别并为之做出了贡献. Hofer 等人^[10]将小样本学习的思想应用于命名实体识别, 使用预训练模型进行初始化以摆脱对大量标注样本的依赖. Fritzier 等人^[13]将小样本命名实体识别定义为半监督学习任务, 通过修改原型网络使其适用于命名实体识别任务. Yang 等人^[11]提出基于结构化最邻近学习方法, 通过计算实体之间的向量表示距离以实现分类任务, 并使用维特比解码器捕捉标签之间的依赖关系. Tong 等人^[21]提出从非实体中学习新实体类的方法, 利用已有的语义知识增强对非实体类的理解, 从非实体中识别出丰富的未定义实体类型, 取得了不错的效果. Ding 等人^[15]设计一个大规模小样本命名实体识别数据集, 同时构建不同侧重点的小样本基准任务, 并使用原型网络与前人的方法进行充分比较, 以说明原型网络在命名实体识别任务中的有效性. Chen 等人^[22]提出基于自描述网络的小样本命名实体识别方法, 有效地利用说明性样本, 并通过使用通用概念集描述实体类型和提及, 从而精确地传递外部资源的知识. 现有的小样本命名实体识别能够很好地应用小样本学习的思想, 然而, 这些方法并没有针对实体与非实体采用不同建模方式, 因此, 对于非实体的标注效果并不是很好. 针对这些问题, 本文提出了细粒度原型网络, 充分考虑非实体的语义特征, 对实体与非实体采用不同的原型建模方式, 从而使得模型对非实体的标注效果有所提升.

2 基础知识

命名实体识别通常被表示为序列标记任务, 对于形如 $x = \{x_1, x_2, \dots, x_T\}$ 的句子输入序列, 命名实体识别旨在为每个字符 x_i 分配标签 $y_i \in Y$, 以表示每个字符所属的类型. 同大多数自然语言处理任务一样, 命名实体识别任务中主流的方法采用有监督神经网络进行分类, 然而, 这种方式的分类效果依赖于大规模标注的训练数据, 并且容易在小数据集中出现过拟合现象. 在现实世界的很多应用场景中, 获取大量标注的数据是相对困难的, 因此, 如何在低资源环境下进行命名实体识别引起广泛的研究, 小样本学习因而被引入到命名实体识别任务中.

2.1 小样本学习

小样本命名实体识别的目标是在标注样本不足的情况下, 通过学习少数样本的特征以对新的类别进行标注.

目前,元学习是小样本学习中使用较多的方法,该方法旨在训练出一种通用的模型,在使用较少训练样本的情况下能够快速适应新领域.基于元学习的方法通常采用 episodic 的训练方式^[23],将数据集划分为多个任务进行训练,每个任务从数据集的 N 个类别中随机采样 K 个样本,即 N 类 K 样本 (N -way K -shot) 的采样方式^[24].给定目标数据集 $T = \{D_{\text{train}}, D_{\text{test}}\}$,训练集 D_{train} 包含支持集 $S_{\text{train}} = \{x_i^s, y_i^s\}_{i=1}^{N_s}$ 和查询集 $Q_{\text{train}} = \{x_j^q, y_j^q\}_{j=1}^{N_q}$,每个集合都由样本 x_j 与标注序列 y_j 组成,且 $S \cap Q = \emptyset$,其中 $N_s = N \times K$, $N_q = N \times K'$,通常情况下 K 与 K' 的取值相等,以便模型在测试过程中共享训练好的参数.在训练阶段,支持集供模型学习样本的语义特征,而查询集用于验证模型性能并做出优化.测试集 D_{test} 由支持集 $S_{\text{test}} = \{x_i^s, y_i^s\}_{i=1}^{N_s}$ 与查询集 $Q_{\text{test}} = \{x_j^q, y_j^q\}_{j=1}^{N_q}$ 组成.在测试阶段,所有样本都是训练阶段中从未见过的,该阶段的查询集 Q_{test} 并不包含样本标签,通过使用较少带有标注序列的支持集样本 S_{test} ,模型需对查询集 Q_{test} 的样本进行标注.然而,命名实体识是一个序列标注问题,输入的句子序列可能包含多个类别的实体,实体之间存在密切的关联.因此,相比自然语言处理中其他小样本学习任务,小样本命名实体识别的采样与分类过程更加困难.为了使采样过程更加随机,Ding 等人^[15]提出 N 类 K - $2K$ 样本的采样方式,每个类别采样 K - $2K$ 个样本,有效缓解因实体数存在差异导致采样不随机的问题.

2.2 原型网络

原型网络最初为图像分类任务所提出,能够在标注样本稀缺的环境中实现分类任务,取得了良好的效果.近年来,原型网络开始应用于信息抽取任务,并在关系抽取与文本分类任务中取得了不错的效果.然而,对于命名实体识别任务而言,其样本包含大量无语义关系的非实体,导致所构造的非实体原型并不准确,因此原型网络为其所带的效果并不显著.原型网络首先将样本对象映射到向量空间,根据空间中样本之间的距离,将样本分离成特定任务所需要集群,从而能够在低资源环境中将输入样本分配到新的类别.对于原型网络而言,首先,支持集的标记样本用于构造每类原型,所得原型表示支持集中每类样本的共同特征,因而查询集能够根据每类原型对看不见的样本分配类别.然后,将支持集样本对象映射到 M 维的向量空间中,每类原型采用平均同类样本表示进行构造,第 K 类原型使用 C_k 表示,代表该类中 K 个样本对象在向量空间中的平均值.在为支持集构造好每类原型后,从查询集中选取样本 X 的向量表示,计算其与所有原型之间的距离,选取与其距离最近的原型.最后,将该原型的类别分配给样本 X .样本对象与原型之间的空间距离通常采用平方欧几里得函数计算,并通过 *Softmax* 函数将空间距离转化为相似度的概率分布,最终使用交叉熵损失函数进行优化,通过不断地训练使得模型能够快速收敛.

3 小样本命名实体识别模型

针对原型网络应用于小样本实体识别存在所构原型不能准确表示同类样本特征的问题,本文提出一种基于细粒度原型网络的小样本命名实体识别 (FNFP) 方法,主要由细粒度原型网络 (FG-Proto) 和不一致性度量模块 (Incon) 两部分构成.细粒度原型网络对实体与非实体采用不同的方式构造原型,对不同查询集样本构造不同的非实体原型,解决因非实体之间不具有较强的语义关系导致所构造的非实体原型不准确的问题,提升模型对非实体类型的识别效果.不一致性度量模块则衡量支持集同类实体特征表示的不一致性,进而通过反向传播更新参数以使得同类型实体的空间表示能够相互靠近,有益于实体原型更好地表达实体特征.FNFP 模型框架如图 2 所示.

FNFP 模型包括嵌入层、细粒度原型网络、不一致性度量模块、维特比解码器与输出层.首先,对于给定的支持集与查询集句子序列,嵌入层使用预训练语言模型对句子中的字符进行上下文编码,得到所有实体与非实体的向量表示;细粒度原型网络为支持集中的实体和非实体构造原型,每类原型代表该类标签的共同特征,通过距离函数计算查询样本与所有原型之间的距离,以表示该样本与每类标签之间的相似度,最后使用 *Softmax* 函数将相似度转化为概率输出,并通过交叉熵函数计算分类损失;不一致性度量模块衡量支持集同类实体之间的不一致性,同时结合细粒度原型网络输出的分类损失,通过反向传播更新嵌入层参数,以减少同类实体之间的特征表示,进一步使得细粒度原型网络所构造的原型更加准确.维特比解码器学习标签之间的转移分布,结合细粒度原型网络的概率输出以优化标签序列的最终分布;输出层使用维特比解码器输出概率最高的标签序列.

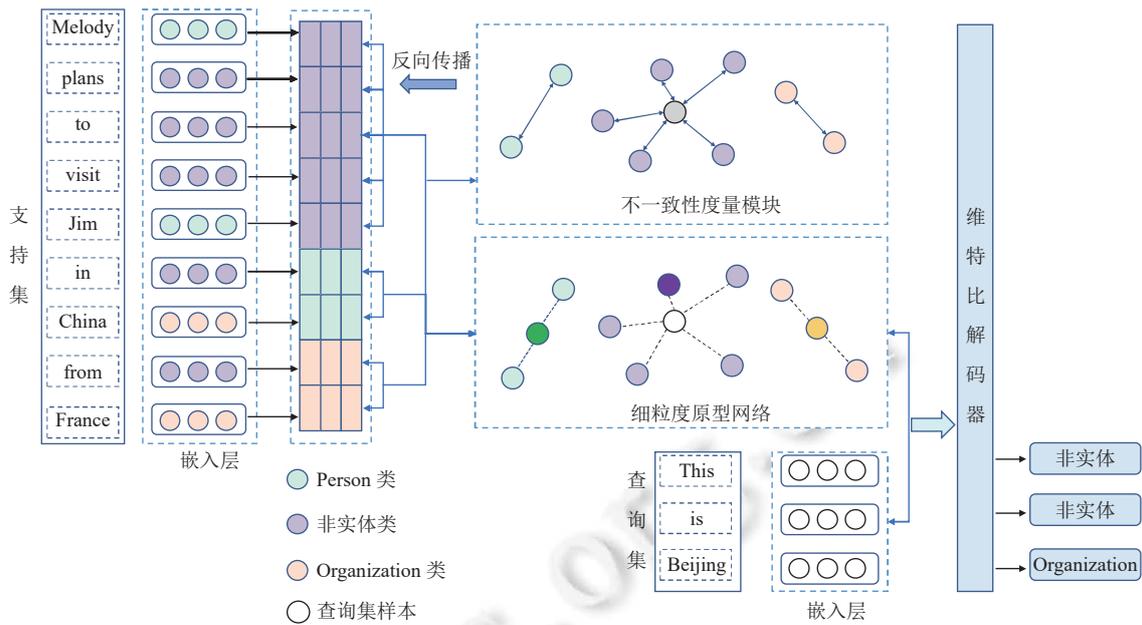


图2 FNFP 模型框架

3.1 嵌入层

嵌入层将输入句子序列的每一字符编码为向量表示, 编码过程考虑到字符上下文的语义关系, 符合命名实体识别这一序列标注任务的实际场景. 研究表明, 大规模预训练语言模型已成为命名实体识别的强大编码器, 使用深度神经网络在大规模的语料库上进行训练, 能够较好地学习到自然语言文本的语义关系. 这些语言模型的参数较大, 需花费大量的时间训练, 通常在自然语言任务中, 使用训练好的语言模型在目标领域进行微调以快速适应.

目前命名实体识别中存在两种流行的有监督预训练语言模型: LSTM 语言模型^[25]与 BERT 语言模型^[26]. 为了与前人的实验进行对比, 本文遵循其嵌入层的设置, 选取 BERT 语言模型作为主干编码器. 将语言模型的参数设为 θ , 使用 f_θ 表示编码器, 对于给定的句子序列 $x = \{x_1, x_2, \dots, x_n\}$, 编码器产生如下的上下文向量表示:

$$h = [h_1, \dots, h_n] = f_\theta([x_1, \dots, x_n]) \quad (1)$$

其中, x_i 表示句子中第 i 个单词, 通过编码器将其转化为具有语义关系的向量表示 h_i , 最终将向量表示的句子序列输入到后续模型中.

3.2 细粒度原型网络

原型网络最初应用于图像分类任务, 越来越多的研究者将其引入到小样本命名实体识别任务. 然而, 现有的研究^[8]仅将其直接应用于命名实体识别, 并没有考虑到命名实体识别任务具有大量非实体的问题. 非实体之间不存在较强的语义关系, 通过平均向量表示的方式所构造的原型, 并不能表示该类语义特征, 从而导致整体的分类效果并不理想.

为解决上述问题, 本节提出细粒度原型网络, 对实体和非实体采用不同的方式构造类原型, 使得非实体原型能够准确表达该类特征. 首先, 对实体与非实体分别构造原型, 从而计算查询样本与所有原型的距离作为相似度, 再选取与该样本最相似的原型, 最终将该原型的类别作为样本的输出标签. 细粒度原型网络的构造流程如图 3 所示.

下面将通过例子来描述细粒度原型网络对样本的标注过程, 给定查询集句子“Melody plans to visit Jim in China from France, they are going to a park and museum”, 其中“Melody”和“Jim”属于人名类别, “China”和“France”属于组织名类别, “park”和“museum”属于地点类别, 其余的单词属于非实体类别, 现需对查询集句子“This is Beijing”中的单词样本“this”进行分类.

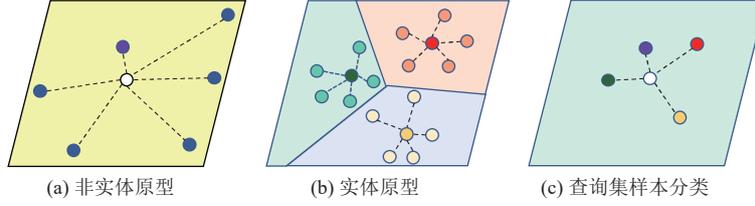


图 3 细粒度原型网络

1) 计算单词样本“this”与支持集中所有非实体之间的距离, 选取与该样本距离最小的一个非实体, 得到的非实体为单词“they”, 将其向量表示作为非实体原型. 如图 3(a) 所示, 白色表示查询集样本, 蓝色表示所有的非实体, 紫色则表示最终得到的非实体原型. 对于不同的查询集样本, 所得非实体原型可能不同. 由于数据集的分类之间存在差异, 所用数据集的非实体中可能存在更细粒度的实体类型, 查询集样本可能与非实体中的某个样本同处一个细粒度类型. 上述例子中, 样本“this”与支持集的“they”都具有指代的含义, 存在较强的语义关系, 而与其他的非实体的语义关系并不大. 因此, 该方法所构建的非实体原型更加准确.

2) 对相同类别的实体采用平均向量表示的方式构建实体原型, 所得原型代表该类实体向量表示的中心点. 如图 3(b) 所示, 每种颜色代表一类实体, 每类实体的中心是对于实体类别的原型. 因同类实体在向量空间中具有聚集的特点, 同类实体向量表示的中心都能够较好划分实体的类别.

3) 结合实体原型与非实体原型, 计算查询样本与所有原型之间的距离, 得到与查询样本距离最近的原型, 使用该原型的类别标注查询样本, 如图 3(c) 所示, 紫色代表非实体原型, 其余颜色代表 3 个类别的实体原型, 查询集单词样本“this”与非实体原型的距离最近, 因此将该样本标注为非实体类.

对于给定的支持集句子序列, 由于句子间所具有的实体类型与个数都大不相同, 且大量的非实体存在于每个句子中. 为方便后续原型的构造, 首先提取出支持集样本中的同类实体:

$$S_i = \{(x_j, y_j) \in S \mid y_j = T_i\} \quad (2)$$

其中, T_i 表示为数据集 T 中第 i 个类别, (x_j, y_j) 属于支持集 S 中的一个实体与标签, 得到第 i 个类别的实体集合 S_i , 同样再提取支持集样本中的非实体集合:

$$S_o = \{(x_j, y_j) \in S \mid y_o = T_o\} \quad (3)$$

其中, y_o 代表非实体这一类别.

在得到支持集样本的非实体集合 S_o 后, 由于非实体之间不具有较强的语义关系, 支持集中的非实体字符在查询集中可能表示为一个实体, 查询集样本也可能与支持集中某个非实体同处一个细粒度实体类型. 因此, 并不是所有支持集的非实体都有益于非实体原型的构造, 对于不同的查询集样本, 计算该样本与支持集中所有非实体之间的距离, 选取与查询集样本距离最小的一个非实体, 将其向量表示作为非实体原型:

$$z_o = \{\hat{x} \mid d(\hat{x}, x_j) = \operatorname{argmin}_{\hat{x} \in S_o} d(\hat{x}, x_j)\} \quad (4)$$

$$d(\hat{x}, x_j) = \|f_{\theta}(\hat{x}) - f_{\theta}(x_j)\|_2^2 \quad (5)$$

其中, \hat{x} 代表支持集中的一个非实体, $d(\hat{x}, x_j)$ 表示两个向量之间的空间距离, 采用平方欧氏距离 d 作为向量空间相似度的距离度量, 并在计算距离之前对向量表示采用 L2 标准化. 对于实体原型的构造, 将支持集中同类实体向量表示的平均值作为每类的原型:

$$z_i = \frac{1}{|S_i|} \sum_{x \in S_i} f_{\theta}(x_j) \quad (6)$$

其中, $|S_i|$ 表示支持集中第 i 类实体的个数, 将实体原型与非实体原型进行融合, 得到支持集中所有原型的表示:

$$z_{\text{all}} = \{z_i, z_o\} \quad (7)$$

最终 z_{all} 代表支持集的所有原型, 接着计算查询集样本 x_j 和 y 类原型 z_y 在向量空间中的相似度距离, 得到与其

距离最小的原型, 使用该原型的类别标注查询集样本:

$$d_y(x_j) = d(f_\theta(x_j), z_y) \tag{8}$$

$$y' = \arg \min_{y \in Y} d_y(x_j) \tag{9}$$

为得到模型的损失, 进而通过反向传播更新参数, 还需将查询集样本 x_j 与每类原型之间的距离转化为概率输出:

$$p(y = c | x_j) = \frac{e^{-d_c(x_j)}}{\sum_{c' \in Y} e^{-d_{c'}(x_j)}} \tag{10}$$

其中, Y 表示所有标注的类别, 包括非实体类型, 模型最终通过交叉熵损失函数进行优化:

$$L(\theta_1) = - \sum_{i=1}^N y_j \log \hat{y}_i \tag{11}$$

其中, \hat{y}_i 为查询集样本实际属于第 i 个类别的概率, \hat{y}_i 表示预测该样本为第 i 个类别的概率.

3.3 不一致性度量模块

细粒度原型网络能较好地改善模型对非实体的识别效果, 而如果支持集中某个实体类的样本表示彼此相距很远, 那么, 所得原型将难以捕捉到样本的共同特征. 因此, 本节设计了一个不一致性度量模块以衡量支持集同类样本之间的一致性, 使得同类样本在空间表示中相互靠近, 进而构造的实体原型能够准确表达该类特征. 由于实体与非实体具有较大区别, 因此, 为实体与非实体设置不同的度量方法, 图 4 刻画了不一致性度量的过程. 实体的一致性使用实体之间的平方欧氏距离作为度量, 如图 4(a) 所示, 每种颜色代表一类实体, 同类实体间的距离作为该类别的不一致性. 相对于实体而言, 非实体之间不具有较强的语义关系, 为此先采用平均非实体向量表示的方式计算出非实体的中心表示, 然后将所有非实体与该中心表示之间的距离作为非实体的不一致性, 如图 4(b) 所示, 紫色代表所有的非实体, 中间的灰色代表非实体的中心表示. 不一致性传入损失函数, 通过反向传播使得同类实体相互靠近, 最终使得细粒度原型网络所构原型更加准确.

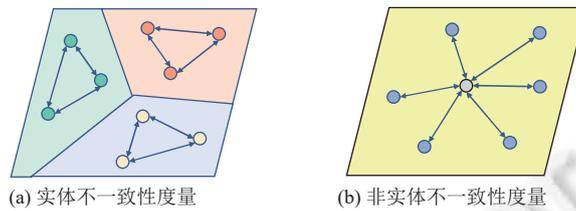


图 4 不一致性度量模块

对于实体的类型, 不一致性度量函数设计为支持集同类样本之间的平均平方欧氏距离, 该值越小表明同类样本的语义特征相对较近, 对于准确地构造原型更有益. 该函数如公式 (12) 和公式 (13) 所示:

$$L(\theta_2) = \frac{1}{\lambda_1} \sum_{n=1}^N \sum_{i=1}^K \sum_{j=1}^K d_i(x_i^n, x_j^n) \tag{12}$$

$$d_i = \text{sum}(\|x_i^n - x_j^n\|_2^2) \tag{13}$$

其中, d_i 表示先计算两个实体表示的欧氏距离, 再将所得结果的所有维度进行求和, λ_1 是一个超参数, 在经过第 4.2.1 节的调参实验后, 将其值取为:

$$\lambda_1 = (N \times K \times (K - 1) \times em) / 2 \tag{14}$$

其中, em 表示词嵌入的维度大小.

为了与前人的研究进行对比, 本文词嵌入大小选取为 768. 通过词嵌入乘于一个系数, 该系数大小为同类样本两两比较的总次数. 综上所述, 最终实体的不一致性函数表示为: 支持集同类实体间词嵌入的单个维度的平均距离.

$$y^* = \operatorname{argmax}_y \prod_{t=1}^T p(y_t | x_j) \times p(y_t | y_{t-1}) \quad (20)$$

发射概率 $p(y_t | x_j)$ 是通过模型输出的, 转移概率 $p(y_t | y_{t-1})$ 是维特比解码器的标签转化概率, 因此, 发射概率和转移概率是独立估计的, 为此引入了超参数 τ 来对转移概率进行归一化, 使得发射概率与转移概率的得分尺度保持一致。

4 实验分析

4.1 实验准备

4.1.1 数据集和评价指标

本文在大规模小样本命名实体识别数据集 FEW-NERD^[15]和跨领域数据集^[27]中进行实验验证。其中, FEW-NERD 是专门为小样本命名实体识别这一任务所设计, 同时也是目前最大的人工标注命名实体识别数据集, 由维基百科中 188 238 个句子组成, 包括 460 160 个单词, 具有 8 个粗粒度实体类型和 66 个细粒度实体类型, 每个粗粒度实体类型包括多个细粒度实体类型, 同一个粗粒度实体类型下的细粒度实体类型存在特定的语义关系。该数据集包括两种不同的基准: Intra 基准根据粗粒度类型构造训练集、验证集和测试集, 不同集合中的细粒度实体不会处于相同的粗粒度类型下, 在这种情况下, 训练集、验证集和测试集之间共享的知识较少, 使得在该基准下进行小样本命名实体识别较为困难; Inter 基准根据细粒度类型构造训练集、验证集和测试集, 所有的细粒度实体类型在训练集、验证集和测试集中互不相交, 共享粗粒度实体类型, 为训练集分配 8 种粗粒度实体类型下 60% 的细粒度类型, 为验证集和测试集都分别分配 20% 的细粒度实体类型, 该基准为探索粗粒度实体特征是否会影响到新类型实体的预测。为了与 Ding 等人^[15]的研究工作进行对比, 本文遵循 N 类 $K-2K$ 样本的采样标准, 并使用其对 FEW-NERD 已采样好的小样本基准数据集进行实验, 以进行较好的对比实验与分析。

Hou 等人^[27]使用 1-shot 和 5-shot 的设置将 News^[28]、Wiki^[29]、Social^[30]、Mixed^[31]等 4 个领域的数据集进行采样以适应小样本命名实体识别任务, 构建了跨领域数据集, 表 1 为采样后的跨领域数据集信息。实验选取 2 个领域作为训练集, 1 个领域作为验证集, 1 个领域作为测试集。

表 1 跨领域数据集信息

| 领域 | 1-shot | | 5-shot | |
|--------|----------|-------|----------|-------|
| | 支持集平均句子数 | 采样样本数 | 支持集平均句子数 | 采样样本数 |
| News | 3.38 | 4 000 | 15.58 | 1 000 |
| Wiki | 6.50 | 4 000 | 27.81 | 1 000 |
| Social | 5.48 | 4 000 | 28.66 | 1 000 |
| Mixed | 14.38 | 2 000 | 62.28 | 1 000 |

精准率 (P)、召回率 (R) 和 $F1$ -score ($F1$) 是通用的评价指标, 由于 $F1$ 值综合了精准率和召回率两个指标, 其值能够综合反应模型整体的性能, 因此, 本文评估小样本实体识别的性能主要以 $F1$ 值作为参照。

4.1.2 基线模型

对于在 FEW-NERD 数据集上的实验, 本文将所提 FNFP 模型与 NNshot 模型^[11]、StructShot 模型^[11]、Proto 模型^[15]、BCL 模型^[32]和 Container 模型^[33]进行比较。

对于在跨领域数据集上的实验, 本文将所提 FNFP 模型与 Bi-LSTM 模型^[27]、TransferBERT 模型^[27]、SimBERT 模型^[27]、Matching Network 模型^[34]、WPZ+BERT 模型^[13]、L-TapNet+CDT 模型^[27]进行比较。

4.1.3 实验设置

对于在 FEW-NERD 数据集下的实验, 选取 5 类 1-2 样本、5 类 5-10 样本、10 类 1-2 样本和 10 类 5-10 样本作为小样本学习中 N 类 K 样本的设置。本文所提超参数 λ_1 与 λ_2 的取值在第 4.2.1 节进行调整, 其余超参数遵循前人^[11,15]的设置, 具体取值如表 2 所示。BERT 隐含层大小设置为 768, 用于生成 768 维的词嵌入向量, 模型训练、验

证、测试的迭代次数分别设置为 10 000、500 和 5 000, 且每训练 1 000 次进行一次验证, 保存验证效果最佳的模型参数供后续测试使用, 采用 Adam 优化器以 0.000 1 的学习率优化模型, 对于模型训练批次、每个句子的最大长度和维特比解码器中的参数 τ , 在 5 类 1-2 样本、5 类 5-10 样本、10 类 1-2 样本和 10 类 5-10 样本的设置中取值各不相同. 对于跨领域数据集下的实验, 模型训练批次设置为 1, 迭代次数设置为 10 次, 学习率取为 $1E-5$. 维特比解码器中的参数 τ 设置为 0.32.

表 2 FEW-NERD 数据集下实验的超参数设置

| 参数 | 参数值 | 参数 | 参数值 |
|--------|---------|------------|-----------------------|
| 学习率 | 0.000 1 | 测试迭代次数 | 5 000 |
| 优化器 | Adam | 验证间隔次数 | 1 000 |
| 词嵌入维度 | 768 | 句子最大长度 | 64/32/64/32 |
| 训练迭代次数 | 10 000 | Batch size | 8/1/4/1 |
| 验证迭代次数 | 500 | τ | 0.32/0.318/0.32/0.434 |

4.2 实验结果与分析

4.2.1 参数调整

本文所提出的不一致性度量模块包含两个超参数: 实体不一致性度量函数的参数 λ_1 和非实体不一致性度量函数的参数 λ_2 . 它们的取值将会影响该模块最终的效果, 进而影响本文所提出的 FNFP 模型的性能. 为了研究这两个参数所带来性能的影响, 使用原型网络作为主要模型并增加不一致性度量模块, 通过依次改变两个参数值的大小以观察模型最终识别的效果, 实验结果如表 3 所示.

表 3 不同 λ_1 与 λ_2 在 FEW-NERD 下的实验结果对比

| λ_1 | λ_2 | Intra (%) | | | | Inter (%) | | | |
|--|---------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | 1-2 shot | | 5-10 shot | | 1-2 shot | | 5-10 shot | |
| | | 5-way | 10-way | 5-way | 10-way | 5-way | 10-way | 5-way | 10-way |
| 1 | 1 | 5.54 | 12.89 | 13.40 | 9.56 | 6.49 | 5.26 | 14.55 | 10.85 |
| em | em | 11.19 | 3.79 | 13.70 | 9.54 | 17.95 | 13.38 | 15.59 | 11.97 |
| $K \times em$ | $K \times em$ | 9.93 | 7.58 | 27.85 | 27.45 | 16.65 | 20.00 | 54.34 | 15.80 |
| $N \times K \times K \times em$ | $N \times K \times em$ | 19.82 | 12.89 | 40.97 | 27.45 | 36.67 | 31.51 | 59.45 | 51.11 |
| $N \times K \times K \times em$ | $N \times K \times K \times em$ | 16.82 | 12.89 | 41.04 | 27.46 | 38.33 | 31.47 | 59.45 | 52.59 |
| $N \times K \times K \times K \times em$ | $N \times K \times K \times em$ | 20.80 | 14.88 | 41.78 | 30.08 | 38.88 | 33.78 | 57.79 | 52.64 |
| $(N \times K \times (K-1) \times em)/2$ | $N \times K \times em$ | 22.13 | 16.25 | 42.07 | 35.10 | 39.55 | 34.76 | 61.03 | 54.72 |

通过调整参数的实验结果可以看出, 参数 λ_1 和 λ_2 的取值过大或过小都会影响模型的识别效果. 当参数的取值过小时, 所得的实体与非实体的不一致性过大, 在进行反向传播更新参数时会导致过模型参数在最优值的附近进行徘徊, 从而导致欠拟合现象的出现. 当参数的选取过大时, 所得的实体与非实体的不一致性过小, 模型的参数更新较慢, 在进行反向传播更新参数时会导致过模型参数难以找到最优值, 最终影响模型的性能.

综上所述, 实验得到的推荐参数值为: $\lambda_1 = (N \times K \times (K-1) \times em)/2$ 和 $\lambda_2 = N \times K \times em$. 我们经过多次尝试后给出的参数配置能够使得模型达到最优的识别效果, 该配置将作为本文后续实验工作的参数配置.

4.2.2 本文方法在 FEW-NERD 下的性能对比分析

为验证本文提出的 FNFP 模型可提高小样本命名实体识别的分类效果, 在大规模小样本命名实体识别数据集 FEW-NERD 上, 将模型与第 4.1.2 节中提到的 5 种基线模型进行实验对比, 结果如表 4 所示, 其中基线模型的实验结果取自于在相关文献中公开的数据.

从表 4 中可以看出, FNFP 在 FEW-NERD 数据集两种基准的 5-10 样本设置中, 性能都优于其他基线方法, 表明 FNFP 更适合多样本的设置. FNFP 模型是基于 Proto 模型进行改进的, 其性能较 Proto 模型得到了较大的提升, 尤其在 Intra 基准的 10-way 5-10 样本设置中, FNFP 模型较 Proto 模型在 $F1$ 值上提高了 12.84%.

表 4 不同方法在 FEW-NERD 下的对比实验结果 (%)

| 模型 | Intra | | | | Inter | | | |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | 1-2 shot | | 5-10 shot | | 1-2 shot | | 5-10 shot | |
| | 5-way | 10-way | 5-way | 10-way | 5-way | 10-way | 5-way | 10-way |
| NNShot | 25.78 | 18.27 | 36.18 | 27.38 | 47.24 | 38.87 | 55.64 | 49.57 |
| StructShot | 30.21 | 21.03 | 38.00 | 26.42 | 51.88 | 43.34 | 57.32 | 49.57 |
| Proto | 20.76 | 15.05 | 42.54 | 35.40 | 38.83 | 32.45 | 58.79 | 52.92 |
| BCL | 31.83 | 26.15 | 50.72 | 43.22 | 47.03 | 41.92 | 63.35 | 56.82 |
| Container | 40.43 | 33.84 | 53.70 | 47.49 | 55.95 | 48.35 | 61.83 | 57.12 |
| FNFP | 33.14 | 24.44 | 54.80 | 48.24 | 51.87 | 46.14 | 63.86 | 60.91 |

实验结果表明, Proto 作为经典的小样本学习模型, 在 5-10 样本中的性能较好, 原因是在多样本的设置下采用平均实体向量表示的方法构造原型, 能够准确地表示实体语义特征. 对于 1-2 样本的设置, 当两个实体的向量表示相差较大, 通过平均向量表示的方式进行原型构造, 其原型表示与实体的向量表示将会相差较大, 因而该原型不能表示实体的语义特征. NNshot 将最邻近分类方法应用到小样本命名实体识别中, 解决了 Proto 在 1-2 样本设置中的不足, 性能较 Proto 在 1-2 样本设置中得到了较好的提升. StructShot 在 NNshot 的输出阶段增加维特比解码器, 整体性能相比 NNshot 有了明显的提升, 表明引入维特比解码器来捕捉标签转移关系能够优化最终的输出, 由于 StructShot 的主干网络是 NNshot, 因而它们相对于 Proto 而言, 性能在 1-2 样本中的表现也相对较好.

BCL 和 Container 主要采用对比学习的思想, 通过对比区分不同类别的实体, 它们的效果相较 Proto、NNshot 和 StructShot 都得到了不错的提升. 相较于 BCL 学习单词的上下文跨度依赖表示以区分实体类, Container 通过计算单词的高斯分布嵌入的方式对区分不同类别的实体更加有效. 本文提出的 FNFP 通过细粒度原型网络改进原有 Proto 中非实体原型的构造方法, 其性能得到了显著的提升, 并在 5-10 样本的实验设置中得到的最佳的效果, 由于 FNFP 的实体类原型构造方式与 Proto 所采用的方式一样, 导致 FNFP 同样具有在 Proto 在 1-2 样本设置中的缺点, 即使通过不一致性度量模块和维特比解码器也很难将其弥补, 因此, FNFP 在 1-2 样本的实验设置中的效果并没有 Container 的效果好.

本文所提的 FNFP 模型在 5-10 样本的设置中效果最佳, 其 F1 值显著提升的原因如下.

1) FNFP 采用细粒度原型网络, 对实体与非实体使用不同的方法构造原型, 解决了因非实体之间不具有较强的语义关系所导致 Proto 构造原型不准确的问题, 提升模型对非实体类型的识别效果, 降低将实体错误分类为非实体的概率.

2) FNFP 融入不一致性度量模块, 使同类型实体的空间表示能够相互靠近, 因此, 构建的原型能够更准确地表示实体的语义特征, 解决了因同类实体表示相差较大所导致原型不准确的问题.

3) 细粒度原型网络与不一致性度量模块相辅相成, 共同促进了实体类原型和非实体类原型所表示的语义特征, 使得样本能够根据原型的表示得到更好的分类.

4) FNFP 融合维特比解码器捕捉标签转移关系后, 解决了部分原型仍然难以表示样本特征, 进而导致分类误差的问题.

4.2.3 本文方法在跨领域数据集下的性能对比分析

为了验证本文方法在不同领域场景下的泛化能力, 在跨领域数据集上, 将模型与第 4.1.2 节中的提到的 6 种基线模型进行实验对比, 结果如表 5 所示, 其中基线模型的实验结果取自于在相关文献中公开的数据.

从实验结果中可以看出, Bi-LSTM 和 TransferBERT 的表现并不是很好, 它们在样本较少的数据集上并不能够很好的优化模型, 可能会造成标签的随机预测, 这表明传统的机器学习方法和迁移学习方法不能够适应样本较少的环境. SimBERT 依靠不经微调的预训练语言模型 BERT 进行编码, 并通过词嵌入的余弦相似性预测标签类型, 效果得到了一定的提升, 这表明基于度量的方法在小样本环境下比传统的机器学习模型更具优势. Matching Network 和 WPZ+BERT 的效果一般, 很大程度上是因为其平等的关注所有支持集的单词, 容易受到支持集中大量非实体

的影响. L-TapNet+CDT 通过在条件随机场中引入标签依赖转移机制, 将标签依赖关系的先验知识转移到其他领域的标签之间, 从而提升了模型对新类别实体的识别效果, 在 1-shot 的 3 个领域和 5-shot 的 1 个领域中获得了最佳的效果. 本文所提的 FNFP 能够根据实体类和非实体类的特点, 采用不同的原型构造方式和不一致性度量函数, 提升原型的表示能力, 在 1-shot 的一个领域和 5-shot 的两个领域中获得了最佳的效果, 由于跨领域数据集的设置中支持集样本较少, FNFP 不能较好的去构造实体类原型, 导致在部分领域中的效果一般. 总体看来, 本文所提 FNFP 方法在不同领域场景中具有一定的泛化能力.

表 5 不同方法在跨领域下的对比实验结果 (%)

| 模型 | 1-shot | | | | 5-shot | | | |
|------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | News | Wiki | Social | Mixed | News | Wiki | Social | Mixed |
| Bi-LSTM | 2.57 | 3.29 | 0.69 | 2.11 | 6.81 | 8.40 | 1.06 | 13.17 |
| TransferBERT | 4.75 | 0.57 | 2.71 | 3.46 | 15.36 | 3.62 | 11.08 | 35.49 |
| SimBERT | 19.22 | 6.91 | 5.18 | 13.99 | 32.01 | 10.63 | 8.20 | 21.14 |
| Matching Network | 19.50 | 4.73 | 17.23 | 15.06 | 19.85 | 5.58 | 6.61 | 8.08 |
| WPZ+BERT | 32.49 | 3.89 | 10.68 | 6.67 | 50.06 | 9.54 | 17.26 | 13.59 |
| L-TapNet+CDT | 44.30 | 12.04 | 20.80 | 15.17 | 45.35 | 11.65 | 23.30 | 20.95 |
| FNFP | 24.50 | 8.61 | 10.89 | 26.90 | 51.58 | 16.10 | 20.62 | 29.91 |

4.2.4 消融实验

FNFP 模型主要有 3 个重要组件: 细粒度原型网络、不一致性度量模块和维特比解码器. 为了验证这 3 个组件对小样本命名实体识别模型 FNFP 的影响程度, 我们分别去掉或替换其中一个组件, 在 FEW-NERD 数据集上进行实验, 并与原模型进行比较.

消融实验结果如图 6 所示. 其中, 细粒度原型网络 FG-Proto 影响最大, 将 FG-Proto 替换为 Proto 后, 所有的实验设置中 $F1$ 值都下降较为明显, 表明该层是 FNFP 的核心组件. FG-Proto 对实体与非实体采用不同的方法构造原型, 解决了因非实体之间不具有较强的语义关系所导致 Proto 构造原型不准确的问题, 提升模型对非实体类型的识别效果. 不一致性度量模块 Incon 通过计算同类实体的差异性作为模型的额外损失, 使得同类型实体的空间表示能够相互靠近, 解决了因同类实体表示相差较大所导致原型不准确的问题. 去掉 Incon 组件后, 模型的性能并没有下降太多, 这是由于 FG-Proto 已经能够较好地构造原型, 使得 Incon 对于构造原型的准确率并没有提升太多. 去掉 Viterbi 解码器后, $F1$ 值有着明显的下降, 表明通过特比解码器捕捉标签转移关系, 能够较好地优化模型的分误差.

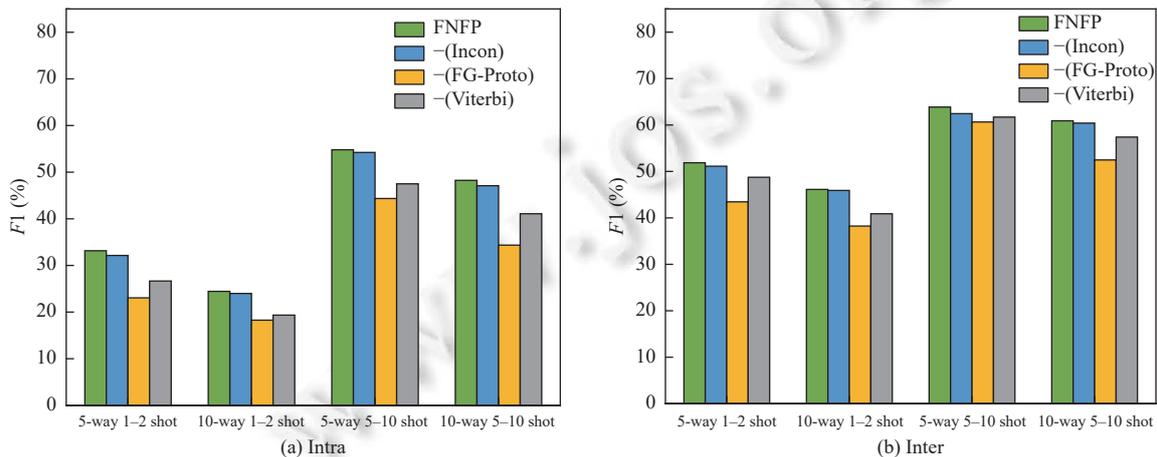


图 6 消融实验结果图

通过对消融实验的分析可知, 细粒度原型网络 FG-Proto、不一致性度量模块 Incon 和 Viterbi 解码器都能够增强 FNFP 模型在小样本命名实体识别任务中的识别效果。

4.3 实验小结

在大规模小样本命名实体识别数据集 FEW-NERD 上的大量实验表明, 基于细粒度原型网络的小样本命名实体识别方法 FNFP 的性能相较于基线方法得到了不错的性能提升, 并在 5–10 样本的实验设置中得到的最佳的效果。FNFP 方法是基于 Proto 方法中构造非实体原型不准确的特点进行改进的, 其性能相较于 Proto 方法在所有的实验设置中都得到了显著的提升, 这充分验证了 FNFP 方法中的细粒度原型网络 FG-Proto 能够较好地捕捉非实体的语义特征, 从而准确的构造非实体类原型。Intra 基准中训练集与测试集并不共享粗粒度实体类型, 这是一个具有挑战性的基准, Ding 等人^[15]在该基准上的实验结果并不理想。然而, 本文提出的 FNFP 模型在 Intra 基准中性能得到显著的提升, 验证了细粒度原型网络更好地捕捉样本的特征, 从而提高命名实体识别的 F1。同时在跨领域数据集上的实验表明, 本文所提 FNFP 方法在不同领域场景中同样具有泛化能力。

由消融实验分析可知:

- 1) 细粒度原型网络 FG-Proto 所构原型能够较好地捕捉非实体语义特征, 提升非实体的识别效果。
- 2) Incon 模块促使同类实体的表示相互靠近, 有益于实体原型更好的表示实体特征。
- 3) Viterbi 解码器通过捕捉标签转移关系, 能够较好地优化模型最终的分类结果。
- 4) Incon 模块的有效性得到验证, 但相较于细粒度原型网络所带来的性能提升一般, 仍有改进空间。

最后, 通过对 FNFP 有效性分析, 验证细粒度原型网络与不一致度量模块是相辅相成的, 使模型构造的实体原型与非实体原型能够较好地表示样本的特征, 提高模型的准确性。

5 总 结

本文分析了现有的小样本命名实体识别方法, 提出了基于细粒度原型网络的小样本命名实体识别方法, 首先对句子序列进行编码并为实体与非实体构建细粒度原型, 然后通过不一致性度量模块以使得同类实体之间的特征表示相互靠近, 最后使用维特比解码器捕捉标签转移关系。实验结果表明, 本文提出的小样本命名实体识别方法在 FEW-NERD 数据集下的 5–10 样本设置中得到最佳的效果, 但在 1–2 样本设置中未能取得最佳效果, 可能是由于在该设置中, 通过平均向量方式构造的原型表示与实体的向量表示相差较大。在未来的工作中, 考虑将本文方法与元学习模型进行结合并改进, 同时引入实例级别的注意力机制构造实体原型, 进一步提高模型的识别效果。

References:

- [1] Nadeau D, Sekine S. A survey of named entity recognition and classification. *Linguisticae Investigationes*, 2007, 30(1): 3–26. [doi: 10.1075/li.30.1.03nad]
- [2] Zhang SD, Elhadad N. Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts. *Journal of Biomedical Informatics*, 2013, 46(6): 1088–1098. [doi: 10.1016/j.jbi.2013.08.004]
- [3] Quimbaya AP, Múnera AS, Rivera RAG, Rodríguez JCD, Velandia OMM, Peña AAG, Labbé C. Named entity recognition over electronic health records through a combined dictionary-based approach. *Procedia Computer Science*, 2016, 100: 55–61. [doi: 10.1016/j.procs.2016.09.123]
- [4] Zhang CY, Hong XG, Peng ZH, Li QZ. Extracting Web entity activities based on SVM and extended conditional random fields. *Ruan Jian Xue Bao/Journal of Software*, 2012, 23(10): 2612–2627 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4189.htm> [doi: 10.3724/SP.J.1001.2012.04189]
- [5] Morwal S, Jahan N, Chopra D. Named entity recognition using hidden Markov model (HMM). *Int'l Journal on Natural Language Computing*, 2012, 1(4): 15–23. [doi: 10.5121/ijnlc.2012.1402]
- [6] Ji ZC, Sun AX, Cong G, Han JL. Joint recognition and linking of fine-grained locations from tweets. In: *Proc. of the 25th Int'l Conf. on World Wide Web*. Montréal: Int'l World Wide Web Conf. Steering Committee, 2016. 1271–1281. [doi: 10.1145/2872427.2883067]
- [7] Li J, Sun AX, Han JL, Li CL. A survey on deep learning for named entity recognition. *IEEE Trans. on Knowledge and Data Engineering*, 2022, 34(1): 50–70. [doi: 10.1109/TKDE.2020.2981314]

- [8] Peng N, Dredze M. Improving named entity recognition for Chinese social media with word segmentation representation learning. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: Association for Computational Linguistics, 2016. 149–155. [doi: [10.18653/v1/P16-2025](https://doi.org/10.18653/v1/P16-2025)]
- [9] Li PH, Dong RP, Wang YS, Chou JC, Ma WY. Leveraging linguistic structures for named entity recognition with bidirectional recursive neural networks. In Proc. of the 2017 Conf. on Empirical Methods in Natural Language Processing. Copenhagen: Association for Computational Linguistics, 2017. 2664–2669. [doi: [10.18653/v1/D17-1282](https://doi.org/10.18653/v1/D17-1282)]
- [10] Hofer M, Kormilitzin A, Goldberg P, Nevado-Holgado A. Few-shot learning for named entity recognition in medical text. arXiv: 1811.05468, 2018.
- [11] Yang Y, Katiyar A. Simple and effective few-shot named entity recognition with structured nearest neighbor learning. In: Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2020. 6365–6375. [doi: [10.18653/v1/2020.emnlp-main.516](https://doi.org/10.18653/v1/2020.emnlp-main.516)]
- [12] Li J, Chiu B, Feng SS, Wang H. Few-shot named entity recognition via meta-learning. IEEE Trans. on Knowledge and Data Engineering, 2022, 34(9): 4245–4256. [doi: [10.1109/TKDE.2020.3038670](https://doi.org/10.1109/TKDE.2020.3038670)]
- [13] Fritzier A, Logacheva V, Kretov M. Few-shot classification in named entity recognition task. In: Proc. of the 34th ACM/SIGAPP Symp. on Applied Computing. Limassol: Association for Computing Machinery, 2019. 993–1000. [doi: [10.1145/3297280.3297378](https://doi.org/10.1145/3297280.3297378)]
- [14] De Lichy C, Glaude H, Campbell W. Meta-learning for few-shot named entity recognition. In: Proc. of the 1st Workshop on Meta Learning and Its Applications to Natural Language Processing. Association for Computational Linguistics, 2021. 44–58. [doi: [10.18653/v1/2021.metanlp-1.6](https://doi.org/10.18653/v1/2021.metanlp-1.6)]
- [15] Ding N, Xu GG, Chen YL, Wang XB, Han X, Xie PJ, Zheng HT, Liu ZY. FEW-NERD: A few-shot named entity recognition dataset. In: Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on Natural Language Processing (Vol. 1: Long Papers). Association for Computational Linguistics, 2021. 3198–3213. [doi: [10.18653/v1/2021.acl-long.248](https://doi.org/10.18653/v1/2021.acl-long.248)]
- [16] Bengio Y. Deep learning of representations for unsupervised and transfer learning. In: Proc. of the 2011 Int'l Conf. on Unsupervised and Transfer Learning Workshop. Washington: JMLR.org, 2011. 17–37.
- [17] Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. In: Proc. of the 32nd Int'l Conf. on Machine Learning Deep Learning Workshop. Lille, 2015.
- [18] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 4080–4090.
- [19] Sun SL, Sun QF, Zhou K, Lv TC. Hierarchical attention prototypical networks for few-shot text classification. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing. Hong Kong: Association for Computational Linguistics, 2019. 476–485. [doi: [10.18653/v1/D19-1045](https://doi.org/10.18653/v1/D19-1045)]
- [20] Han X, Zhu H, Yu PF, Wang ZY, Yao Y, Liu ZY, Sun MS. Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 4803–4809. [doi: [10.18653/v1/D18-1514](https://doi.org/10.18653/v1/D18-1514)]
- [21] Tong MH, Wang SA, Xu B, Cao YX, Liu MH, Hou L, Li JZ. Learning from miscellaneous other-class words for few-shot named entity recognition. In: Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on Natural Language Processing. Association for Computational Linguistics, 2021. 6236–6247. [doi: [10.18653/v1/2021.acl-long.487](https://doi.org/10.18653/v1/2021.acl-long.487)]
- [22] Chen JW, Liu Q, Lin HY, Han XP, Sun L. Few-shot named entity recognition with self-describing networks. In: Proc. of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin: Association for Computational Linguistics, 2022. 5711–5722. [doi: [10.18653/v1/2022.acl-long.392](https://doi.org/10.18653/v1/2022.acl-long.392)]
- [23] Zhao KL, Jin XL, Wang YZ. Survey on few-shot learning. Ruan Jian Xue Bao/Journal of Software, 2021, 32(2): 349–369 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6138.htm> [doi: [10.13328/j.cnki.jos.006138](https://doi.org/10.13328/j.cnki.jos.006138)]
- [24] Yang MD. A survey on few-shot learning in natural language processing. In: Proc. of the 2021 Int'l Conf. on Artificial Intelligence and Electromechanical Automation. Guangzhou: IEEE, 2021. 294–297. [doi: [10.1109/AIEA53260.2021.00069](https://doi.org/10.1109/AIEA53260.2021.00069)]
- [25] Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Neural architectures for named entity recognition. In: Proc. of the 2016 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego: Association for Computational Linguistics, 2016. 260–270. [doi: [10.18653/v1/N16-1030](https://doi.org/10.18653/v1/N16-1030)]
- [26] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2018. 4171–4186. [doi: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423)]
- [27] Hou YT, Che WX, Lai YK, Zhou ZH, Liu YJ, Liu H, Liu T. Few-shot slot tagging with collapsed dependency transfer and label-

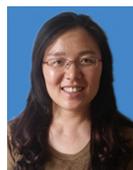
- enhanced task-adaptive projection network. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020. 1381–1393. [doi: [10.18653/v1/2020.acl-main.128](https://doi.org/10.18653/v1/2020.acl-main.128)]
- [28] Sang EFTK, De Meulder F. Introduction to the conll-2003 shared task: Language-independent named entity recognition. In: Proc. of the 7th Conf. on Natural Language Learning at HLT-NAACL 2003. 2003. 142–147.
- [29] Zeldes A. The GUM corpus: Creating multilayer resources in the classroom. Language Resources and Evaluation, 2017, 51(3): 581–612. [doi: [10.1007/s10579-016-9343-x](https://doi.org/10.1007/s10579-016-9343-x)]
- [30] Derczynski L, Nichols E, van Erp M, Limsopatham N. Results of the wnut2017 shared task on novel and emerging entity recognition. In: Proc. of the 3rd Workshop on Noisy User-generated Text. Copenhagen: Association for Computational Linguistics, 2017. 140–147. [doi: [10.18653/v1/W17-4418](https://doi.org/10.18653/v1/W17-4418)]
- [31] Pradhan S, Moschitti A, Xue NW, Ng HT, Björkelund A, Uryupina O, Zhang YC, Zhong Z. Towards robust linguistic analysis using ontonotes. In: Proc. of the 17th Conf. on Computational Natural Language Learning. Sofia: Association for Computational Linguistics, 2013. 143–152.
- [32] Ming H, Yang JY, Jiang LL, Pan Y, An N. Few-shot nested named entity recognition. arXiv:2212.00953, 2022.
- [33] Das SSS, Katiyar A, Passonneau R J, Zhang R. Container: Few-shot named entity recognition via contrastive learning. In: Proc. of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin: Association for Computational Linguistics, 2021. 6338–6353. [doi: [10.18653/v1/2022.acl-long.439](https://doi.org/10.18653/v1/2022.acl-long.439)]
- [34] Vinyals O, Blundell C, Lillicrap T, Kavukcuoglu K, Wierstra D. Matching networks for one shot learning. In: Proc. of the 30th Int'l Conf. on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 3637–3645.

附中文参考文献:

- [4] 张传岩, 洪晓光, 彭朝晖, 李庆忠. 基于SVM和扩展条件随机场的Web实体活动抽取. 软件学报, 2012, 23(10): 2612–2627. <http://www.jos.org.cn/1000-9825/4189.htm> [doi: [10.3724/SP.J.1001.2012.04189](https://doi.org/10.3724/SP.J.1001.2012.04189)]
- [23] 赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述. 软件学报, 2021, 32(2): 349–369. <http://www.jos.org.cn/1000-9825/6138.htm> [doi: [10.13328/j.cnki.jos.006138](https://doi.org/10.13328/j.cnki.jos.006138)]



戚荣志(1980—), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为实体识别, 关系抽取, 智能软件工程.



李水艳(1980—), 女, 讲师, CCF 专业会员, 主要研究领域为实体识别, 关系抽取.



周俊宇(1999—), 男, 硕士, 主要研究领域为小样本命名实体识别.



毛莺池(1976—), 女, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为分布式数据处理, 边缘智能计算.