一种基于特征迁移的跨领域中文分词模型

张韬政，张家健
(中国传媒大学 信息与通信工程学院，北京 100024)

摘要：中文分词是自然语言处理的常见任务之一。在跨领域分词任务中，目标领域的数据分布不同及数据量不足通常导致分词效果急剧下降。基于该问题，本文通过引入了迁移学习、对抗学习和正交约束以减轻共享和私有特征之间的干扰,提出了一种基于特征迁移的跨领域中文分词模型，能够在跨领域和小数据量条件下，借鉴数据量较大的源领域的知识来进行学习，实验证明该模型最终获得了出色的表现。
关键词：迁移学习；对抗学习；正交约束；中文分词
中图分类号：TP391.1　　**文献标识码：**A

A cross-domain Chinese word segmentation model based on feature transfer

ZHANG Tao-zheng, ZHANG Jia-jian

(School of information and communication engineering, Communication University of China, Beijing, 100024, China)

Absrtact: Chinese word segmentation is one of the common tasks in natural language processing. In cross-domain Chinese word segmentation tasks, the different distributions between two different domains and the lack of enough training data often result the low performance. For this problem, we propose a cross-domain Chinese word segmentation model based on feature transfer, which introduces transfer learning, adversarial learning and orthogonal constraints to reduce the interferences between shared and private features. This model can learn from the knowledge of source domain with large amount of data under the premise of small amount of data and cross-domain. Experimental results show that the scheme achieves excellent performance.
Key words: transfer learning; adversarial learning; orthogonal constraints; Chinese word segmentation

1 引言

中文分词一直被视为是其他中文自然语言处理任务的前提，由于汉语不像英语一样在书写过程中天生带有空格来表明词与词之间的分界，所以许多自然语言处理技术在中文领域中不能正常运行，由此，如何将中文进行快速、准确的分词成为了大家共同所关注的问题。

2003 年，隐马尔可夫模型 [1](Hidden Markov Model, HMM)和最大熵模型 [2](Maximum Entropy Model, ME)是主要的分词方法。2004 年，Peng[3]提出了将条件随机场(Conditional Random Fields, CRF)和分词相结合，解决了输出关联之间的问题。近几年人工神经网络飞速发展，深度学习也开始应用于解决中文分词问题，学者们开始使用循环神经网络

基金项目：中国传媒大学中央高校基本科研业务费专项资金资助(3132018XNG1829)
作者简介：张韬政（1982-），女（汉族），山西太原人，中国传媒大学副教授，博士，主要从事机器学习和自然语言处理研究，zhangtaozheng@cuc.edu.cn；张家健（1997-），男（汉族），辽宁沈阳人，中国传媒大学硕士研究生，zhangjiajian@cuc.edu.cn.

解决句子的长期依赖问题，Chen[4]将长短期记忆网络(Long Short-Term Memory Networks, LSTM)引入到中文分词，使分词的准确率有了很大突破。在之后的几年里，门控循环单元和 LSTM 网络等及其双向结构，结合 CRF，便成了中文分词的经典方案，一直沿用至今。2019 年，谷歌 [5]发布并开源了一种新的语言表征模型：BERT，即来自 Transformer 的双向编码器表征，其性能使自然语言处理领域的各个任务都有突破性的进步。至此，中文分词在封闭数据范围内的成绩已经十分可观。

然而，随着互联网时代飞速发展，各种新词层出不穷，旧词新用的现象屡见不鲜，这对中文分词又提出了新的挑战。同其他序列标注类任务一样，中文分词对于专业性强的法律、新闻、医疗、文学作品等特有领域来说，未登录词很可能出现激增的可能；在微博等互联网平台上，一方面，每时每刻都有新词新意不断诞生，另一方面，这些文本的句式、格式、语言习惯也和其他领域有着很大的不同；传统的语料库很难对各个领域进行完整的覆盖，因此对于特定领域分词的表现总是不尽如人意。同时，在特定领域中的分词任务往往伴随着数据量不足的情况，而拥有足够数据量的开放性语料又很难针对特定领域有很高的提升，盲目的合并训练只会产生负迁移，对于深度学习来说，没有足够的数据很难支撑起庞大的参数迭代，模型在训练过程中退化严重。

基于这个背景，成于思等人 [6]使用源域数据对模型的参数进行预训练，并使用目标域数据进行微调，但该类方法容易使模型产生过拟合，并且模型仍然使用小数据量进行训练，结果鲁棒性不强。武惠等人 [7]通过文本相似性从源域中筛选出和目标域相似的样本进行有针对性的迁移学习，但这类方法需要两个领域的数据分布有交集，并且高度依赖衡量文本相似性的算法的精度。Guo[8]提出了共享参数层的迁移方法，但是该模型只会盲目地对两个领域的数据进行学习，并不对提取的特征加以区分，学习效率较低。因此，总结前人的工作，我们采取了一种迁移和对抗的方法来解决特定领域的小样本深度学习问题，该模型鲜少用于解决序列标注类任务，这个方法通过共享层实现迁移学习，并运用了目前飞速发展的生成对抗网络技术，来进一步提升模型的表现，实验证明该模型确实对 F1 分数有不错的提升。

2 基础结构

我们采用了经典的 Word Embedding + Bi-LSTM + CRF 模型 [9]作为对抗迁移模型的基础结构，如图 1 所示。该模型在绝大多数自然语言处理任务中都有着稳定的表现，我们将在此结构上进行改进，完成进一步提升。本节主要分析模型基础结构的一些技术特点。

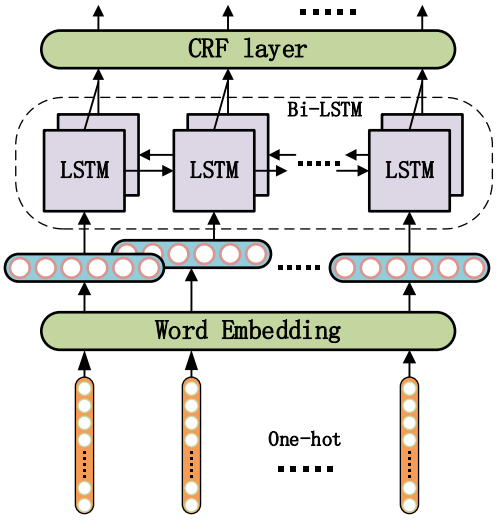

图1 中文分词基础模型

2.1 词嵌入技术

Bengio 在 2003 年 [10]，根据 Hinton 提出的分布式表示概念 [11]，提出了词嵌入技术(Word Embedding, WE)，并用之于神经网络，Zheng[12]在 2013 年中将该技术用于分词。WE 是将每一个字或词语映射到向量空间之中，用稠密的向量对其进行表示，将语义相近的词语之间距离拉近，这种表示方式改变了传统 one-hot 的数据输入模式，降低了数据维度以加快计算，提高语料利用率，并且 WE 也拓展了迁移的可能，极大地加强了模型的泛化能力。

2.2 Bi-LSTM 网络

LSTM 网络，是基于循环神经网络的深度学习模型。由 Schmidhube 在 1997 年提出 [13]，其计算单元如图 2 所示。

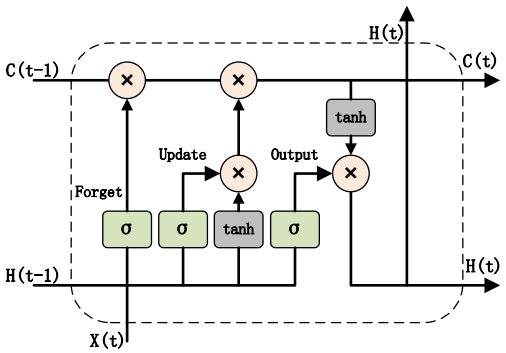

图2 LSTM计算单元

它的特点是在模型中加入了记忆细胞，用于保存当前输入的状态，这个细胞将在传输中一直传递下去，同时用一系列的门函数来判断是否处理细胞中的信息、是否更新细胞信息、以多大程度使用细胞提供的信息，从而将信息较好地传递到后方，达成长输入学习，公式如下：

$$\boldsymbol{\Gamma}_f^{<t>} = \sigma(\boldsymbol{W}_f[\boldsymbol{h}^{<t-1>}, \boldsymbol{x}^{<t>}] + b_f) \quad (1)$$

$$\boldsymbol{\Gamma}_u^{<t>} = \sigma(\boldsymbol{W}_u[\boldsymbol{h}^{<t-1>}, \boldsymbol{x}^{<t>}] + b_u) \quad (2)$$

$$\tilde{\boldsymbol{c}}^{<t>} = tanh(\boldsymbol{W}_c[\boldsymbol{h}^{<t-1>}, \boldsymbol{x}^{<t>}] + b_c) \quad (3)$$

$$\boldsymbol{c}^{<t>} = \boldsymbol{\Gamma}_f^{<t>} * \boldsymbol{c}^{<t-1>} + \boldsymbol{\Gamma}_u^{<t>} * \tilde{\boldsymbol{c}}^{<t>} \quad (4)$$

$$\boldsymbol{\Gamma}_o^{<t>} = \sigma(\boldsymbol{W}_o[\boldsymbol{h}^{<t-1>}, \boldsymbol{x}^{<t>}] + b_o) \quad (5)$$

$$\boldsymbol{h}^{<t>} = \boldsymbol{\Gamma}_o^{<t>} * tanh(\boldsymbol{c}^{<t>}) \quad (6)$$

其中：$\boldsymbol{W}_f$、$\boldsymbol{W}_u$、$\boldsymbol{W}_c$、$\boldsymbol{W}_o$ 是权重参数，b_f、b_u、b_c、b_o 为偏置参数，$\boldsymbol{c}^{<t>}$ 为时间步t的细胞函数，$\boldsymbol{x}^{<t>}$ 为输入向量，$\boldsymbol{h}^{<t>}$ 是隐藏层向量，$\boldsymbol{\Gamma}_f^{<t>}$、$\boldsymbol{\Gamma}_u^{<t>}$、$\boldsymbol{\Gamma}_o^{<t>}$ 分别为遗忘门、更新门、输出门。

1997 年 Schuster 和 Paliwal 提出双向的循环神经网络 [14]，双向的 LSTM 网络--Bi-LSTM 应运而生。单向的网络中将信息从头传至尾，而许多信息并不是简单的单向关联，因此双向网络的引入能使模型更好的理解上下文信息。

2.3 CRF 层

CRF 是 Lafferty 于 2001 年提出的一种无向图模型 [15]，如图 3 所示，它同时拥有 HMM 和 ME 的优点，在序列标注类任务中有举足轻重的作用。深度学习中通常使用 Softmax 函数来进行分类预测，但是序列标注的输出之间有着紧密的逻辑关系，因此 CRF 层通过发射分数矩阵和转移分数矩阵全面衡量输出之间的关系，最终给出符合逻辑的输出序列，大大的提高准确率。

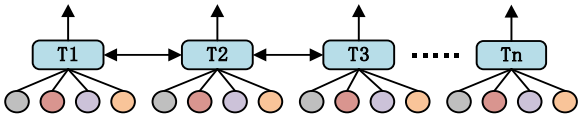

图3 CRF层

3 对抗迁移模型

本文提出的模型的整体结构如图 4 所示。我们采用多任务学习中的共享-私有结构 [16]，并在其基础上做出进一步改动。

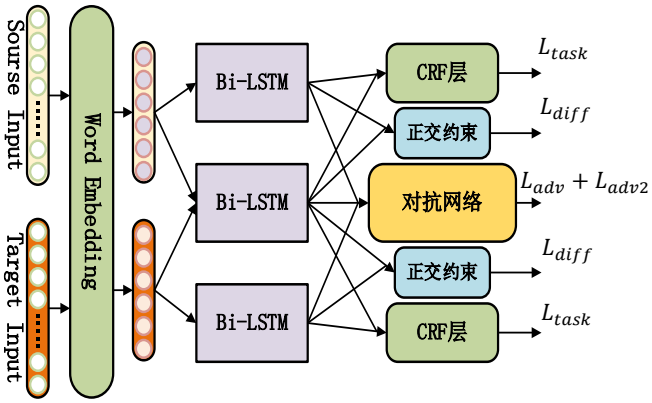

图4 对抗迁移模型

3.1 基于共享-私有结构的迁移

迁移学习的核心思想是在数据量充足的源领域中学习一些有用的知识信息来帮助目标领域更好地学习，接下来对共享-私有结构做出简单介绍。

我们将源域和目标域中学习到的特征分为两类：私有特征和共享特征。该模型将在源域和目标域学习到的特征分别引入两类独立的特征空间：私有特征空间和共享特征空间。私有特征强调领域本身的知识，共享特征更加蕴含对于分词方法的通性知识。由此可见，我们希望源域在神经网络学习的过程中，将对于目标域有用的知识信息通过共享特征空间传递给目标域。因此我们的目标就是让共享特征空间中出现更多有用的共享特征，同时让私有特征尽量少地出现在共享特征空间中，避免特征污染和冗余。

最后，给出分词任务的损失函数计算过程：

$$\hat{y}_t^k = \begin{cases} softmax(\boldsymbol{W}_s\boldsymbol{h}_s + \boldsymbol{W}_{sc}\boldsymbol{h}_c + b_s), if\ k=0 \\ softmax(\boldsymbol{W}_s\boldsymbol{h}_s + \boldsymbol{W}_{st}\boldsymbol{h}_t + b_t), if\ k=1 \end{cases} \quad (7)$$

$$L_{task} = \frac{1}{n}\sum_{i=1}^{n} CRF(\boldsymbol{Y}^k, \hat{\boldsymbol{Y}}^k) \quad (8)$$

其中：$\hat{y}_t^k$ 是对于 t 时间步的预测结果，$\boldsymbol{h}_s$、$\boldsymbol{h}_t$ 代表了源域和目标域的私有特征，$\boldsymbol{h}_c$ 代表共享特征，$\boldsymbol{W}_s$、$\boldsymbol{W}_t$、$\boldsymbol{W}_{sc}$、$\boldsymbol{W}_{tc}$ 是权重参数，b_s、b_t 为偏置参数。k 代表了领域标签（0 代表源域，1 代表目标域），$\hat{\boldsymbol{Y}}^k = (\hat{y}_1^k, \hat{y}_2^k, \hat{y}_3^k, ...)$ 是由各个时间步合并组成的单个样本。

3.2 正交约束

正交约束由 Konstantinos 于 2016 提出 [17]，是指通过将源域和目标域的私有特征与共享特征进行正交乘积，并在结果中按一定比例加入损失函数，完成共享特征和私有特征的差异化，计算公式如下：

$$L_{diff} = \left\| \boldsymbol{H}_s^T \boldsymbol{H}_{sc} + \boldsymbol{H}_t^T \boldsymbol{H}_{tc} \right\| \quad (9)$$

其中：$\boldsymbol{H}_s$ 和 $\boldsymbol{H}_{sc}$ 分别是源域的私有特征和共享特征构成的参数矩阵，同样，$\boldsymbol{H}_t$ 和 $\boldsymbol{H}_{tc}$ 是目标域的相应参数矩阵，这些矩阵均为 batch-size 行、特征维度列，batch-size 是批大小。

3.3 GAN 网络

Goodfellow [18]首先提出了 GAN 网络的概念，Ganin[19]在 2016 年将对抗思想用于解决迁移学习中的领域自适应问题。

正交约束的作用仅是增加共享特征和私有特征的差异，并不能明确保留在空间内的特征是共享特征还是私有特征，所以需要其他网络来完成分类任务。因为共享特征空间提取的是两个领域交叉的共有信息，因此在理想状态下，两个领域的输入经过共享层所提取的特征应当为不带有私有特征的共享特征，将这些特征输入神经网络和分类器后，模型应当无法给出可靠的预测。带有梯度反转的 GAN 网络便适用于这种情况，在正向传播过程中，LSTM 网络给出的共享特征试图误导分类器，而分类器则努力判断该特征来自哪个领域，在反向传播时梯度反转不断优化特征，使其变为不带有私有特征的共享特征，经过不断迭代后，对抗网络达到平衡。对抗网络的损失函数：

$$p(k \mid h_c) = \sigma(\boldsymbol{W}_c \boldsymbol{h}_c + b_c) \tag{10}$$

$$L_{adv} = -\frac{1}{n}\sum_{i=1}^{n}[k * log\, p(k \mid \boldsymbol{h}_{ic}) + (1-k) * log(1 - p(k \mid \boldsymbol{h}_{ic}))] \tag{11}$$

其中：$\boldsymbol{W}_c$ 是权重参数，b_c 为偏置参数 $p(k \mid \boldsymbol{h}_{ic})$。是对目标域 $(k=1)$ 的预测，n 为 batch-size 的大小，即样本个数。

3.4 辅助对抗学习

Chen[20]在论文中指出，Liu[16]的模型中对抗学习并没有起到作用。在实验之后我们也发现其模型相比于简单的多任务学习并没有明显提升，加大对抗网络和正交约束的损失函数比例反而会使模型退化更加严重，Chen 分析原因认为共享特征并不能很好的训练分类器，因此将私有特征也加入到分类器的训练当中，完善了模型。辅助对抗学习损失函数：

$$p(k \mid h_p) = \sigma(\boldsymbol{W}_c \boldsymbol{h}_p + b_p) \tag{12}$$

$$L_{adv2} = -\frac{1}{n}\sum_{i=1}^{n}[k * log\, p(k \mid \boldsymbol{h}_{ip}) + (1-k) * log(1 - p(k \mid \boldsymbol{h}_{ip}))] \tag{13}$$

其中：$\boldsymbol{W}_c$ 是权重参数，b_p 为偏置参数。$\boldsymbol{h}_p$ 为私有特征，是 $\boldsymbol{h}_s$、$\boldsymbol{h}_t$ 的集合。

最终的损失函数是：

$$L = L_{task} + \lambda_1 L_{diff} + \lambda_2 L_{adv} + \lambda_1 L_{adv2} \tag{14}$$

其中：λ_1、λ_2、λ_3 是可调节的超参数。λ_1 代表正交约束损失函数的比重，该值越大，训练后领域中的共享特征和私有特征差异越大。λ_2 是对抗训练损失函数的比重，需要根据目标领域和源领域的特征分布重合程度调节。λ_3 代表辅助对抗训练的损失比重，该值大小决定了模型对分类器性能关注程度，如果训练后对抗网络的分类器效果不好，便可进一步调整该参数。

4 实验

4.1 实验配置

我们尽可能收集了网络上的开源中文分词数据库，分别为：sighan2005 的 PKU、MSR、AS、CITYU 和 SXU、CTB、UDC、CNC 以及微博（WTB）和文学作品《诛仙》(ZX)共 10 个语料库。将它们的嵌入通过 t-SNE 可视化出来如图 5 所示，因为词嵌入向量的数值并没有具体意义，因此图 5 的坐标轴并未标出具体单位。通过图 5 发现各个语料库之间几乎没有显著的相似性 [21]。

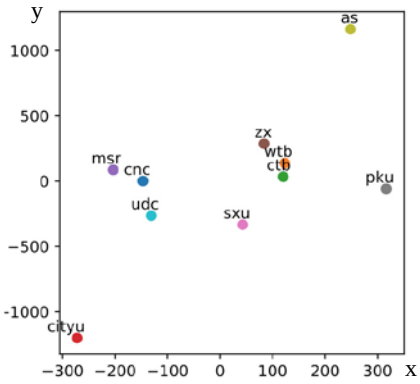

图5　语料库领域分布

经过多次实验，我们得出损失函数的超参数为 $\lambda_1 = \lambda_2 = 0.03, \lambda_3 = 0.05$ 时模型的表现最好。由于对抗训练需要一定的学习时间，所以迭代次数设定不能过少，本次实验 batch-size 设置为 128，迭代次数为 800。当仅使用正交约束进行训练时，λ_2 在 0~0.1 的范围内时模型性能有不错的提升，经过对比实验，正交约束的输入加入正则化后模型效果会进一步提高。当我们仅使用对抗和辅助对抗进行训练时，发现模型 F1 分数是否上升十分依赖损失函数的超参数。最后，由于有多个损失函数项，实验结果存在波动并不稳定，需要多次实验观察避免偶然性。

4.2 实验结果

本次实验取 10 万个样本作为源域数据，3000 样本作为目标域数据（其中 WTB 数据库不足 3000，用 700 训练集和 300 测试集样本进行实验），评价标准为 F1 分数（通过准确率和召回率计算得出，该指标能够更全面地衡量模型结果），按照数据量选择了 5 个语料库作为源域，5 个语料库作为目标域，对抗迁移模型在各个语料库的表现如下方表 1。

表1　模型在各个领域实验的F1分数

源域 \ 目标域	SXU	CTB	UDC	WTB	ZX
PKU	89.16	89.41	88.94	83.53	92.47
MSR	87.00	89.13	88.65	83.27	91.69
AS	88.54	89.13	88.92	83.47	92.79
CITYU	89.03	88.54	89.35	78.60	92.34
CNC	89.23	89.73	89.01	82.90	93.28

为了更直观地展现迁移结果，我们从 ZX 的预测结果中选出一个典型例子，如表 2 所示。

表2　语料库ZX中的实例

基础模型	入/青云门/后，阴差阳错/受到/佛道/魔三/方面/的/影响。
我们的模型	入/青云门/后，阴差阳错/受到/佛道魔/三方面/的/影响。

观察发现，在“佛道魔三方面”的分词过程中，数据量稀缺的基础模型认为“佛道”、“方面”经常作为完整的词独立出现，因而造成了分词错误。但在我们的模型中，大量的源域数据中含有更多“佛道魔”、“三方面”相关的分词样本，这些样本通过共享层迁移至目标域，使模型最后做出更准确的预测，同时分词的粒度也更为精细。

我们同时也进行了多组对比实验，分别为：基础模型，包括：使用源域语料训练模型后将其应用于目标域(source-only)、使用目标域语料训练模型后将其应用于目标域(target-only)、源域和目标域语料混合训练模型后将其应用于目标域(mix)等；目前的强基线模型 [16]为仅共享 Bi-LSTM 的迁移模型（SP-MTL），本文提出的对抗迁移模型(Ad-Tr)，我们随机选择了 5 种情况进行实验，S→T 代表从源域 S 向目标域 T 迁移，结果如表 3 所示。

表3　对比实验结果

模型		PKU→SXU	MSR→CTB	AS→UDC	CITYU→ZX	CNC→ZX
Basic model	Sourse-Only	87.84	83.00	81.41	84.14	85.71
	Target-Only	86.95	87.38	87.16	91.52	91.52
	Mix	88.75	85.14	84.46	89.90	89.19
SP-MTL		88.87	88.16	88.71	92.17	91.42
Ad-Tr		**89.16**	**89.13**	**88.92**	**92.51**	**93.28**

通过观察我们发现 CNC 语料对 ZX 迁移效果最好，AS 对 UDC 效果最差，因此我们进一步做了 CNC 和 AS 对目标域语料的对比实验。通过观察结果图 6 和图 7，我们发现 AS 对各个语料的提升微弱，CNC 对各个语料均有不错的提升，分析原因可能是 AS 的语料内容比较单一，在迁移学习中没有足够的知识可以帮助目标域，而 CNC 的语料相比之下更加丰富，在各个领域都有涉及，可以更好地完成知识共享。

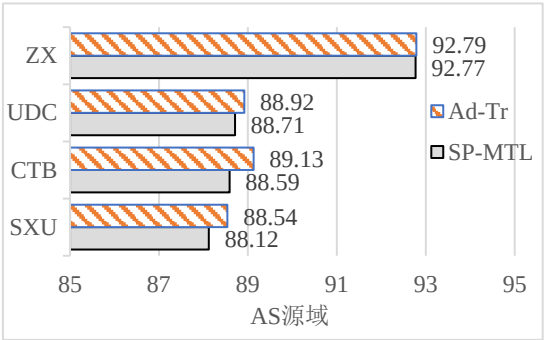

图6　AS作为源域时的F1分数

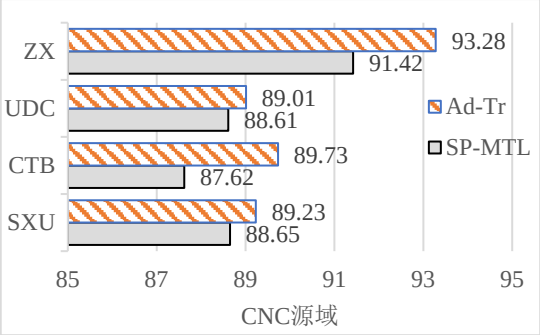

图7　CNC作为源域时的F1分数

5 结论与展望

本文实验结果说明，虽然中文分词各语料领域分布和分词标准不同，但语料间仍存在可用于迁移学习的信息，同时，对抗学习和正交约束可提升共享特征的纯净度。本文模型同样适用于诸如命名实体识别、词性标注等其他序列标注类任务。此外，由于对抗学习可引入未标注数据，故今后可在对抗学习中引入无标注语料进一步提升效果。

参考文献

[1] 俞鸿魁, 张华平, 刘群, 等. 基于层叠隐马尔可夫模型的中文命名实体识别[J]. 通信学报. 2006, 027(002):87-94.

[2] Jaynes E T . On the Rationale of Maximum-Entropy Methods[C]. Proceedings of the Institute of Electrical and Electronics Engineers, 1982, 70(9):939-952.

[3] Fuchun Peng, Fangfang Feng and Andrew McCallum. Chinese Segmentation and New Word Detection Using Conditional Random Fields[C]. Proceedings of the International Conference on Computational Linguistics, 2004, 562-569.

[4] Xinchi Chen, Xipeng Qiu, Chenxi Zhu, et al. Long Short-Term Memory Neural Networks for Chinese Word Segmentation[C]. Conference on Empirical Methods in Natural Language Processing, 2015, 1197-1206.

[5] Jacob Devlin, MingWei Chang, Kenton Lee, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]. North American Chapter of the Association for Computational Linguistics, 2019.

[6] 成于思, 施云涛. 基于深度学习和迁移学习的领域自适应中文分词[J]. 中文信息学报, 2019, 33(09): 16-23.

[7] 武惠, 吕立, 于碧辉. 基于迁移学习和BiLSTM-CRF 的中文命名实体识别[J]. 小型微型计算机系统, 2019, 40(06): 1142-1147.

[8] Jiang Guo, Wanxiang Che, Haifeng Wang, et al. A Universal Framework for Inductive Transfer Parsing across Multi-typed Treebanks[C]. International Conference on Computational Linguistics, 2016.

[9] Zhiheng Huang, Wei Xu and Kai Yu. Bidirectional LSTM-CRF Models for Sequence Tagging[J]. Computer Science, 2015.

[10] Yoshua Bengio, Holger Schwenk, Jean-Sébastien, et al. Neural Probabilistic Language Models[J]. The Journal of Machine Learning Research, 2003, 3(6):1137-1155.

[11] Alberto Paccanaro, Geoffrey Hinton. Learning Distributed Representations of Concepts using Linear Relational Embedding[J]. IEEE Transactions on Knowledge & Data Engineering, 1986.

[12] Xiaoqing Zheng, Hanyang Chen and Tianyu Xu. Deep Learning for Chinese Word Segmentation and POS Tagging[C]. Conference on Empirical Methods in Natural Language Processing, 2013, 647-657.

[13] Sepp Hochreiter, Jurgen Schmidhuber. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.

[14] Mike Schuster, Kuldip Paliwal. Bidirectional Recurrent Neural Networks[J]. IEEE Transactions on Signal Processing, 2002, 45(11):2673-2681.

[15] John Lafferty, Andrew Mccallum and Fernando Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data[C]. 18th International Conference on Machine Learning, 2001, 282-289.

[16] Pengfei Liu, Xipeng Qiu and Xuanjing Huang. Adversarial Multi-task Learning for Text Classification[C]. Proceedings of the 55th Annual Meeting of the Association for Computational

Linguistics, 2017.

[17] Bousmalis Konstantinos, Trigeorgis George and Silberman Nathan. Domain Separation Networks[J].
Neural Information Processing Systems, 2016.

[18] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, 2672-2680.

[19] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, et al. Domain-adversarial Training of Neural Networks[J]. The Journal of Machine Learning Research, 2016, 17(59):1-35.

[20] Cen Chen, Yinfei Yang, Jun Zhou, et al. Cross-Domain Review Helpfulness Prediction based on Convolutional Neural Networks with Auxiliary Domain Discriminators[C]. Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2018.

[21] He Han, Wu Lei, Yan Hua, et al. Effective Neural Solution for Multi-Criteria Word Segmentation[C]. Proceedings of the Second International Conference on Smart Computing and Informatics, 2018.