小样本命名实体识别方法研究

文献综述

计算机科学与技术学院 学硕2102 2212523 胡怡淳

指导教师 卢婷

1. 课题背景

命名实体识别旨在从非结构化的文本中识别出有意义的实体，并将其分类为一组预定义的实体类型，如人名、机构名、地名、医疗代码、日期、数量等[1]。命名实体通常含有丰富的语义信息，并与语料中的关键信息有着紧密的联系，因此命名实体识别是许多下游任务的重要组成部分，如关系抽取[2-3]、事件抽取[4]、问答[5]、知识图谱构建[6]等。命名实体识别的结果会对后续任务的性能产生重大影响，可能会导致误差传递[7]等问题，因此命名实体识别一直以来都是热门的研究课题。

最早提出的是基于词典和规则的方法，这种方法依赖领域内的实体词典以及人工构建的规则库。基于规则的方法的不需要标注语料，但规则库依赖人工构建同样存在代价大和耗时的问题，并且规则库中的规则难以涵盖所有情况，容易出现错误，而且这些规则依赖于具体的领域、文本风格和语言，可移植性差[8]。伴随着机器学习的广泛使用，提出了许多基于机器学习的命名实体识别方法，将命名实体识别问题转换为了序列标注问题或者分类问题，通过机器学习方法将标注语料转化为特征向量，再通过分类模型来识别实体[9]。这种方法利用人工标注的语料数据进行训练，标注语料时不需要广博的语言学知识，并且可以通过在新的领域中继续训练来进行移植，在一定程度上克服了基于词典和规则的方法的局限性。当前随着人工智能技术的日新月异，基于深度学习和大规模预训练语言模型的命名实体识别方法是目前应用最广泛的方法，通过预训练语言模型获取语料的表征，利用神经网络自动提取特征，最后通过解码器来预测实体类别[10]。

尽管基于深度学习的命名实体识别方法已经取得了非常优异的性能，但是这些方法通常都需要大量的人工标注语料数据才能达到足够好的效果。对于那些标注数据就比较少的领域，比如金融、餐饮、影视等，从头开始构建一个高质量的标注数据集需要花费巨大的代价，需要丰富的领域内知识以及大量的人力和时间成本[1]。因此，现在有越来越多的研究聚焦于更具挑战性的小样本命名实体识别任务。在小样本命名实体识别中，模型只通过非常少的标注数据来学习如何识别和分类命名实体，这样大幅降低了数据的标注成本，也更加符合现实当中的实际情况。

1. 研究现状

小样本命名实体识别发展至今已经提出了许多方法，概括来说可以分为元学习方法、迁移学习方法、提示学习方法等。

2.1 元学习方法

元学习[11]方法在小样本学习任务中取得了非常好的效果，也常被用于解决小样本命名实体识别问题[1,12]。其根本思想是通过原型网络将每个类别的数据映射到一个相同的空间中，并取这些向量的均值作为该类型的原型来代表这个类型，通过测试数据与各原型之间的距离来决定最终的类别。

具体来说，每次从数据集中随机采样包含种实体类型的支持集和查询集，其中是输入的句子，是句子中每个词的实体类别，可以小于或者等于数据集的实体类型的数量，支持集中每个类型包含个例子，查询集中每个类型包含个例子。对于类别，定义为中属于该实体类别的词的集合，以此来构建类的原型

|  |  |  |
| --- | --- | --- |
|  |  | （1） |

其中为将输入映射到向量空间的函数。对于查询集中的输入，通过计算与原型之间的距离作为与类别之间的相似度（通常采用欧式距离），最后使用函数将相似度转化为类的分布。

在命名实体识别任务中，存在一种非实体类型，它通常含有混杂的语义，很难用一个原型去概括，并且非实体类型的词不需要也不应该在向量空间中彼此靠近[12]。为了减轻非实体类型的影响，大部分方法都会使用固定标量或其他方法来代替非实体类型的原型。近期有研究提出对非实体类型进行再次分类，并为其建立多个原型，以充分利用非实体语料，并提前学习潜在的看不见的类型[13]。

2.2 迁移学习方法

迁移学习[14]方法通常是先在一个拥有较多标注数据的高资源领域中进行预训练，学习一般的表征，然后再将其迁移到标注数据较少的低资源领域，在少量数据上进行微调，使模型适应新的任务[15]。迁移学习方法在标注数据非常少的情况下特别有效。

由于预训练时的源领域和微调时的目标领域各自包含的实体类型通常并不一致，需要在微调时舍弃原来的分类层并添加新的分类层重新训练[16]，这样导致分类层的训练并不充分，最终模型的性能不够理想。最近有许多研究针对这一问题，提出对类别信息进行统一的建模，使得模型能够适应不同类别的情况。

SMXM[17]模型提出利用类别描述进行零样本的命名实体识别，对于每个类别，将对应的类别描述拼接在输入句子之后，经过一个交叉注意力编码器生成每个词的表征，最后经过一系列线性变换得到每个词属于该类的概率。由于是零样本命名实体识别任务，该方法采用的类别描述来自外部数据，包括数据集的注释文档以及专业的百科数据库等。类似的，Label Semantics[18]提出利用类名进行小样本命名实体识别，将类名与begin、inside等词相结合，将BIO形式的实体标签（如B-PER，I-ORG等）转化为自然语言形式（如begin person，inside organization等）。模型采用双塔结构，由两个编码器组成，一个用来编码输入的句子另一个用来编码类名，最后利用类名和词的表征之间的点积作为相似度判断每个词所属的类别。

SDNet[19]提出了一种用于小样本命名实体识别的自描述机制，通过一个共通的概念集合来概括实体提及（mention），并将概念映射到实体类别。具体而言，SDNet是一个序列到序列的生成式模型，依次执行两个生成任务：生成实体提及的概念描述；自适应地生成与新类型一一对应的实体提及。

NNShot[20]提出了一种使用最近邻进行小样本命名实体识别的方法，基于经过预训练之后相同实体类型的词会在向量空间中靠的更近的想法，对于输入中的每个词，在支持集中寻找在向量空间中与其欧氏距离最近的词，并赋予它相同的实体标签。CONTaiNER[21]模型也采用了类似的方法，采用了基于高斯分布的嵌入，并引入了对比学习。NNShot还在TapNet[22]的基础上提出了一种简单但有效的方法来捕获实体标签之间的依赖，取代了原先耗时的条件随机场[23]的训练，并且可以任意目标领域。

2.3 提示学习方法

随着大规模预训练语言模型的发展，许多任务在预训练模型上加以微调就能达到不错的效果。但一般的微调方法是添加一个额外的分类器（通常是线性层和softmax），这样会引入新的参数，而且模型越大参数量也会越大，从少量样本中学习会比较困难[24]。为了不引入新的参数，同时充分利用预训练语言模型的先验知识，提出了提示学习的方法[25]，借鉴了遮蔽语言模型的方式，将下游任务转化为完形填空式的问题。具体来说就是为任务构建提示模板，然后通过语言模型对模板进行补全，从而得到预测结果。

TemplateNER[26]第一个将提示学习思想用于小样本命名实体识别任务，提出了一种基于模板的方法，利用生成式预训练语言模型来进行小样本命名实体识别。该方法为实体和非实体分别构造了模板，对于输入中的每个片段（span），填入模板中的实体插槽，并使用生成式语言模型对所有模板进行打分。

提示学习方法的主要问题在于模板和候选词的构造。PLET[27]提出对每个类别使用标签词集，加入类名及其近义词，优化单词在标签词集上分布的相似性。ProtoVerb[28]提出了从训练数据构建原型表达器的方法，将从训练数据预测得到的候选词投射到嵌入空间，并通过对比学习得到原型。Prompt-based FT[24]提出使用训练数据预测得到的预测词的前k个作为候选词，并枚举所有情况选出最好的标签词组合，再使用生成式模型自动生成模板。PET[29]提出构建多个模板和候选词的组合，使用多个预训练语言模型在不同的组合上进行微调，然后将这些微调模型进行集成，最后将集成模型的知识进行蒸馏得到输出模型。

EntLM[30]提出了一种无模板的方法，将命名实体识别问题转化为一个以实体为导向的语言模型问题，实体词预测为标签词，非实体词预测为原本的单词，不再需要枚举所有片段，提升了推理速度。

提示学习方法还会通过在输入的同时引入一些示例，通过上下文学习来进一步提升效果[23]。论文[31]对示例的选择策略、示例包含的内容、示例的形式等问题进行了系统的研究。Structured Prompting[32]引入了结构化提示，打破了传统上下文学习的长度限制，将上下文学习扩展到数千个例子。

1. 前沿研究的主要问题

（1）类别描述信息的获取。对于类别信息进行建模的方法已经提出了很多种，但其中的类别描述信息大多数使用的是类名，类名作为最基本的类别描述，获取容易但包含的信息量也较少。也有一些研究使用的是比较详细的类别描述信息，包含精确的定义，有的还有一些列子，但这种描述通常来自于数据集本身提供的注释指南，或者是一些专业可靠的百科信息，需要借助外部信息才能获取，并不适用于所有情况。

（2）实体标签之间的依赖。条件随机场对于命名实体识别任务有非常大的提升，但在小样本上训练它是十分困难的，所以提出了利用抽象转移矩阵概括一些抽象标签之间的依赖关系，再将其推广到任意目标领域，而概括与推广的方法会对模型的最终性能产生很大的影响。

（3）数据增强。数据增强一直以来都是解决资源稀缺问题最主要的手段，现有研究针对命名实体识别问题的数据增强方法也已经提出了许多，但主要还是针对低资源情况，对于小样本这种每个类只有几个样本的极低资源的情况并没有非常合适的解决方案。

参考文献

[1] HUANG J, LI C, SUBUDHI K, et al. Few-shot named entity recognition: An empirical baseline study[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021: 10408-10423.

[2] PENG H, GAO T, HAN X, et al. Learning from Context or Names? An Empirical Study on Neural Relation Extraction[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020: 3661-3672.

[3] BALDINI SOARES L, FITZGERALD N, LING J, et al. Matching the blanks: Distributional similarity for relation learning[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 2895-2905.

[4] RITTER A, ETZIONI O, CLARK S. Open domain event extraction from twitter[C]//Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. 2012: 1104-1112.

[5] MOLLÁ D, VAN ZAANEN M, SMITH D. Named entity recognition for question answering[C]//Proceedings of the Australasian language technology workshop 2006. 2006: 51-58.

[6] JI S, PAN S, CAMBRIA E, et al. A survey on knowledge graphs: Representation, acquisition, and applications[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(2): 494-514.

[7] ZHONG Z, CHEN D. A frustratingly easy approach for entity and relation extraction[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2021: 50-61.

[8] WEN Y, FAN C, CHEN G, et al. A survey on named entity recognition[C]//Communications, Signal Processing, and Systems. 2020: 1803-1810.

[9] ROY A. Recent trends in named entity recognition (ner): abs/2101.11420[A]. 2021.

[10] LI J, SUN A, HAN J, et al. A survey on deep learning for named entity recognition[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(1): 50-70.

[11] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning[C]//Advances in Neural Information Processing Systems: volume 30. 2017.

[12] FRITZLER A, LOGACHEVA V, KRETOV M. Few-shot classification in named entity recognition task[C]//Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing. 2019: 993-1000.

[13] TONG M, WANG S, XU B, et al. Learning from miscellaneous other-class words for few-shot named entity recognition[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021: 6236-6247.

[14] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

[15] ZHUANG F, QI Z, DUAN K, et al. A comprehensive survey on transfer learning[J]. Proceedings of the IEEE, 2021, 109(1): 43-76.

[16] ZHUANG F, LUO P, XIONG H, et al. Cross-domain learning from multiple sources: A consensus regularization perspective[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(12): 1664-1678.

[17] ALY R, VLACHOS A, MCDONALD R. Leveraging type descriptions for zero-shot named entity recognition and classification[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021: 1516-1528.

[18] MA J, BALLESTEROS M, DOSS S, et al. Label semantics for few shot named entity recognition[C]//Findings of the Association for Computational Linguistics: ACL 2022. 2022: 1956-1971.

[19] CHEN J, LIU Q, LIN H, et al. Few-shot named entity recognition with self-describing networks[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 5711-5722.

[20] YANG Y, KATIYAR A. Simple and effective few-shot named entity recognition with structured nearest neighbor learning[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 6365-6375.

[21] DAS S S S, KATIYAR A, PASSONNEAU R, et al. CONTaiNER: Few-shot named entity recognition via contrastive learning[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022: 6338-6353.

[22] HOU Y, CHE W, LAI Y, et al. Few-shot slot tagging with collapsed dependency transfer and label-enhanced task-adaptive projection network[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 1381-1393.

[23] LAFFERTY J, MCCALLUM A, PEREIRA F C. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[C]//Proc. 18th International Conference on Machine Learning, 2001. 2001: 282-289.

[24] GAO T, FISCH A, CHEN D. Making pre-trained language models better few-shot learners[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 3816-3830.

[25] HU S, DING N, WANG H, et al. Knowledgeable prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022: 2225-2240.

[26] CUI L, WU Y, LIU J, et al. Template-based named entity recognition using BART[C]//Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. 2021: 1835-1845.

[27] DING N, CHEN Y, HAN X, et al. Prompt-learning for fine-grained entity typing[C]//Conference on Empirical Methods in Natural Language Processing. 2021.

[28] CUI G, HU S, DING N, et al. Prototypical verbalizer for prompt-based few-shot tuning[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022: 7014-7024.

[29] SCHICK T, SCHÜTZE H. Exploiting cloze-questions for few-shot text classification and natural language inference[C]//Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. 2021: 255-269.

[30] MA R, ZHOU X, GUI T, et al. Template-free prompt tuning for few-shot NER[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2022: 5721-5732.

[31] LEE D H, KADAKIA A, TAN K, et al. Good examples make a faster learner: Simple demonstration-based learning for low-resource NER[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022: 2687-2700.

[32] HAO Y, SUN Y, DONG L, et al. Structured prompting: Scaling in-context learning to 1, 000 examples: abs/2212.06713[A]. 2022.