半监督跨领域语义依存分析技术研究

毛达展　　李华勇　　邵艳秋*
国家语言资源监测与研究平面媒体中心，信息科学学院
北京语言大学，北京市海淀区学院路 15 号，100083，中国
maodazhan@foxmail.com　lihuayong@blcu.edu.cn　yqshao163@163.com

摘要

近年来，尽管深度学习给语义依存分析带来了长足的进步，但由于语义依存分析数据标注代价非常高昂，并且在单领域上性能较好的依存分析器迁移到其他领域时，其性能会大幅度下降。因此为了使其走向实用，就必须解决领域适应问题。本文提出一个新的基于对抗学习的领域适应依存分析模型，我们提出了基于对抗学习的共享双编码器结构，并引入领域私有辅助任务和正交约束，同时也探究了多种预训练模型在跨领域依存分析任务上的效果和性能。

关键词： 语义依存分析 ； 领域适应 ； 对抗学习 ； 预训练模型

Semi-supervised Domain Adaptation for Semantic Dependency Parsing

Dazhan Mao　　Huayong Li　　Yanqiu Shao*
Language Resources Monitoring and Research Center,
Information Science School, Beijing Language and Culture University,
15 Xueyuan Road, HaiDian District, Beijing, 100083, China
maodazhan@foxmail.com　lihuayong@blcu.edu.cn　yqshao163@163.com

Abstract

Recently, although deep learning has brought significant progress to semantic dependency parsing, the semantic annotation data is very expensive to label, and when a dependency parser with better performance in a single domain is migrated to other domains, its performance will decline largely. Therefore, in order to make it practical, it is necessary to solve the problem of domain adaptation. This paper proposes a new domain adaptation dependency parsing model based on adversarial learning. We proposed a shared dual encoder structure based on adversarial learning, and introduced domain private auxiliary tasks and orthogonal constraints. At the same time, we also explored a variety of pre-training models in the cross domain dependency parsing task about the effectiveness and performance.

* 通讯作者　Corresponding Author

第十九届中国计算语言学大会论文集，第783页-第794页，海南，中国，2020年10月30日至11月1日。
(c) 2020 中国中文信息学会计算语言学专业委员会

Keywords: Semantic dependency parsing , Domain adaptation , Adversarial learning , Pre-training model

1 引言

依存分析是一种句子结构的解析方式，其将句子的句法或语义结构解析为一系列二元、非对称依存关系，这些依存关系构成了句子的依存树（或依存图）。不同于句法依存树分析，语义依存图分析是一种深层次的语义解析，其描述的是句子各个组成部分间的语义关系 (Che et al., 2012)，如图 1 所示，其允许更复杂的依存结构（如多父节点、非投射等等）。由于其能够直接表达深层语义信息，因此应用价值更大。然而，现有的语义依存分析研究使用的数据集往往来自课本或者新闻等单个领域，这样即使依存分析器在数据集上取得了较高的性能，当迁移到其他目标领域时，分析器的性能也会大幅度下降。

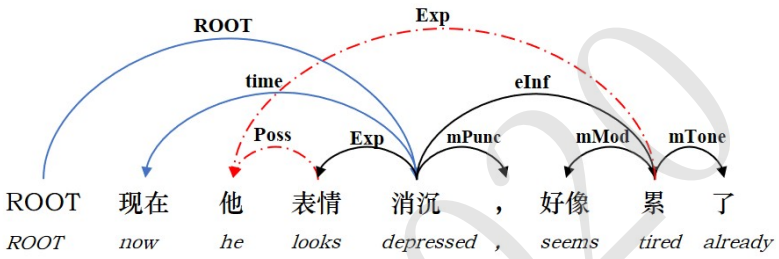

图 1. 语义依存分析示例。图中红色依存弧为多父节点现象，蓝色依存弧为非投射现象

根据目标领域的数据有无标注，领域适应可以划分为无监督领域适应（目标领域完全没有标注数据）和半监督领域适应（目标领域存在少量标注数据，同时也有大量无标注数据）(Kouw and Loog, 2018)。由于语义依存分析本身的复杂性，目前纯粹基于无监督的跨领域语义依存分析的研究进展相对滞后。而半监督的领域适应虽然仍需要少量的数据标注，但是其可以利用一定的监督信号指导领域适应，领域迁移效果更好，迁移后的模型实用价值更大，也能更好地和语义依存分析任务结合。因此本文关注于针对语义依存分析任务的半监督领域适应。本文的主要工作总结如下：

- 本文提出了一个新的基于对抗学习的领域适应框架。该框架支持一个模型同时解决面向多个目标领域的领域适应问题。该框架在实验数据集上明显优于基线模型。
- 本文将预训练语言模型融合到了对抗领域适应框架中，从而进一步提升了模型的领域适应能力。同时我们详尽讨论分析了应用预训练语言模型解决语义依存分析任务以及领域适应时的一系列细节问题。

2 相关工作

2.1 依存分析

现有的依存分析方法主要有两种，分别是基于转移的算法 (Chen and Manning, 2014);(Dyer et al., 2015) 和基于图的算法 (Chen et al., 2013);(Wang and Chang, 2016)。早期的这两种依存

©2020 中国计算语言学大会
根据《Creative Commons Attribution 4.0 International License》许可出版

分析器需要手动定义复杂的特征模板，这费时费力且需要很强的背景知识，限制了分析器的进一步发展 (Koo and Collins, 2010);(Koo and Collins, 2010)。

近年来，神经网络方法被广泛应用在依存分析中 (Chen and Manning, 2014);(Dozat et al., 2017)。在这些基于神经网络的依存分析器的研究工作当中，(Dozat and Manning, 2016) 双仿网络依存分析器取得了目前最优的性能。因此，双仿网络依存分析器在本文中，将作为后续对依存分析进行领域适应研究的基础。

2.2 领域适应

最近，随着 (Peters et al., 2018)ELMO;(Devlin et al., 2018)BERT 等上下文表示的兴起，大量工作开始研究基于预训练上下文表示的领域适应方法，并且取得了较好的结果，展示了预训练上下文表示在领域适应任务上的巨大潜力。(Liu et al., 2019a) 分析了上下文表示中的语言学知识和可迁移性。(Mulcaire et al., 2019) 使用上下文表示提升了跨语言任务的迁移效果。受到这些工作的启发，本工作将把预训练模型融入依存分析的领域适应模型中，探究上下文信息对跨领域依存分析是否有帮助。

对抗学习已经被证明可以明显提升跨领域依存分析器的性能 (Bousmalis et al., 2016);(Ganin and Lempitsky, 2014)。但是大部分的工作为了抽取不同领域之间的无关特征，都是把领域无关的特征和领域私有的特征混合在一起，这就不可避免地损失一些领域私有的信息 (Sato et al., 2017)。(Chen et al., 2017) 针对中文多粒度分词任务，提出了一个 Shared-Private 模型。在这个模型的基础上，本文对私有编码器进行简化，不同领域的私有编码器统一成一个，并增加领域预测的辅助任务。(Liu et al., 2017);(Shi et al., 2018) 引入了正交约束来消除共享空间和私有空间之间的冗余信息。在本文中也将把正交约束应用到领域无关编码器和领域私有编码器之间。

3 基于对抗学习的领域适应依存分析模型

与一般基于对抗的跨领域依存分析做法一样，都是混合源领域和目标领域的数据输入到 Biaffine 编码器，但本模型增加了 BERT 通用编码层、领域共享双编码器、领域分类辅助任务以及正交约束等可能对模型性能有提升作用的组件。

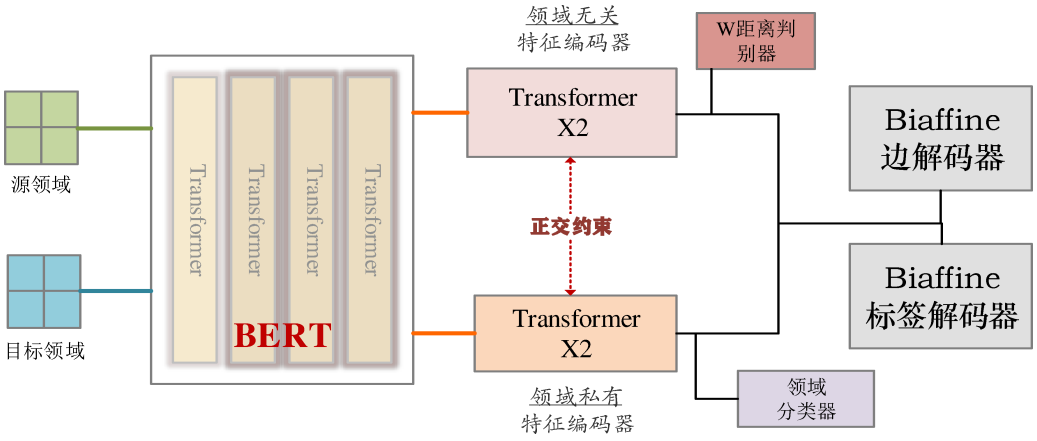

图 2. 基于对抗学习的领域适应依存分析模型结构

(c) 2020 中国中文信息学会计算语言学专业委员会

3.1 BERT 通用编码层

经典的依存分析器采用词向量加词性向量的静态表征，有时也会以字符向量表示加以辅助，这种经典的组合方式不能为每个词提供基于上下文的正确表示，也无法很好地解决未登录词问题。近年来，随着 BERT 等预训练语言模型的涌现，越来越多的研究开始使用预训练语言模型替换经典的词向量输入，同时也有大量研究表明 BERT 等预训练语言模型对于跨领域迁移有着很好的帮助，因此本文使用 BERT 作为底层编码。BERT 是多层 Transformer 神经网络 (Vaswani et al., 2017) 的堆叠，形式化地，BERT 每层的处理过程可表示为：

$$h_{i,j} = BERT_j(x_i) \tag{1}$$

其中，i 表示第 i 个输入，j 表示第 j 层 BERT，x_i 是输入的字符。

BERT 默认选择使用最后一层 BERT 输出作为整体输出，但是已有大量研究表明 BERT 等预训练语言模型每层的编码信息并不相同，一般 BERT 底层涉及一些语言基础知识，BERT 中层编码了一定的句法结构知识，BERT 高层则编码了语义知识，且 BERT 和训练时的任务相关度很高。因此直接使用最后一层 BERT 输出可能不是最好的方案，为此，本文引入了层加权机制，以一种可训练的方式加权平均不同 BERT 层的输出。层加权机制可形式化为：

$$h_i = c\sum_i BERT_{j,i} \cdot softmax(w_j) \tag{2}$$

其中 w_j 是一个可训练的“权重”标量，用来对应每一层 BERT 输出；c 是一个可训练的缩放标量，用了缩放最后的加权表示；$BERT_{j,i}$ 表示第 j 层 BERT 在第 i 个位置的输出。

经过层加权机制后，可以得到对应输入的字符序列表示，由于依存分析是基于词语级别的，所以需要从字符序列映射到词语序列，我们采用简单的尾字表示法完成映射，即对于每个词语只选择尾字对应的表示来作为整个词语的表示。

3.2 领域共享双编码器

在预训练语言模型之后，又连接了两个领域共享编码器，一个是领域无关特征编码器 $f_{share}^{E(x)}$，一个是领域私有信息编码器 $f_{private}^{E(x)}$，分别负责提取领域无关特征和领域私有特征。两个编码器均使用两层 Transformer 神经网络实现，每层 Transformer 网络可以形式化地表示为：

$$Transformer(X) = Skip(FF, Skip(MultiHead, X)) \tag{3}$$

$$Skip(f, h) = LayerNorm(h + Dropout(f(h))) \tag{4}$$

$$FF(h) = GELU(hW_1^T + b_1)W_2^T + b_2 \tag{5}$$

其中，(Hendrycks and Gimpel, 2016)GELU 代表高斯误差线性单元激活（Gaussian error linear units）函数。为了保证领域无关特征编码器可以提取到领域共享的特征，我们在领域无关编码器上额外连接了一个对抗判别器，基于对抗学习的方式强制编码器编码领域无关特征。同时为了保证领域私有编码器可以提取到每个领域的私有信息，我们在领域私有编码器上也额外连接了一个领域分类辅助任务。

第十九届中国计算语言学大会论文集，第783页-第794页，海南，中国，2020年10月30日至11月1日。
(c) 2020 中国中文信息学会计算语言学专业委员会

3.3 对抗判别器

“领域无关”特征编码器除了连接依存任务所需的解码器 $Biaffine^{edge}$ 和 $Biaffine^{label}$ 外，还额外连接一个对抗判别器 $D_{adv}(x)$，负责提取领域之间的不变特征。

参考 WGAN 的实现 (Arjovsky et al., 2017);(Arjovsky and Bottou, 2017)，本文采用基于 Wasserstein 距离的对抗判别器。在使用基于 Wasserstein 距离的损失作为对抗损失时，对抗判别器实际上是一个 Wasserstein 距离回归网络。

形式化地，对于源领域的输入数据 X_{source} 和目标领域的输入数据 X_{target}，经过领域特征编码器后，我们分别得到对应的表示分布 P_s 和 P_t，则 P_s 和 P_t 之间的 Wasserstein 距离等于：

$$W(P_s, P_t) = \sup_{\|f\|_L \leq 1} E_{x \sim P_s}[f(x)] - E_{x \sim P_t}[f(x)] \tag{6}$$

其中，f 是一个 Lipschitz-1 连续函数，注意为了求解 Wasserstein 距离，这里依据 WGAN 对其定义公式做了转换。根据 WGAN 中的要求，我们使用一层全连接神经网络 f^W 近似该 Lipschitz-1 连续函数，同时将该网络的参数取值范围固定到 [-0.01,0.01] 之间。

进而可以计算得到 Wasserstein 距离对抗损失 L_{adv}^W：

$$L_{adv}^W(S^s, S^t) = f^W(S^s) - f^W(S^t) \tag{7}$$

在训练时，一方面我们需要优化“判别器”以产生最准确的 Wasserstein 距离，为此需要在“判别器”的参数上最小化 Wasserstein 距离对抗损失 $L_{adv}^W(S^s, S^t)$；另一方面，本文需要使领域无关特征编码器产生的两个表示分布尽可能“迷惑”Wasserstein 距离判别器，为此，本文需要在领域无关特征编码器的参数上最大化 Wasserstein 距离对抗损失 $L_{adv}^W(S^s, S^t)$。

由上述可知基于 Wasserstein 距离的对抗学习过程是一个 minmax 训练，即：

$$\min_{\Theta^{dis}} \max_{\Theta^{share}} L_{adv}^W \tag{8}$$

其中，Θ^{dis} 表示判别器的参数，Θ^{share} 表示判别器的参数。

在训练时我们通过先进行 $\min_{\Theta^{dis}}$ 训练，然后再进行 $\max_{\Theta^{share}}$ 训练的方式交替完成整个训练过程

3.4 Biaffine 解码层

本文使用双仿解码器来分别预测两个词语之间的依存弧关系和依存标签。首先将编码输出的词语级别的表示向量 h_i^{lstm} 传入两个前馈神经网络层 (FNN)，分别得到该词语的“头表示“和“尾表示”：

$$h_i^{edge-head} = FNN^{edge-head}(h_i^{lstm}) \tag{9}$$

$$h_i^{edge-dep} = FNN^{edge-dep}(h_i^{lstm}) \tag{10}$$

随后使用双仿变换整个句子中可能存在的依存弧的得分矩阵 $s_{i,j}^{edge}$：

$$Biaffine(x_1, x_2) = x_1^T U x_2 + W(x_1 \otimes x_2) + b \tag{11}$$

$$s_{i,j}^{edge} = Biaffine^{edge}(h_i^{edge-dep}, h_j^{edge-head}) \tag{12}$$

$$p_{i,j}^{edge} = sigmoid(s_{i,j}^{edge}) \tag{13}$$

(c) 2020 中国中文信息学会计算语言学专业委员会

训练时，依存弧解码器的损失定义为：

$$J_{edge}(\Theta^p) = -p_{i,j}^{edge} \log p_{i,j}^{edge} - (1 - p_{i,j}^{edge}) \log(1 - p_{i,j}^{edge}) \tag{14}$$

依存标签的方式和预测依存弧的方式非常相似，唯一不同的就是两个词语之间的依存标签的分类空间比较大，因此这里使用 softmax(Grave et al., 2016) 而不是 sigmoid 函数处理，最终得到依存标签概率 $p_{i,j}^{label}$。

$$p_{i,j}^{label} = softmax(s_{i,j}^{label}) \tag{15}$$

训练时，依存标签解码的损失定义为：

$$J_{label}(\Theta^p) = -\sum_{label} \log p_{i,j}^{label} \tag{16}$$

最后将依存弧概率和依存标签概率传给解码算法，就能得到最后的依存图。

在训练时，通过最小化依存损失 $J_{parser}(\Theta^p)$ 从而训练得到一个领域内依存分析器，依存分析损失由依存弧损失和依存标签损失相加得到：

$$J_{parser}(\Theta^p) = \beta J_{label}(\Theta^p) + (1 - \beta) J_{edge}(\Theta^p) \tag{17}$$

其中，β 是一个超参数，用来控制最终损失中两个解码器损失的相对大小。

3.5 领域分类辅助任务

我们希望私有编码器能够提取领域私有的信息，但仅通过最小化依存任务的损失 L_{parser} 无法保证私有特征编码器真正提取到了对应领域的私有信息，因此本工作又额外引入了一个私有辅助任务，即领域分类任务，负责判断编码器编码的特征属于哪一个领域。这一辅助任务类似于文本领域分类，由一个领域分类器 $f^c(x)$ 实现，其包括一层全连接神经网络和一个 softmax 层：

$$f^c(p^T, \theta_C) = softmax(b + UP^T) \tag{18}$$

其中，b 和 U 代表全连接层的参数，P 代表私有信息编码层 $f_{private}^E$ 的输出特征。

训练时，领域分类器的交叉熵损失 $L_{classify}$ 定义为：

$$L_{classify} = -\sum_{i=1}^{N} \sum_{j=1}^{2} y_i^j \log(\hat{y}_i^j) \tag{19}$$

其中，$\hat{y}_i^j$ 为 softmax 层的预测标签，$\hat{y}_i^j$ 为真实标签。

通过通过最小化 $L_{classify}$，可以迫使领域私有特征编码器编码对应领域的私有特征。

3.6 正交约束

加入辅助任务后可以保证领域私有特征编码器学习到了领域的私有信息，但是私有特征编码器可能会学习到一部分领域无关特征，造成特征冗余表达。为了确保这两个编码器之间不存在冗余的特征，本工作为两个编码器之间增加了一个正交约束，在训练时惩罚领域私有编码器中和“领域无关”编码器重合的特征，从而促使领域私有信息编码器不提取领域间的不变特征。正交约束损失的定义如下：

$$L_{diff} = \|S^T P\|_F^2 \tag{20}$$

(c) 2020 中国中文信息学会计算语言学专业委员会

这里 S 代表领域无关编码器 f^E_{share} 的输出，P 代表领域私有信息编码层 $f^E_{private}$ 的输出，$\|\cdot\|^2_F$ 代表平方 Frobenius 范数。

矩阵 A 的 Frobenius 范数 $\|A\|_F$ 定义为：

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{i,j}|^2} \tag{21}$$

由上述公式可知，Frobenius 范数代表了矩阵的所有元素平方和的开方。因此，通过最小化正交约束 L_{diff}，就迫使 S^TP 的乘积最小化，进而等价于迫使两个矩阵相互“正交”，而从使得两个编码器的输出特征互不重叠。

3.7 联合训练

通过将上述的所有任务损失整合起来，得到了总共 4 个损失，分别是对抗损失 L^C_{adv}（或者 L^W_{adv}）、依存分析任务损失 L_{parser}、领域私有信息编码层辅助任务损失 $L_{classify}$、领域无关信息编码器和领域私有信息编码器之间的正交损失 L_{diff}。我们定义最终的训练目标损失 L 为：

$$L = L_{parser} + \lambda L_{adv} + \gamma L_{classify} + \eta L_{diff} \tag{22}$$

其中，依存分析的任务损失定义为：

$$L_{parser}(\Theta^p) = \beta L_{label}(\Theta^p) + (1-\beta)L_{edge}(\Theta^p) \tag{23}$$

上述 β、λ、γ、η 均为控制损失大小的超参数。注意，当使用目标领域的无标注数据时，L_{parser} 只在源领域的数据上计算。

4 实验部分

4.1 数据集介绍

本研究的源领域数据集来自 the SemEval-2016 task9(Che et al., 2012) 和《博雅汉语》。经过调研，选择两大类四小类目标领域，一大类是文学风格，主要包括散文（《文化苦旅》）、小说（《小王子》、《少女小渔》）、剧本（《武林外传》）三个子目标领域。另一大类是下游应用，主要包括医疗诊断文本子目标领域。

依据中文语义依存图标注规范，依托语义依存图标注平台，我们组织了 6 名语言学专业的学生做了数据标注。对于每个目标领域，我们只标注了少部分数据，并将其划分为训练集、验证集、测试集，并对剩余的无标注数据做了清洗和筛选，如表 1 所示。

表 1. 数据集划分

领域说明			人工标注数据集			无标注数据
			训练集	验证集	测试集	
源领域		平衡语料	38000	2000	2000	0
目标领域	文学	散文	3000	1000	1000	20000
		小说	3000	1000	1000	30000
		剧本	3000	1000	1000	8000
	应用	医疗	2000	500	500	30000

(c) 2020 中国中文信息学会计算语言学专业委员会

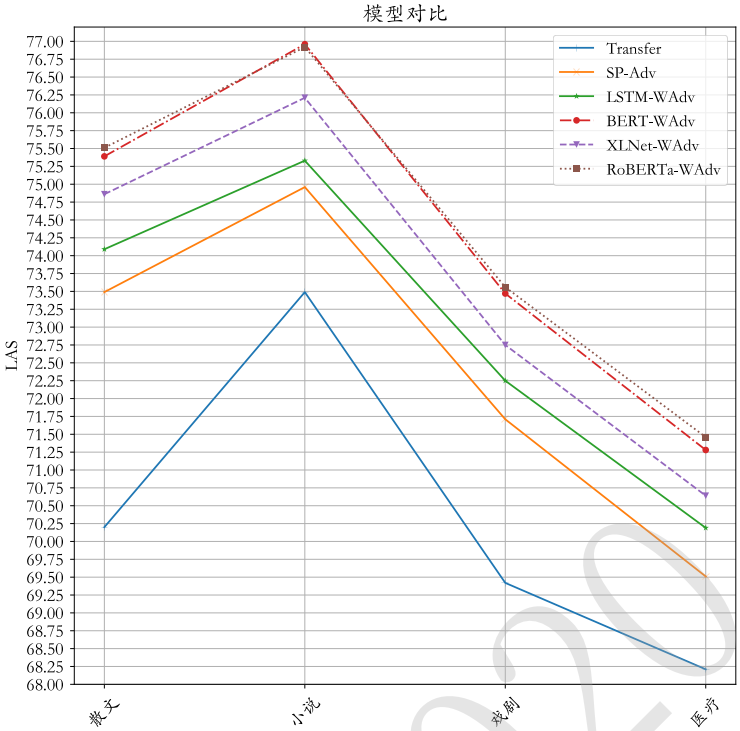

图 3. 本工作的模型和基线模型的对比

4.2 实验设置

我们尝试了多种预训练语言模型，其层数均为为 12，隐层向量维度均为 768。领域私有特征编码器和领域无关特征编码器都使用两层 Transformer 神经网络，其中 Transformer 层的注意力头数为 8，隐层向量维度为 768，dropout 比例为 0.2，使用 Relu 激活函数。对抗损失的控制参数 λ 为 0.5；领域分类辅助任务损失的控制参数 γ 为 0.05；正交约束损失的控制参数 η 为 0.001；依存损失的控制参数 β 为 0.5。对抗判别器的学习率设置为 0.0001，模型的其他部分的学习率设置为 0.001。在训练时使用带 L2 正则的 Adam 优化算法，min 训练和 max 训练的交替比例为 5:1。输入最大句长为 100，超过此长度的句子将被跳过。本文使用 4 张 NVIDIA Tesla V100-16GB 的显卡完成训练，单卡的批量大小设置为 32。

4.3 基线模型

为了更好地比较提出的模型的领域适应能力，我们选择了两个基线模型，分别是迁移模型**Transfer**和基于领域分类对抗损失的“共享 -私有”模型 **SP-Adv**：

- **Transfer**：Transfer 使用基于 LSTM+Biaffine 的单领域依存分析模型，在训练时，Transfer 模型先在源领域的数据上预训练，然后再在对应的目标领域上进一步微调。
- **SP-Adv**：模型使用经典的“共享 -私有”框架，同样使用对抗训练，但是其不采用正交约束，也不采用领域预测的辅助任务。

此外，为了进一步对比基于预训练语言模型的动态表征和传统的基于词向量的静态表征之

(c) 2020 中国中文信息学会计算语言学专业委员会

间的差别，我们将预训练语言模型替换为词向量加词性向量，模型其他部分保持不变，得到另一个基线模型，称为 **LSTM-WAdv**。

4.4 实验结果

4.4.1 与基线模型的对比

表 2 展示了我们的模型和基线模型在 4 个目标领域上的 LAS 指标，其中 **Transfer**、**SP-Adv**分别代表两个基线模型的结果，**LSTM-WAdv**代表在本文提出的模型上去掉预训练语言模型之后的结果，**BERT-Wadv**(Devlin et al., 2018);**XLNet-WAdv**(Yang et al., 2019);**RoBERTa-WAdv**(Liu et al., 2019b) 分别代表使用 **BERT**、**XLNET**、**RoBERTa**预训练语言模型的结果。

为了更加直观地比较差异，我们绘制了模型之间的对比折线图（如图 3），由图 3 可以看出，我们提出的基于预训练语言模型和对抗学习的领域适应框架都明显优于两个基线模型。同时使用预训练语言模型的领域适应框架也要优于使用词向量的框架。同时在三种预训练语言模型中，**RoBERTa**展现了最好的领域适应性能。

表 2. 本工作的模型和基线模型在 4 个目标领域上的 LAS 指标

模型	散文	小说	戏剧	医疗
Transfer	70.20	73.49	69.42	68.21
SP-Adv	73.49	74.96	71.71	69.51
LSTM-WAdv	74.09	75.33	72.25	70.19
BERT-WAdv	75.39	**76.96**	73.47	71.28
XLNet-WAdv	74.86	76.21	72.75	70.64
RoBERTa-WAdv	**75.51**	76.92	**73.56**	**71.46**

4.4.2 无标注数据对领域适应的影响

为了进一步探索无监督数据量在半监督学习中的影响，我们又做了两组对比实验。这两组实验分别选择前述实验中 LAS 最高的小说目标领域和 LAS 最低的医疗目标领域。本文将这两个领域的所有无标注数据划分为相等的 10 份，从不使用无标注数据到使用全部的无标注数据，逐步增加无标注数据的数量训练模型，并记录对应的 LAS 指标。如图 3 所示，无论是医疗领域还是小说领域，LAS 指标都随着无标注数据量的增加呈现接近线性关系的增长。注意，在小说领域上，当无标注数据使用超过七成的时候，LAS 指标的提升已经非常微弱，这说明此时两个编码器已经基本收敛，无法进一步提升。

4.4.3 消融实验

为了进一步分析本文提出的不同组件对最终模型领域适应性能的影响，我们在 **LSTM-WAdv**的基础上又做了相应的消融实验，如表 3 所示，分别记录了去掉对抗损失、去掉正交约束、去掉领域预测辅助任务以及去掉私有特征编码器时的实验结果。

(c) 2020 中国中文信息学会计算语言学专业委员会

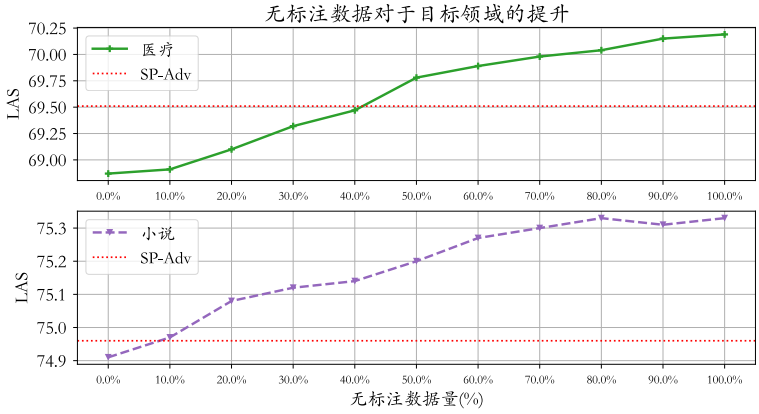

图 4. 无标注数据量对领域适应的影响

表 3. 消融实验

实验	散文	小说	剧本	医疗	平均下降
LSTM-WAdv	74.09	75.33	72.25	70.19	——
去掉对抗	72.82	74.90	71.10	69.48	0.890
去掉正交约束	73.90	75.16	71.81	69.84	0.288
去掉辅助任务	73.62	75.20	72.15	69.91	0.245
去掉私有特征	73.41	75.01	71.60	69.74	0.525

从表中可以看出，以上四个组件中，对模型最终效果影响最大的是对抗损失，去掉其之后模型在 4 个目标领域上平均 LAS 下降了 0.89，这再次证明了对抗学习技术在领域适应任务中的重要作用；其次影响模型性能的组件是私有特征，去掉其之后模型 LAS 平均下降了 0.525，这里需要注意一旦去掉私有编码器，正交约束和辅助任务也相应地失去了作用，因此私有特征的影响要大于其他两个组件。同时从表中可以看出，四个组件均对模型最终的性能有积极作用，其中影响最小的辅助任务也有 0.245 的平均共享。上述实验充分证明了本章提出的模型方法是有效的。

5 结论

在之前提到的跨领域分析数据集上，本文提出的基于预训练语言模型和对抗学习的领域适应框架都明显优于两个基线模型，在尝试的三种预训练模型中，RoBERTa 展现了最好的领域适应性能。在消融实验中，也验证了本文提出的领域适应框架的各个组件对模型最终性能的提升是有积极作用的。

致谢

本成果受国家自然科学基金项目（61872402），教育部人文社科规划基金项目 (17YJAZH068)，北京语言大学校级项目（中央高校基本科研业务费专项资金）（18ZDJ03），模式识别国家重点实验室开放课题基金资助。

(c) 2020 中国中文信息学会计算语言学专业委员会

参考文献

Martin Arjovsky and Leon Bottou. 2017. Towards principled methods for training generative adversarial networks. *Stat*, 1050.

Martin Arjovsky, Soumith Chintala, and Leon Bottou. 2017. Wasserstein gan.

Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. 2016. Domain separation networks. *CoRR*, abs/1608.06019.

Wanxiang Che, Meishan Zhang, Yanqiu Shao, and Ting Liu. 2012. Semeval-2016 task 9: Chinese semantic dependency parsing. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*.

D. Chen and C. D. Manning. 2014. A fast and accurate dependency parser using neural networks.

Wenliang Chen, Zhang Min, and Haizhou Li. 2013. Utilizing dependency language models for graph-based dependency parsing models. In *Meeting of the Association for Computational Linguistics: Long Papers*.

Xinchi Chen, Zhan Shi, Xipeng Qiu, and Xuanjing Huang. 2017. Adversarial multi-criteria learning for Chinese word segmentation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1193–1203, Vancouver, Canada, July. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding.

Timothy Dozat and Christopher D Manning. 2016. Deep biaffine attention for neural dependency parsing.

Timothy Dozat, Peng Qi, and Christoper Manning. 2017. Stanford's graph-based neural dependency parser at the conll 2017 shared task. pages 20–30, 01.

Chris Dyer, Miguel Ballesteros, Wang Ling, Austin Matthews, and Noah A. Smith. 2015. Transition-based dependency parsing with stack long short-term memory. *Computer Science*, 37(2):321–C332.

Yaroslav Ganin and Victor Lempitsky. 2014. Unsupervised domain adaptation by backpropagation.

Edouard Grave, Armand Joulin, Moustapha Cisse, David Grangier, and Herve Jegou. 2016. Efficient softmax approximation for gpus.

Dan Hendrycks and Kevin Gimpel. 2016. Gaussian error linear units (gelus).

Terry Koo and Michael Collins. 2010. Efficient third-order dependency parsers. In *ACL 2010, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden*.

Wouter M. Kouw and Marco Loog. 2018. An introduction to domain adaptation and transfer learning.

Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. 2017. Adversarial multi-task learning for text classification.

Nelson F Liu, Matt Gardner, Yonatan Belinkov, Matthew E Peters, and Noah A Smith. 2019a. Linguistic knowledge and transferability of contextual representations.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019b. Roberta: A robustly optimized bert pretraining approach.

Phoebe Mulcaire, Jungo Kasai, and Noah A. Smith. 2019. Polyglot contextual representations improve crosslingual transfer. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3912–3918, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations.

(c) 2020 中国中文信息学会计算语言学专业委员会

Motoki Sato, Hitoshi Manabe, Hiroshi Noji, and Yuji Matsumoto. 2017. Adversarial training for cross-domain universal dependency parsing. In *Conll Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*.

Ge Shi, Chong Feng, Lifu Huang, Boliang Zhang, and Heyan Huang. 2018. Genre separation network with adversarial training for cross-genre relation extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.

Wenhui Wang and Baobao Chang. 2016. Graph-based dependency parsing with bidirectional lstm. In *Meeting of the Association for Computational Linguistics*.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding.

CCL2020

(c) 2020 中国中文信息学会计算语言学专业委员会