



计算机科学与探索

Journal of Frontiers of Computer Science and Technology

ISSN 1673-9418, CN 11-5602/TP

《计算机科学与探索》网络首发论文

题目：命名实体识别的迁移学习研究综述
作者：李猛，李艳玲，林民
网络首发日期：2020-09-25
引用格式：李猛，李艳玲，林民. 命名实体识别的迁移学习研究综述. 计算机科学与探索. <https://kns.cnki.net/kcms/detail/11.5602.TP.20200925.1522.010.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

命名实体识别的迁移学习研究综述*

李 猛, 李艳玲⁺, 林 民

内蒙古师范大学 计算机科学技术学院, 呼和浩特 010022

⁺ 通讯作者 E-mail: liyanling7871397@163.com

摘 要: 命名实体识别是自然语言处理的核心应用任务之一。传统和深度命名实体识别方法严重依赖于大量具有相同分布的标注训练数据, 模型可移植性差。然而在实际应用中数据往往都是小数据、个性化数据, 收集足够的训练数据是非常困难的。在命名实体识别中引入迁移学习, 利用源域数据和模型完成目标域任务模型构建, 提高目标领域的标注数据量和降低目标域模型对标注数据数量的需求, 在处理资源匮乏命名实体识别任务上, 具有非常好的效果。本文首先对命名实体识别方法和难点以及迁移学习的方法进行概述, 然后对近些年应用于命名实体识别的迁移学习方法包括基于数据迁移学习、基于模型迁移学习和对抗迁移学习进行全面综述, 重点阐述了对抗迁移学习方法, 最后进一步思考当前存在的问题并对未来的研究方向进行了展望。

关键词: 命名实体识别; 迁移学习; 对抗迁移学习; 深度学习

文献标志码: A **中图分类号:** TP391

李猛, 李艳玲, 林民. 命名实体识别的迁移学习研究综述[J]. 计算机科学与探索

LI M, LI Y L, LIN M. Review of Transfer Learning for Named Entity Recognition[J]. Journal of Frontiers of Computer Science and Technology

Review of Transfer Learning for Named Entity Recognition *

LI Meng, LI Yanling⁺, LIN Min

College of Computer Science and Technology, Inner Mongolia Normal University, Hohhot 010022, China

Abstract: Named entity recognition (NER) is one of the core application tasks of natural language processing. Tra-

* The National Natural Science Foundation of China under Grant Nos. 61806103, 61562068 (国家自然科学基金); the Natural Science Foundation of Inner Mongolia under Grant No. 2017MS0607 (内蒙古自然科学基金); the Special Project of Inner Mongolia Ethnic Affairs under Grant No. MW-2014-MGYWXXH-01 (内蒙古民委蒙古文信息化专项扶持子项目); the Youth Innovation and Entrepreneurship Talents of Inner Mongolia "grassland talents" Project (内蒙古自治区“草原英才”工程青年创新创业人才项目); the Graduate Innovation Fund of Inner Mongolia Normal University under Grant No. CXJJS19151 (内蒙古师范大学研究生创新基金); the Inner Mongolia Autonomous Region Science and Technology Planning Project Mongolian MOOC Online Interactive Learning Platform R&D Operation and Ethnic Education Big Data Analysis and Research No. JH20180175 (内蒙古自治区科技计划项目蒙古文 MOOC 在线互动学习平台研发运营及民族教育大数据分析研究).

ditional and deep NER methods rely heavily on a large amount of labeled training data with the same distribution, and the portability of the model is very poor. However, data is small and personalized in practical applications, and it is very difficult to collect enough training data. Transfer learning is used in NER, the data and model of the source domain are utilized to complete the target task model construction, increase the amount of labeled data in the target domain and reduce the demand for the amount of labeled data of the target model. It has a very good effect in dealing with the task of low-resource NER. Firstly, the method and difficulty of NER and the method of transfer learning are summarized. Then, transfer learning methods applied to NER include data-based transfer learning, model-based transfer learning and adversarial transfer learning in recent years are comprehensive reviewed, and adversarial transfer learning method is mainly described. Finally, we further expound the current problems and look forward to the future research directions.

Key words: named entity recognition; transfer learning; adversarial transfer learning; deep learning

1 引言

命名实体识别 (Named Entity Recognition, NER) 是自然语言处理 (Natural Language Processing, NLP) 一项非常重要的基础任务, 旨在自动检测文本中的命名实体并将其分类为预定义实体类型, 例如: 人名、地名、组织机构名等, 是人机对话系统、机器翻译、关系抽取等的前置任务^[1]。传统以及深度 NER^[2-6]方法已经取得了非常高的识别精度, 但是训练模型需要大规模标注数据, 模型性能与标注数据量成正比, 在训练语料匮乏的特定领域 (如: 医学、生物等) 和小语种 (如: 蒙古语、维吾尔语等) 上, 性能差强人意。由于训练集和测试集要求独立同分布, 所以导致将已有的模型运用到其他领域或者语言上性能差强人意。

迁移学习旨在利用源域中大量标注数据和预训练模型提高目标任务学习性能, 凭借其对数据、标注依赖性小和放宽了独立同分布约束条件等优点^[7,8], 已经成为解决资源匮乏 NER 的强大工具。NER 的迁移学习方法早期工作主要集中在基于数据的方法, 利用并行语料库、双语词典等作为桥梁将知识 (如: 标注、特征表示等) 从高资源语言投影到低资源语言, 主要用于跨语言 NER 迁移。后来研究者将源模型部分参数或特征表示迁移到目标模型上, 不需要额外的对齐信息, 实现了跨领域和跨应用 NER 迁移并取得了非常好的效果。最近 NER 对抗

迁移学习受到越来越多研究人员的关注, 引入由生成对抗网络 (Generative Adversarial Networks, GAN)^[9]启发的对抗技术, 生成一种“域无关特征”, 进而实现源域知识到目标域的迁移, 帮助目标任务提高学习性能, 同时有效缓解了负迁移问题。

刘浏等人^[10]从 NER 定义、评测会议、主流研究方法等角度, 介绍了 NER 任务的发展历程; Li 等人^[8]详细地总结和分析了 NER 的深度学习方法, 以及深度 NER 任务面临的挑战和未来发展方向。Pan 和 Yang^[11]的综述是一项开创性工作, 对迁移学习进行了定义和分类, 并回顾了 2010 年之前的研究进展; Weiss 等人^[12]介绍并总结了许多同构和异构迁移学习方法; Zhuang 等人^[8]从数据和模型角度对 40 多种具有代表性的迁移学习模型进行了介绍和对比。以上综述都是对 NER 或迁移学习单方面的阐述, 没有详细地介绍两者结合的方法。本文从基于数据迁移学习、基于模型迁移学习、对抗迁移学习这三个方面对目前 NER 任务的迁移学习方法进行了研究和调查。

2 命名实体识别

2.1 传统 NER 方法

传统的 NER 方法大致有三类: 基于规则、无监督学习和基于特征的有监督学习^[1,12]。基于规则的方法依赖语言学家和领域专家手工制定的语义语法规

则,通过规则匹配识别各种类型的命名实体,基于规则的方法虽然能够在特定语料上(字典详尽且大小有限)获得很好的效果,但是构建这些规则不仅耗时、难以覆盖所有规则,而且可扩展性和可移植性比较差。无监督学习方法利用在大型语料库上获得的词汇资源、词汇模型和统计信息,使用聚类^[13]推断命名实体类型。基于特征的有监督学习方法通过监督学习,将NER任务转换为序列标注任务,根据标注数据,将每个训练样本用精心设计的特征表示出来,然后利用机器学习算法训练模型,从看不见的数据中学习相似的模式,其缺点是需要大量人工标注训练数据和人为构造、选择的有效特征^[14]。

基于规则的方法主要有:LaSIE-II^[15]、NetOwl^[16]、Facile^[17]。基于特征的有监督学习算法主要有:隐马尔可夫模型(Hidden Markov Model, HMM)^[18]、决策树(Decision Trees)^[19]、最大熵模型(Maximum Entropy Model, MEM)^[20]、支持向量机(Support Vector Machines, SVM)^[21]、条件随机场(Conditional Random Fields, CRF)^[22]等。

2.2 深度NER方法

近年来,随着深度神经网络的迅猛发展,NER的深度学习方法已成为主流。深度学习有三个优势:(1)深度神经网络可以进行非线性变换,进而从数据中学习更复杂的特征;(2)深度学习可以从原始数据中自动学习特征;(3)深度NER模型属于端到端模型^[8]。

NER的深度学习方法一般分为三个阶段^[8],如图1所示,分别是:分布式表示、特征提取器和标签解码器。分布式表示把单词或字符映射到低维实值密集向量中,其中每个维度代表一个隐藏特征,它可以自动从文本中学习语义和句法特征,分布式表示有三种:词向量(Word Embedding)^[23]、字符向量(Character Embedding)^[14]和混合表示^[3]。特征提取器通过接收上一层的向量对上下文特征表示进行学习,常用的特征提取器有:CNN^[2]、Transformer^[24]、RNN^[25]以及它的两种变体:门控循环单元(Gated Recurrent Unit, GRU)^[26]和长短时记忆(Long Short-Term Memory, LSTM)^[3]。标签解码

器是最后一个阶段,以上下文特征为输入,生成标签序列,常见的解码方式有:Softmax^[23]、CRF、RNN。

B-PER I-PER O B-LOC I-LOC I-LOC I-LOC O O O
李 明 在 呼 和 浩 特 出 生 。

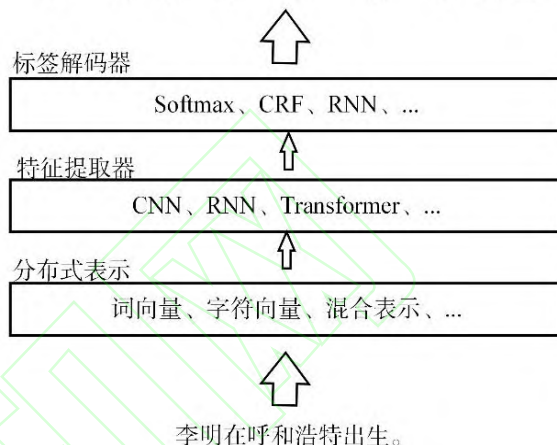


Fig.1 Deep learning for NER flowchart
图1 NER的深度学习方法流程图

2.3 深度NER的难点

(1) 资源匮乏。深度学习需要大规模标注数据才能很好地训练模型,但数据标注非常耗时且昂贵,尤其对于许多资源匮乏的语言和特定领域,如:蒙古语,医学、军事领域。当标注数据较少时,由于无法充分学习隐藏特征,深度学习的性能会大大降低,而且深度学习模型可移植性很差,无法将已有数据和模型应用到资源匮乏领域^[7]。因此,采用半监督学习和无监督学习实现资源的自动构建和补足,以及迁移学习等方法都可作为解决该问题的研究方向^[27]。

(2) 非正式文本。非正式文本即表达不符合书面语法规则的文本,如人机对话系统用户提问、Twitter和微博等社交媒体上的文章评论等用户生成文本,由于其语句简短、口语化、内容宽泛、语意含糊、包含谐音字,使NER变得非常困难,甚至无法识别。例如:“我中奖了”写成“我中了”;“杯具”表示“悲剧”;“xswl”代表“笑死我了”。Li等人使用主动学习将微博文本NER的F1值提高到了67.23%,但是仍需要人工标记^[28]。可以使用注意力机制和迁移学习结合深度学习完成对非正式文本的识别。

(3) 命名实体多样性。随着信息化时代的来临, 移动互联网的普及, 数据规模海量化, 命名实体及其类型趋于多样化, 同时也在不断演变^[29]。传统的实体类型只有人名、地名和组织机构名, 但是现实生活中实体类型复杂多样, 不同领域存在不同的实体类型, 需要识别更详细的实体类型, 例如: 交通查询领域, 需要出发地、目的地、时间、交通工具等实体类型。可以使用迁移学习技术, 重复利用已有数据和模型, 实现细粒度 NER。

(4) 命名实体歧义性。自然语言中存在大量歧义问题, 这给 NER 带来很大挑战。在不同文化、背景、领域下, 同一实体可能含有不同的含义, 例如: “香格里拉”可能是“香格里拉市”也有可能是“香格里拉酒店”。因此需要充分理解上下文语义关系进行识别, 可以使用实体链接、注意力机制、特征融合、图神经网络等方法, 挖掘更详细、更深层次的语义信息, 从而消除命名实体的歧义性^[27]。

(5) 实体嵌套。实体嵌套指实体内部有一个或多个其它实体, 例如: “中国驻俄罗斯大使馆”这一组织机构名中包含了“中国”和“俄罗斯”两个地名。嵌套实体中包含了实体与实体之间丰富的语义关系, 充分利用嵌套实体的嵌套信息, 可以帮助我们更详细、更深层次地理解文本。Xu 等人使用 SVM 和 CNN 抽取中文嵌套实体的语义关系^[30]。Xia 等人使用 MGNER 模型, 重构了命名实体识别的流程, 对嵌套实体进行识别^[31]。

3 迁移学习

随着信息化时代的来临, 传统机器学习以及深度学习已经取得了巨大的成功, 并已经应用到许多实际生活中。但是它们严重依赖于大量具有相同数据分布的标记训练数据, 然而实际应用中, 收集足够的训练数据是非常困难的。半监督学习、无监督学习可以放宽对大量标注数据的需求进而可以解决部分问题, 但是训练的模型性能不尽如人意^[8]。迁移学习是机器学习中解决训练数据不足这一基本问题的重要方法, 旨在利用来自源域的知识提高目标任务学习性能^[8]。它放宽了机器学习中的两个

基本假设: (1) 用于学习的训练样本与新测试样本满足独立同分布条件; (2) 必须有足够可利用的训练样本才能学习得到一个性能不错的模型^[32]。在迁移学习中, 给定源域 D_S (含有大量标注数据, 如英语、新闻领域等) 和源任务 T_S , 目标域 D_T (只有少量或完全没有标注数据, 如蒙古语、社交媒体领域、医学领域等) 和目标任务 T_T , 其中 $D_S \neq D_T$ 或者 $T_S \neq T_T$ 。迁移学习已经广泛地应用于 NLP、计算机视觉等领域, 是当前机器学习中的研究热点。

3.1 传统迁移学习方法

传统迁移学习方法分为基于数据和基于模型。基于数据方法主要使用实例加权和特征转换, 以减小源域样本和目标域样本之间的分布差异。Dai 等人提出了 TrAdaboost^[33], 将 Adaboost 算法扩展到了迁移学习中, 提高可用源域实例权重, 降低不可用源域实例权重。Huang 等人提出核均值匹配法 (Kernel Mean Matching, KMM), 通过再生核希尔伯特空间 (Reproducing Kernel Hilbert Space, RKHS) 中匹配源域和目标域实例之间的均值, 估计源域和目标域概率分布, 使得带权源域和目标域概率分布尽可能相近^[34]。Pan 等人提出迁移成分分析 (Transfer Component Analysis, TCA)^[35], 采用最大均值误差 (Maximum Mean Discrepancy, MMD)^[36]作为度量准则测量边缘分布差异, 以分散矩阵作为约束条件, 将源域和目标域之间的分布差异最小化。

基于模型方法是指利用源域和目标域之间的相似性和相关性, 将已训练好的部分源域模型或特征表示迁移到目标模型上, 以提高目标模型的性能。Duan 等人提出领域自适应机 (Domain Adaptation Machine, DAM) 通用框架, 借助在多个源域上分别预先训练的基本分类器, 为目标域构造一个鲁棒的分类器^[37]。为了解决分类问题, Tommasi 等人提出了一种单模型知识迁移方法 (Single-Model Knowledge Transfer, SMKL), 该方法基于最小二乘 SVM, 从源域选择一个预先获得的二进制决策函数, 然后迁移其参数到目标模型^[38]。Yao 等人等在 TrAdaBoost 的基础上做了多源扩展, 提出了

TaskTrAdaBoost, 首先在每个源域上执行 AdaBoost 构造一组候选分类器, 然后每次迭代挑选出在目标域上具有最低分类误差的候选分类器并为其分配权重, 最后将所选分类器组合产生最终预测^[39]。

3.2 深度迁移学习方法

深度学习目前是机器学习领域最流行的方法, 许多研究者利用深度学习技术构建迁移学习模型, 已经成为解决深度学习数据依赖和训练数据不足等问题的重要方法。深度迁移学习分为两类: 非对抗方法和对抗方法。

非对抗方法复用源域中预先训练好的部分深度神经网络, 将其迁移至目标模型中。Tzeng 等人提出深度域混淆 (Deep Domain Confusion, DDC) 解决深度网络的自适应问题, 在源域与目标域之间添加了一层适应层和 MMD, 让深度迁移网络在学习如何分类的同时减小源域实例和目标域样实例之间的分布差异^[40]。Long 等人对 DDC 进行了扩展, 提出了深度自适应网络 (Deep Adaptation Network, DAN)^[41], 在深度神经网络中加入多层适应层和表征能力更好的 MK-MMD^[42]。Long 等人对 DAN 网络的基础上又提出了联合自适应网络 (Joint Adaptation Network, JAN), 相比于 DAN 只考虑边缘分布自适应, JAN 使用效果更好的多层联合分布自适应^[43]。

对抗方法引入由 GAN 网络启发的对抗技术, 使模型无法识别特征来自源域还是目标域, 进而完成源域知识到目标域的迁移, 同时有效缓解了负迁移问题。Ganin 等人提出了结构简单的领域对抗神经网络 (Domain-adversarial Neural Networks, DANN)^[44], 其由特征提取器、标签预测器和领域分类器组成, 特征提取器的作用类似于 GAN 网络生成器, 其目的是生成“域无关”的特征表示, 领域分类器起着类似于判别器的作用, 试图检测提取的特征是来自源域还是目标域, 通过在领域分类器和特征提取器之间使用对抗技术, 学习一种“域无关特征”。Tzeng 等人提出了一种对抗领域自适应的通用框架 ADDA (Adversarial Discriminative Domain Adaptation)^[45], 利用判别模型、无条件权重共享和 GAN

损失, 解决领域之间的数据分布问题。Zhang 等人提出了一种部分领域自适应的方法, 称为基于重要性加权对抗网络的领域自适应 (Importance Weighted Adversarial Nets-Based Domain Adaptation, IWANDA)^[46], 不再使用一个共享特征提取器, 而是分别为源域和目标域提供特定域的特征提取器。

4 命名实体识别的迁移学习方法

4.1 基于数据迁移学习

NER 的基于数据迁移学习方法大多利用额外高资源语言标注数据作为迁移学习的弱监督训练, 以对齐信息作为桥梁, 如双语词典^[47]、并行语料库^[48]和单词对齐^[49]等, 将知识(标注、词向量、特征表示等)从高资源语言投影到低资源语言。基于数据方法在跨语言 NER 中显示出相当大的优越性, 但是对高资源语言标注数据和对齐信息的规模和质量非常敏感, 并且仅限于跨语言迁移。

4.1.1 标注和表示投影法

为了提高跨语言 NER 的性能以及针对目标语言中没有人工标注, Ni 等人提出两种弱监督跨语言 NER 方法^[48]: 标注和表示投影法, 以及两种共解码方案: 基于排除-O 置信度和基于等级的共解码方案。

标注投影法利用并行语料库、翻译等对齐语料, 将高资源语言中的标注迁移到对应目标语言上, 并开发了一种独立于语言的数据选择方案, 可以从嘈杂的数据中选择高质量标注投影数据。设: 给定目标语言句子 y , 以及质量得分阈值 q 和实体数量阈值 n , 其投影质量得分 $q(y)$, 如公式(1)所示:

$$q(y) = \frac{\sum_{e \in y} \hat{P}(l'(e)|e)}{n(y)} \quad (1)$$

式中 e 代表 y 中的每个实体, $\hat{P}(l'(e)|e)$ 代表 e 用投影标注 $l'(e)$ 标记的相对频率, $n(y)$ 是 y 中的实体总数。数据选择方案必须满足: $q(y) \geq q$; $n(y) \geq n$ 。

表示投影法, 首先使用以词向量为输入的前馈神经网络模型训练英语 NER 系统, 然后将目标语言的词向量通过线性映射 $M_{f \rightarrow e}$ 投影到英语向量空间中; 最后使用训练好的英语 NER 系统对目标语言进行标记。可通过加权最小二乘法得到线性映射

$M_{f \rightarrow e}$, 如公式(2)所示:

$$M_{f \rightarrow e} = \arg \min_M \sum_{i=1}^n w_i \|u_i - Mv_i\|^2 \quad (2)$$

其中 w_i 表示训练词典中英语目标语言单词对 (x_i, y_i) 的权重, u_i 、 v_i 分别表示英语单词 x_i 和目标语言单词 y_i 的词向量。

共解码方案可以有效地结合两种投影法的输出, 提高识别精度。基于排除-O 置信度的共解码方案是选择置信度分数较高的标注投影法或表示投影法生成的标签, 优先选择一种方法的非 O 标签(即实体标签)。基于等级的共解码方案, 给予标注投影法更高优先级, 即组合输出包括标注投影法检测到的所有实体, 以及与标注投影法不冲突的所有表示投影法检测到的实体。当标注投影法为一段 x 都生成了 O 标签, 则表示投影法检测到 x 的实体标签不会与标注投影冲突。例如: 标注投影法的输出标签序列为 (B-PER, O, O, O, O), 表示投影法的输出标签序列为 (B-ORG, I-ORG, O, B-LOC, I-LOC), 那么基于等级的共解码方案合并输出为 (B-PER, O, O, B-LOC, I-LOC)。

Ni 的贡献在于为标注投影法开发了一种语言无关数据选择方案, 以及两种共解码方案, 有效地提高了 NER 的识别精度。两种投影法都具有较高灵活性, 但是容易受到双语单词对的对齐准确率和英语 NER 系统准确率的影响。

4.1.2 双语词典特征表示迁移法

为了丰富低资源语言的语义表示以及缓解词典外单词问题, Feng 等人提出了双语词典特征表示迁移法, 将双语词典特征表示和词级实体类型分布特征作为目标 NER 模型的额外输入, 并设计一个词典扩展策略估计词典外单词的特征表示^[50]。

双语词典特征表示: 根据来自高资源语言的翻译, 对每个低资源语言单词的所有翻译词向量使用 BiLSTM 或注意力机制提取特征表示 vec_i 。每个翻译项目 T 都由一个或多个高资源语言单词组成, 例如: 中文单词“美国”有四个英文翻译“America”、“UnitedStates”、“USA”和“TheUnitedStates of

America”。

词典扩展策略, 用于估计词典外单词的双语词典特征表示。给定低资源语言单词 x_i 及其对应词向量 w_i 和双语词典特征表示 vec_i 。使用线性映射函数, 如公式(3)所示, 作为两个语义空间之间的转换, 最小化公式(4)以优化映射矩阵 M , 在获得 M 之后, 对每个词典外单词 o_i 用公式(5)估算其特征表示 veo_i :

$$vec_i = Mw_i \quad (3)$$

$$loss_M = \sum_{i=1}^f \|vec_i - Mw_i\|_2 \quad (4)$$

$$veo_i = Mo_i \quad (5)$$

单词实体类型的分布特征是每个单词被标记为每种实体类型的概率。实验中只使用了三种最常见的命名实体类型, 即 P(人名)、L(地名)、O(组织名)以及随机生成一个表示非实体的类型 N, 因此构造了四个实体类型向量 $\{E_P, E_O, E_L, E_N\}$, $E_j \in R^d$ 。然后, 使用标准余弦函数计算低资源词向量 w_i 与实体类型向量 E_j 之间的语义相关性, 如公式(6)所示, 最后, 每个低资源和高资源语言单词都分配有一个维数为 4 的实体类型分布特征, $e_{ij} = \{e_P, e_O, e_L, e_N\}$ 。

$$e_{ij} = \frac{w_i^T \cdot E_j}{\|w_i\| \times \|E_j\|} \quad (6)$$

最后, 将低资源词向量 w_i 、低资源字符向量 c_i 、双语词典翻译特征 vec_i 或 veo_i 以及实体类型分布特征 e_{ij} 的连接词向量 $W_i = \{w_i, c_i, vec_i, e_{ij}\}$ 作为 BiLSTM-CRF 模型的输入。

该方法开创性地使用双语词典特征表示和单词实体类型的分布特征表示, 丰富了低资源语言的语义表示, 并设计了一种词典扩展策略有效地缓解了词典外单词问题, 在低资源 NER 性能上取得很大的提升。该方法具有非常好的可扩展性, 可以将高资源语言的其它知识(例如: WordNet、知识图谱等)整合到体系结构中, 还可以扩展到其它 NLP 任务(例如: 意图识别、情感分析)。

4.2 基于模型迁移学习

NER 的基于模型迁移学习不需要额外的高资源语言对齐信息, 主要利用源域和目标域之间的相

似性和相关性,将源模型部分参数或特征表示迁移到目标模型,并自适应地调整目标模型^[51]。例如:Ando 和 Zhang^[52]提出了一种迁移学习框架,该框架在多个任务之间共享结构参数,并提高了包括 NER 在内多种任务的性能。Collobert 等人^[2]提出一个独立于任务的卷积神经网络,并采用联合训练将知识从 NER 和 POS 任务迁移到组块识别任务。Wang 等人^[53]利用标签感知 MMD 完成特征迁移,实现了跨医学专业 NER 系统。Lin 等人^[54]在现有的深度迁移神经网络结构上引入单词和句子适应层,弥合两个输入空间之间的间隙,在 LSTM 和 CRF 层之间也引入了输出适应层,以捕获两个输出空间中的变化。考虑到目标数据的领域相关性差异, Yang 等人^[55]受知识蒸馏(Knowledge distillation, KD)的启发,提出了一种用于序列标记领域自适应的细粒度知识融合模型,首先对目标域句子和单词的领域相关性进行建模,然后在句子和单词级别上对源域和目标域进行知识融合,有效平衡了目标模型从目标域数据学习和从源模型学习之间的关系。

4.2.1 基于 RNN 的迁移学习

RNN 及其变体已被大量应用于 NER 任务,并取得了非常高的识别精度。Yang 等人利用神经网络通用性,提出了一种基于 RNN 的序列标注迁移学习框架(RNN-TL)^[51],通过源任务和目标任务之间共享模型参数和特征表示,提高目标任务的学习性能。并利用不同级别的共享方案,在一个统一的模型框架下处理跨域、跨应用和跨语言迁移。

该方法开发的三种级别共享方案框架分别为 T-A、T-B 和 T-C,如图 2 所示。T-A 共享所有模型参数,最后在源任务和目标任务的 CRF 层上执行一个标签映射,应用于两个具有相同标签集领域的跨域迁移。T-B 共享 CRF 层前的所有模型参数,即每个任务单独训练一个 CRF,用于两个具有不同标签集领域的跨域迁移和跨应用迁移。T-C 只共享字符向量和字符特征表示,用于跨语言迁移,侧重于字母相似的语言(如:英语和西班牙语)。该方法的共享参数由两个任务共同优化,源任务和目标任务具有不同的收敛速度,所以提前停止目标任务。

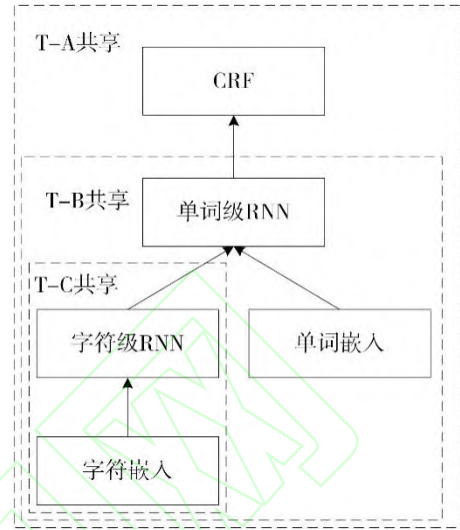


Fig.2 RNN-based transfer learning model

图 2 基于 RNN 的迁移学习模型

该迁移方法在一个统一的 RNN-CRF 框架下,在低资源跨域、跨应用和跨语言的序列标注任务上取得了不错的效果,尤其是在跨领域方面。但是还存在一定不足:跨语言迁移只能是字母相似的语言;对于迁移的参数和特征表示没有进行任何筛选工作,这使得负迁移对模型性能产生消极影响。

4.2.2 参数和神经适配器迁移

NER 的实体类型随时间在不断变化,为了解决目标领域出现新的实体类型而导致重新标注数据和训练模型的问题,Chen 等人^[56]提出了一种解决方案:在目标模型的输出层添加新的神经元并迁移源模型部分参数,然后在目标数据上进行微调(Fine-tuned),此外还设计一种神经适配器学习源数据和目标数据之间的标签分布差异,迁移过程如图 3 所示。

在目标模型输出层扩展 nM 个神经元用于学习新实体类型,其中 n 取决于数据集标签格式(例如:如果数据集为 BIO 格式,则 $n=2$,因为对于每个命名实体类型,将有两种输出标签 B-NE 和 I-NE), M 是新命名实体类型的数量。迁移源模型参数时,目标模型输出层参数用正态分布 $X \sim N(\mu, \sigma^2)$ 得出的权重进行初始化;其它参数都用源模型中相对应参数进行初始化。

神经适配器使用 BiLSTM 实现,将源模型输出

层输出连接到目标模型相应输出上,作为目标 CRF 的附加输入。可以为目标模型学习两个任务之间的标签分布差异,以减少数据标签不一致的影响。

该方案使用源模型参数和神经适配器实现模型迁移,是一种非常简单的迁移方法,解决了目标领域出现新实体类型而导致重新标注数据和训练模型的问题。同时神经适配器可以解决标签不一致,且具有提高迁移模型性能的能力。

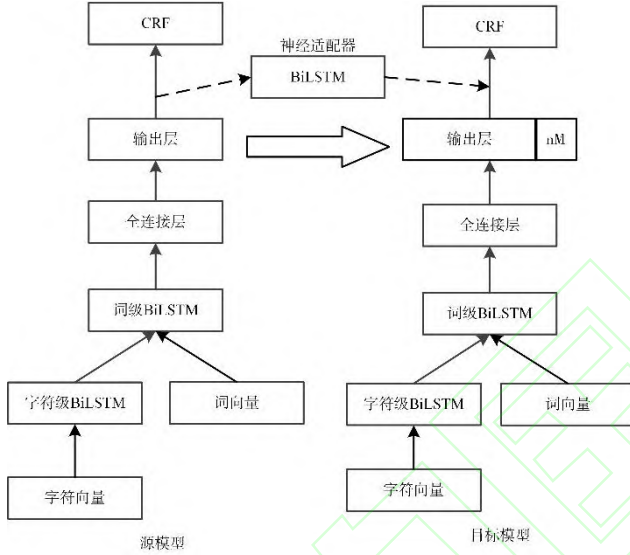


Fig.3 Source model parameters and neural adapter model

图 3 源模型参数和神经适配器模型

4.3 对抗迁移学习

NER 的基于模型迁移学习方法虽然取得了很好的性能,但是还存在以下问题有待解决:(1)没有考虑资源间的差异,强制在语言或领域之间共享模型参数和特征表示;(2)资源数据不平衡,高资源语言或领域的训练集规模通常比低资源训练集规模大得多^[57],忽略了领域间的这些差异,导致泛化能力差。所以研究者引入受 GAN 网络启发的对抗技术,学习一种“域无关特征”,实现源域知识到目标域的迁移,同时有效缓解了负迁移问题。

NER 的对抗迁移学习流程如图 4 所示。对抗鉴别器选择有利于提高目标任务性能的源任务特征,同时防止源任务的特定信息进入共享空间。训练完成之后,对抗鉴别器和共享特征提取器达到平衡:

对抗鉴别器无法区分共享特征提取器中的特征表示来自源域还是目标域。但是训练达到这个平衡点需要花费大量时间,还有可能发生模型崩溃。

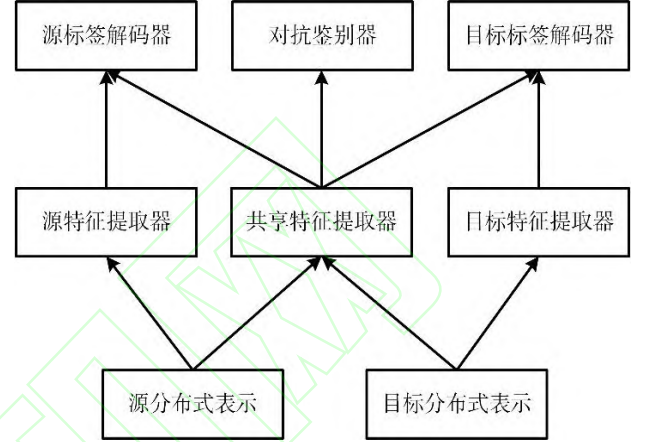


Fig.4 NER adversarial transfer learning flowchart
图 4 NER 对抗迁移学习流程图

4.3.1 自注意力机制对抗迁移网络

Cao 等人首次将对抗迁移学习应用于 NER 任务,提出自注意力机制的中文 NER 对抗迁移学习模型^[58]。充分利用中文分词(Chinese Word Segmentation, CWS)任务更加丰富的词边界信息,并通过任务鉴别器和对抗损失函数过滤中文分词任务的特有信息,以提高中文 NER 任务性能。同时在 BiLSTM 层后加入自注意力机制明确捕获两个字符之间的长距离依赖关系并学习句子内部结构信息。

任务鉴别器,通过 Max-pooling 层和 softmax 层识别特征来自哪个领域,可以表示为公式(7)(8):

$$S = \text{Maxpooling}(H) \quad (7)$$

$$D(S; \theta_d) = \text{softmax}(W_d S + b_d) \quad (8)$$

其中 H 表示共享自注意力的输出, θ_d 表示任务鉴别器的参数, $W_d \in R^{K \times 2d_h}$ 和 $b_d \in R^K$ 是可训练参数, K 是任务数。

通过引入对抗损失函数 L_{Adv} , 如公式(9)所示,防止中文分词任务的特定信息进入共享空间。 L_{Adv} 训练共享特征提取器以产生共享特征,使得任务鉴别器无法可靠地判断特征的领域。通过在 softmax 层下方添加一个梯度反转层完成 minimax 优化,使共享 BiLSTM 生成一个特征表示来误导任务鉴别器。

$$L_{Adv} = \min_{\theta_s} \left(\max_{\theta_d} \sum_{k=1}^K \sum_{i=1}^{T_k} \log D(E_s(x_k^{(i)})) \right) \quad (9)$$

式中 θ_s 表示共享特征提取器可训练参数, E_s 表示共享特征提取器, T_k 是任务 k 训练实例的数量, $x_k^{(i)}$ 是任务 k 的第 i 个实例。

该模型首次将对抗迁移学习应用于NER任务, 在WeiboNER数据集^[59]和SighanNER数据集^[60]上的F1值分别达到53.08%和90.64%, 比BiLSTM-CRF模型提升了4.67%和1.43%, 并通过实验验证了迁移学习、对抗训练、自注意力机制各个方法对于模型的有效性。

4.3.2 双重对抗迁移网络

Zhou等人提出了双重对抗迁移网络(Dual Adversarial Transfer Network, DATNet)^[57], 在通用深度迁移单元上引入两种对抗学习: 一是用广义资源对抗鉴别器(Generalized Resource-Adversarial Discriminator, GRAD), 解决资源数据不均衡和资源差异问题; 二是对抗训练, 分别在字符向量和词向量层添加以一个小范数 ϵ 为界的扰动, 以提高模型的泛化能力和鲁棒性。

DATNet根据特征提取器的差异, 有两种体系结构: 一是特征提取器有共享BiLSTM和资源相关BiLSTM的DATNet-F; 二是特征提取器只有共享BiLSTM而没有资源相关BiLSTM的DATNet-P。

GRAD通过权重 α 平衡高低资源的训练规模差异较大的影响, 使源域和目标域中提取的特征表示更加兼容, 共享BiLSTM的输出与领域无关, 并为每个样本提供自适应权重, 从而使模型训练的重点放在困难样本上。为了计算GRAD的损失函数, 如公式(10)所示, 共享BiLSTM的输出序列首先通过自注意力机制编码为单个向量, 然后通过线性变换投影到标量 r :

$$l_{GRAD} = - \sum_i \{ I_{i \in D_s} \alpha (1 - r_i)^\gamma \log r_i + I_{i \in D_T} (1 - \alpha) r_i^\gamma \log (1 - r_i) \} \quad (10)$$

式中 $I_{i \in D_s}$ 和 $I_{i \in D_T}$ 是标识函数, 分别表示特征来自源域还是目标域; 参数 γ 衡量困难和简单样本损失贡

献对比, $(1 - r_i)^\gamma$ (或 r_i^γ) 通过测量预测值与真实标签之间的差异控制各个样本的损失贡献。权重 α 和 $(1 - r_i)^\gamma$ (或 r_i^γ) 分别减少了高资源样本和简单样本的损失贡献。

对抗训练就是在原始样本的基础上添加以一个小范数 ϵ 为界的扰动 η_x , 计算如公式(11)所示:

$$\eta_x = \arg \max_{\eta: \|\eta\|_2 \leq \epsilon} \log p(y|\theta; x + \eta) \quad (11)$$

其中 θ 是当前模型参数集。 η_x 按照文献[61]中的策略, 通过如下线性化方法近似估算: $\eta_x = \epsilon \frac{g}{\|g\|_2}$, $g = \nabla_x \log p(y|\theta; x)$, ϵ 可在验证集上确定。在每个训练步骤中, 由 θ 参数化当前模型找到的扰动 η_x , 并通过 $x_{adv} = x + \eta_x$ 构造一个对抗样本, 然后进行原始样本和对抗样本混合训练以提高模型泛化能力。对抗训练损失函数 l_{AT} 的计算如公式(12)所示:

$$l_{AT} = \log p(y|\theta; x) + \log p(y|\theta; x_{adv}) \quad (12)$$

其中 $\log p(y|\theta; x)$, $\log p(y|\theta; x_{adv})$ 分别表示原始样本及其对抗样本的损失。DATNet分别在字符级和词级向量层使用对抗训练, 可以根据公式(11)计算字符向量扰动 η_c 、源域词向量扰动 η_{w_s} 和目标域词向量扰动 η_{w_T} 。

DATNet很好地解决了表示差异和数据资源不平衡的问题, 提高了模型的泛化能力, 并在跨语言和跨域NER迁移上取得显著改进。通过实验, DATNet-P架构更适合具有相对更多训练数据的跨语言迁移, 而DATNet-F架构更适合具有极低资源和跨域迁移的跨语言迁移。

5 评价指标及NER方法比较

5.1 命名实体识别评价指标

目前, NER最常用的评价标准有精确率(Precision), 召回率(Recall)和F1值(F1-score)等。

精确率: 在给定数据集中, 标注正确实体数占所有被标注实体数的比例, 如公式(13)所示:

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

召回率: 在给定数据集中, 标注正确实体数占数据集中所有实体数的比例, 如公式(14)所示:

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

F1 值: 同时考虑精确率和召回率, 是平衡精确率和召回率的综合指标, 如公式(15)所示:

$$F1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (15)$$

其中, TP (True Positive) 表示真阳性, 识别出的正确实体数; FP (False Positive) 表示假阳性, 识别出的错误实体数; FN (False Negative) 表示假阴性, 未被识别出实体数。

5.2 迁移学习评价指标

为了验证迁移学习模型的性能, 通常是在同一

数据集下, 将实验模型与深度神经网络模型、其它迁移学习模型的精确率、F1 值等进行比较。

5.3 NER 方法性能比较

为了评估 NER 迁移学习方法的性能, 以 CoNLL2003 英语 NER 数据集^[62]为源域, CoNLL2002 西班牙语和荷兰语 NER 数据集^[63]以及 WNUT-2017 英语 Twitter NER 数据集^[64]为目标域, 进行实验。这些数据集的统计信息, 如表 1 所示, 使用官方的训练集、验证集和测试集的划分方法。实验中使用 30 维字符向量, 50 维词向量, LSTM 隐状态的数量设置为 100 维。

Table 1 Statistics of NER Datasets

表 1 NER 数据集统计信息

数据集	语言	训练集 (实体数)	验证集 (实体数)	测试集 (实体数)
CoNLL2003	English	20,4567(23,499)	51,578 (5,942)	46,666 (5,648)
CoNLL2002	Spanish	207,484 (18,797)	51,645 (4,351)	52,098 (3,558)
CoNLL2002	Dutch	202,931 (13,344)	37,761 (2,616)	68,994 (3,941)
WNUT-2017	English	62,730 (3,160)	15,733 (1,250)	23,394 (1,740)

表 2 是一些 NER 方法的 F1 值比较, 可以看出双语词典特征表示在跨语言迁移方面取得了非常好的效果, 相比于 LSTM-CRF 在西班牙语和荷兰语上分别提升了 6.65%和 3.01%, 由于领域之间没有标准的对齐信息, 该方法无法进行跨领域迁移; DATNet 的两个变体优于其它方法, 三个数据集上都有很大提升, DATNet-P 更适合跨域迁移, 而 DATNet-P 在跨语言上更有优势。与多任务学习^[65]相比, 迁移学习最关心的是目标任务, 而不是同时提升所有的源任务和目标任务。

Table 2 NER method performance comparison (F1-score)

表 2 NER 方法性能比较 (F1 值)

Model	Spanish	Dutch	Twitter
LSTM-CRF	81.74	83.41	40.42
双语词典特征表示	86.42	88.39	-
RNN-TL	85.77	85.19	43.24
多任务学习	85.88	86.55	45.62
DATNet-P	87.04	87.77	53.43
DATNet-F	88.16	88.32	50.85

6 总结与展望

本文主要对应用于 NER 任务的迁移学习方法从基于数据、基于模型和对抗迁移学习三个方面进行了归纳总结。迁移学习对于解决 NER 任务的资源匮乏、实体类型多样化等问题, 取得了非常好的效果。基于数据迁移学习在跨语言迁移任务中取得很大成功, 但也仅限于跨语言迁移。基于模型迁移学习不需要额外的高资源语言表示, 将源模型的部分参数和特征迁移到目标模型上。对抗迁移学习以其独特的对抗训练思想, 生成一种“域无关特征”, 实现源域知识到目标域的迁移, 帮助目标任务提高学习性能, 同时有效缓解了负迁移问题, 是目前发展潜力最大的 NER 迁移学习方法, 也是今后的研究重点。

随着对 NER 迁移学习的深入研究, 还有一些新的问题需要解决:

(1) 负迁移是迁移学习道路上最大的阻碍, 虽然寻找源域和目标域之间相关性衡量标准以及对抗迁移学习可以缓解该问题, 但是这些方法都有

其自身的局限性。如何更好地解决负迁移问题，有待进一步深入研究。

(2) 对于多步传导式迁移学习，如何寻找一个或几个既能照顾到目标域也能照顾到源域的中间领域，帮助相关性不大的两个领域之间实现迁移学习，以达到充分利用已有的大量数据，是迁移学习未来的一个研究方向。

(3) 在 NER 对抗迁移学习中如何构建更加强大的对抗鉴别器，帮助共享特征提取器和对抗鉴别器之间更快达到平衡点，处理多源域 NER 对抗迁移学习任务，以及更好地解决负迁移问题，是下一步研究的主要工作。

References:

- [1] Yadav V, Bethard S. A survey on recent advances in named entity recognition from deep learning models[C]//COLING2018:2145-2158
- [2] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch[J]. Journal of Machine Learning Research, 2011, 12(Aug): 2493-2537.
- [3] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging[J]. ArXiv Preprint ArXiv:1508.01991, 2015.
- [4] Lample G, Ballesteros M, Subramanian S, et al. Neural architectures for named entity recognition[C]//HLT-NAACL, 2016:260-270.
- [5] Chiu J P C, Nichols E. Named entity recognition with bi-directional LSTM-CNNs[C]//Transactions of the Association for Computational Linguistics, 2016, 4: 357-370.
- [6] Ma X, Hovy E. End-to-end sequence labeling via bi-directional lstm-cnns-crf[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016.
- [7] Zhang B, Pan X, Wang T, et al. Name tagging for low-resource incident languages based on expectation-driven learning[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016: 249-259.
- [8] Zhuang F Z, Qi Z, Duan K, et al. A Comprehensive Survey on Transfer Learning[J]. ArXiv Preprint ArXiv:1911.02685, 2019.
- [9] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems, 2014: 2672-2680.
- [10] Liu Liu, Wang Dongb. A Review on Named Entity Recognition[J]. Journal of the China Society for Scientific and Technical Information, Mar. 2018, 37(3): 329-340.
- [11] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 22(10): 1345-1359.
- [12] Weiss K, Khoshgoftaar T M, Wang D D. A survey of transfer learning[J]. Journal of Big Data, 2016, 3(1): 9.
- [13] Nadeau D, Sekine S. A survey of named entity recognition and classification [J]. Lingvisticae Investigationes, 2007, 30(1):3-26.
- [14] Li J, Sun A, Joty S R. SegBot: A Generic Neural Text Segmentation Model with Pointer Network[C]//IJCAI. 2018: 4166-4172.
- [15] Humphreys K, Gaizauskas R, Azzam S, et al. University of Sheffield: Description of the LaSIE-II system as used for MUC-7[C]//Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax, Virginia, April 29-May 1, 1998. 1998.
- [16] Krupka G R, IsoQuest K H. Description of the nerowl extractor system as used for muc-7[C]//Proceedings of the 7th Message Understanding Conference, Virginia. 2005: 21-28.
- [17] Black W J, Rinaldi F, Mowatt D. FACILE: Description of the NE System Used for MUC-7[C]//Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax, Virginia, April 29-May 1, 1998.
- [18] Eddy S R. Hidden markov models[J]. Current Opinion in Structural Biology, 1996, 6(3): 361-365.
- [19] Quinlan J R. Induction of decision trees[J]. Machine Learning, 1986, 1(1): 81-106.
- [20] Kapur J N. Maximum-entropy models in science and engineering[M]. John Wiley & Sons, 1989.
- [21] Shawe-Taylor J, Cristianini N. Support vector machines[J]. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, 2000: 93-112.
- [22] Lafferty J, McCallum A, Pereira F C N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[C]// ICML, 2001:282-289.

- [23] Strubell E, Verga P, Belanger D, et al. Fast and accurate entity recognition with iterated dilated convolutions[C]//EMNLP,2017: 2670-2680.
- [24] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. 2017: 5998-6008.
- [25] Shen Y, Yun H, Lipton Z C, et al. Deep Active Learning for Named Entity Recognition[C]//Meeting of the Association for Computational Linguistics, 2017: 252-256.
- [26] Yang Z, Salakhutdinov R, Cohen W. Multi-task cross-lingual sequence tagging from scratch[J]. ArXiv Preprint ArXiv:1603.06270, 2016.
- [27] Chen S D, Ouyang X Y. Overview of Named Entity Recognition Technology[J]. Radio Communications Technology, 2020,46(3) : 251-260.
- [28] Li G,Huang Y F. An approach to named entity recognition towards micro-blog[J]. Application of Electronic Technique , 2018, 44(1): 118-120, 124.
- [29] Sheng J. Transfer learning in named entity recognition[D]. Harbin Institute of Technology, 2019.
- [30] Xu H L, Li Y Q, He Y Q, et al. Research on Chinese Nested Named Entity Relation Extraction [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, Vol. 55, No. 1: 8-14
- [31] Xia C, Zhang C, Yang T, et al. Multi-Grained Named Entity Recognition[C]//Proceedings of the 57th Conference of the Association for Computational Linguistics, 2019, 1430-1440.
- [32] Zhuang F Z, Luo P, He Q, Shi Z Z. Survey on transfer learning research[J]. Ruan Jian XueBao/Journal of Software, 2015, 26(1): 26-39.
- [33] Dai W, Yang Q, Xue G, et al. Boosting for transfer learning[C]//International Conference on Machine Learning, 2007: 193-200.
- [34] Huang J, Gretton A, Borgwardt K M, et al. Correcting Sample Selection Bias by Unlabeled Data[C]//Neural Information Processing Systems, 2006: 601-608.
- [35] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis[C]//International Joint Conference on Artificial Intelligence, 2009: 1187-1192.
- [36] Borgwardt K M, Gretton A, Rasch M J, et al. Integrating structured biological data by Kernel Maximum Mean Discrepancy[J]. Bioinformatics, 2006, 22(14): 49-57.
- [37] Duan L, Xu D, Tsang I W, et al. Domain Adaptation From Multiple Sources: A Domain-Dependent Regularization Approach[J]. IEEE Transactions on Neural Networks, 2012, 23(3): 504-518.
- [38] Tommasi T, Caputo B. The more you know, the less you learn: from knowledge transfer to one-shot learning of object categories[C]//British Machine Vision Conference, 2009: 1-11.
- [39] Yao Y, Doretto G. Boosting for transfer learning with multiple sources[C]//Computer Vision and Pattern Recognition, 2010: 1855-1862.
- [40] Tzeng E, Hoffman J, Zhang Ning, et al. Deep domain confusion: maximizing for domain invariance[J]. Computer Science, 2014.
- [41] Long M, Cao Y, Wang J, et al. Learning Transferable Features with Deep Adaptation Networks[C]// International Conference on Machine Learning, 2015: 97-105.
- [42] Gretton A, Borgwardt K M, Rasch M J, et al. A kernel two-sample test[J]. Journal of Machine Learning Research, 2012, 13(Mar): 723-773.
- [43] Long M, Zhu H, Wang J, et al. Deep transfer learning with joint adaptation networks[C]//Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017: 2208-2217.
- [44] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks[M]//Domain Adaptation in Computer Vision Applications. Springer, Cham, 2017: 189-209.
- [45] Tzeng E, Hoffman J, Saenko K, et al. Adversarial Discriminative Domain Adaptation[C]// Computer Vision and Pattern Recognition, 2017: 2962-2971.
- [46] Zhang J, Ding Z, Li W, et al. Importance Weighted Adversarial Nets for Partial Domain Adaptation[C]// Computer Vision and Pattern Recognition, 2018: 8156-8164.
- [47] Zirikly A, Hagiwara M. Cross-lingual transfer of named entity recognizers without parallel corpora[C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). 2015: 390-396.
- [48] Ni J, Dinu G, Florian R, et al. Weakly Supervised Cross-Lingual Named Entity Recognition via Effective Annotation and Representation Projection[C]//Meeting of

- the Association for Computational Linguistics, 2017: 1470-1480.
- [49] Yarowsky D, Ngai G, Wicentowski R. Inducing multilingual text analysis tools via robust projection across aligned corpora[C]//Proceedings of the First International Conference on Human Language Technology Research. Association for Computational Linguistics, 2001: 1-8.
- [50] Feng X, Feng X, Qin B, et al. Improving Low Resource Named Entity Recognition using Cross-lingual Knowledge Transfer[C]//IJCAI. 2018: 4071-4077.
- [51] Yang Z, Salakhutdinov R, Cohen W W, et al. Transfer Learning for Sequence Tagging with Hierarchical Recurrent Networks[C]//International Conference on Learning Representations, 2017.
- [52] Ando R K, Zhang T. A framework for learning predictive structures from multiple tasks and unlabeled data[J]. Journal of Machine Learning Research, 2005, 6(Nov): 1817-1853.
- [53] Wang Z, Qu Y, Shen J, et al. LABEL-AWARE DOUBLE TRANSFER LEARNING FOR CROSS SPECIALTY MEDICAL NAMED ENTITY RECOGNITION[C]//North American Chapter of the Association for Computational Linguistics, 2018: 1-15.
- [54] Lin B Y, Lu W. Neural Adaptation Layers for Cross-domain Named Entity Recognition[C]//Empirical Methods in Natural Language Processing, 2018: 2012-2022.
- [55] Yang H, Huang S, Dai X, et al. Fine-grained Knowledge Fusion for Sequence Labeling Domain Adaptation[C]//EMNLP/IJNLP(1), 2019:4195-4204.
- [56] Chen L, Moschitti A. Transfer Learning for Sequence Labeling Using Source Model and Target Data[J]. ArXiv Preprint ArXiv:1902.05309, 2019.
- [57] Zhou J T, Zhang H, Jin D, et al. Dual adversarial neural transfer for low-resource named entity recognition[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 3461-3471.
- [58] Cao P, Chen Y, Liu K, et al. Adversarial Transfer Learning for Chinese Named Entity Recognition with Self-Attention Mechanism[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018: 182-192.
- [59] Peng N, Dredze M. Named entity recognition for chinese social media with jointly trained embeddings[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015: 548-554.
- [60] Levow G A. The third international Chinese language processing bakeoff: Word segmentation and named entity recognition[C]//Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing. 2006: 108-117.
- [61] Goodfellow I, Shlens J, Szegedy C, et al. Explaining and Harnessing Adversarial Examples[C]//International Conference on Learning Representations, 2015.
- [62] Yang P, Liu W, Yang J. Positive Unlabeled Learning Via Wrapper-based Adaptive Sampling[C]//IJCAI, 2017: 3273-3279.
- [63] Sang E F, De Meulder F. Introduction to the CoNLL-2003 shared task: language-independent named entity recognition[C]//North American Chapter of the Association for Computational Linguistics, 2003: 142-147.
- [64] Zeman D, Popel M, Straka M, et al. CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies[C]//Conference on Computational Natural Language Learning, 2017, 1(1): 1-19.
- [65] Lin Y, Yang S, Stoyanov V S, et al. A Multi-lingual Multi-task Architecture for Low-resource Sequence Labeling[C]//Meeting of the Association for Computational Linguistics, 2018: 799-809.
- 附中文参考文献:
- [10] 刘浏,王东波.命名实体识别研究综述[J]. 情报学报, 2018, 37(3): 329-340.
- [27] 陈曙东,欧阳小叶.命名实体识别技术综述[J].无线电通信技术,2020,46(3):251-260.
- [28] 李刚,黄永峰.一种面向微博文本的命名实体识别方法[J].电子技术应用,2018,44(1):118-120+124.
- [30] 许浩亮,李雁群,何云琪,钱龙华.中文嵌套命名实体关系抽取研究[J]. 北京大学学报(自然科学版),2019,55(1):8-14.
- [32] 庄福振,罗平,何清,史忠植.迁移学习研究进展[J]. 软件学报,2015,26(1):26-39.



LI Meng was born in 1993. He is an M.S. candidate at Inner Mongolia Normal University. His research interests include natural language processing, spoken language understanding.

李猛 (1993—), 男, 河北张家口人, 内蒙古师范大学硕士研究生, 主要研究领域为自然语言处理, 口语理解。



LI Yanling was born in 1978. She received the Ph.D. degree from Institute of Acoustics, Chinese Academy of Sciences. Now she is an associate professor at College of Computer Science and Technology, Inner Mongolia Normal University, and the member of CCF. Her research interests include natural language processing, spoken language understanding and machine learning, etc. She published more than ten papers in journals and conferences, and presided several projects includes the National Natural Science Foundation of China, the Natural Science Foundation of Inner Mongolia, the Scientific Researches of Higher Education Institution of Inner Mongolia Autonomous Region and the School Fund of Inner Mongolia Normal University.

李艳玲 (1978—), 女, 内蒙古呼和浩特人, 中国科学院声学研究所博士, 内蒙古师范大学计算机科学技术学院副教授, CCF 会员, 主要研究领域为自然语言处理, 口语理解, 机器学习等。发表学术论文 10 余篇, 主持国家自然科学基金 2 项、内蒙古自然科学基金 1 项、内蒙古自治区高等学校科学研究项目 1 项、内蒙古师范大学校级项目 3 项。



LIN Min was born in 1969. He received the Ph.D. degree from Beijing University of Technology. Now he is an professor at College of Computer Science and Technology, Inner Mongolia Normal University. His research interests include natural language processing, text mining, etc. He published more than ten papers in journals and conferences, and presided several projects includes the National Natural Science Foundation of China, the Special Project of Inner Mongolia Ethnic Affairs and the Natural Science Foundation of Inner Mongolia.

林民 (1969—), 男, 内蒙古呼和浩特人, 北京工业大学博士, 内蒙古师范大学计算机科学技术学院教授, 主要研究领域为自然语言处理, 文本挖掘等。发表学术论文 10 余篇, 主持国家自然科学基金 2 项、内蒙古民委蒙古文信息化专项扶持项目 1 项、内蒙古自然科学基金 3 项。