

基于桥的主动域自适应特征提取

陈康平

摘要

作为一项细粒度的任务，方面术语提取的注释成本非常高。最近的尝试使用跨域传输公共知识的域适应来缓解这个问题。由于大多数方面术语都是特定于领域的，因此它们不能直接转移。现有方法通过将方面术语与枢轴词相关联来解决这个问题（我们称之为被动域适应，因为方面术语的转移依赖于到枢轴的链接）。然而，所有这些方法都需要手动标记关键词或昂贵的计算资源来建立关联。在本论文提出了一种新颖的主动域适应方法。其目标是通过积极补充可转移知识来转移方面术语。为此，可以通过将句法角色识别为枢纽而不是枢纽的链接来构建句法桥梁。还通过检索可转移的语义原型来构建语义桥梁。

关键词：序列标记任务；语义桥；句法桥；端对端标记器

1 引言

方面术语提取（ATE）是基于方面的情感分析中的一项基本任务。给定一条评论语句“这里的披萨也绝对好吃。”，ATE 旨在提取披萨这个词。最近的研究将 ATE 定义为序列标注任务，并提出了有监督的标注器^[1]。然而，由于令牌级标注成本高昂，缺乏标注数据成为主要障碍^[2]。为了缓解数据不足问题，提出了无监督域自适应将知识从标记的源域转移到未标记的目标域。由于 ATE 是 token 级别的任务，自然要进行 token 级别的域适配。

方面术语是特定领域的，不能直接转移。在图 1 (R (Restaurant)、L (Laptop) 和 d (Device) 是来自不同域的三个数据集。) 中展示了也出现在目标测试数据中的源方面术语的比例。可以看出，在 R→L 等远距离传输对中，只有不到 10% 的源方面术语出现在目标数据中。即使在紧对的 L→D 中，比例也不超过 40%。也就是说，不同领域的数据存在很大差异，很多方面的术语都需要在适当的引用指导下进行转换。^[3]

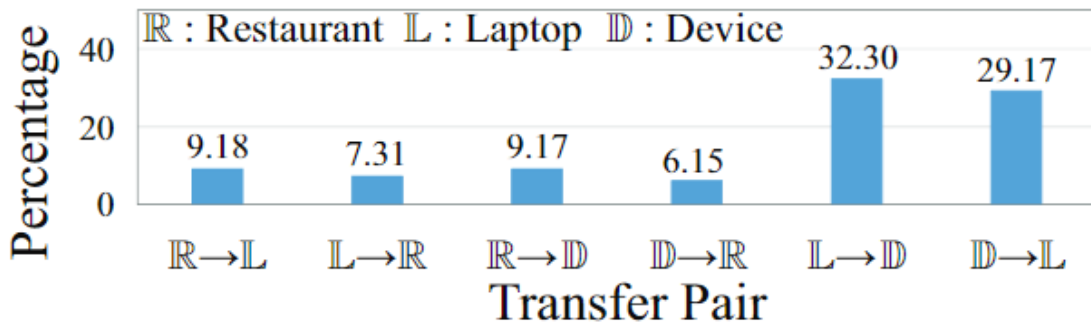


图 1: 目标数据中出现的源方面术语的比例

2 相关工作

为了解决这个问题，以前的研究试图将方面术语与特定的枢轴词联系起来。本论文将这些方法命名为被动域适应，因为方面术语的转移取决于它们与枢轴的链接，沿着这条线有两种方法。

2.1 基于意见条款

由于方面和意见项通常成对出现，因此可以直接从意见项中提取带有指示的方面项^[4]。早期的研究使用常见的意见种子（例如，good、fancy）和预定义的规则（例如，good→amod→NN）来跨域提取方面术语^[5]。然而，很难收集到一套完整的种子或定义高质量的规则，因此这些方法往往会产生较差的性能。

2.2 基于上下文术语

由于像 BERT 这样的预训练语言模型 (PLM) 可以根据上下文来表示单词，因此最近的研究利用 PLM 来传输具有常见上下文术语的方面术语^[6]。然而，并不是所有的上下文术语都可以作为枢纽（例如，吃）。此外，像 BERT 这样的 PLM 主要基于共现上下文中的语义相似性来构建单词关联。对于像披萨这样的方面术语，BERT 倾向于通过像披萨 → 吃 → 汉堡这样的流程将其链接到汉堡包。因此，这些方法很难根据源域中的标签词比萨来识别目标域中的键盘。

3 本文方法

3.1 本文方法概述

ATE 是一个序列标记任务，即预测每一个词元的 BIO 标签。复现的论文中采用了无监督的领域自适应方法解决 ATE 问题，即在目标域没有带标记的训练数据。具体来说，给定源域中的一组标记数据 $D^s = (x_s^j, y_s^j)_{j=1}^{N_s}$ 和一组目标域中的未标记数据 $D^U = (x_j^U)_{j=1}^{N_U}$ ，目标是预测测试数据集 $D^T = (x_j^T)_{j=1}^{N_T}$ 中的标签 y^T 。

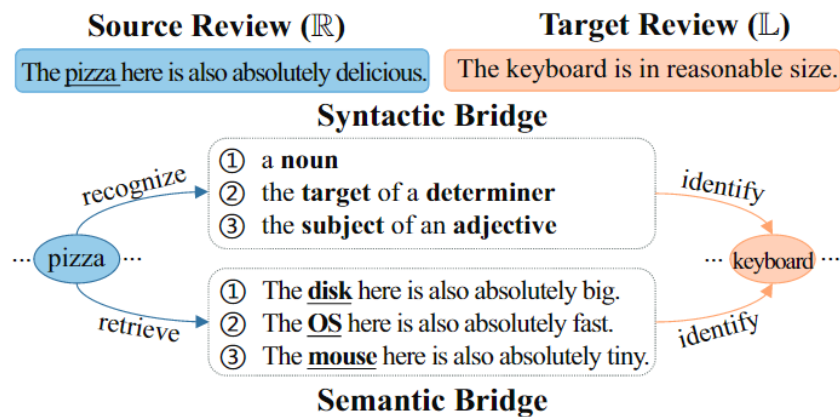


图 2: 句法和语义桥梁图解

为了完成 ATE 任务，首先要构建句法桥和语义桥。假设给定一个来自两个域的评论句子 x ，用一个查找表 (lookup table)，用 $E \in R^{de \times n}$ 来映射它，并生成单词嵌入 $E = \{e_i, ..., e_n\} \in R_{de \times n}$ ，其中 $|V|$ 是词汇量大小， de 是嵌入维度。对于跨域 ATE，构建了用于评审的桥梁，以帮助跨两个域直接转移属性词。

3.2 句法桥

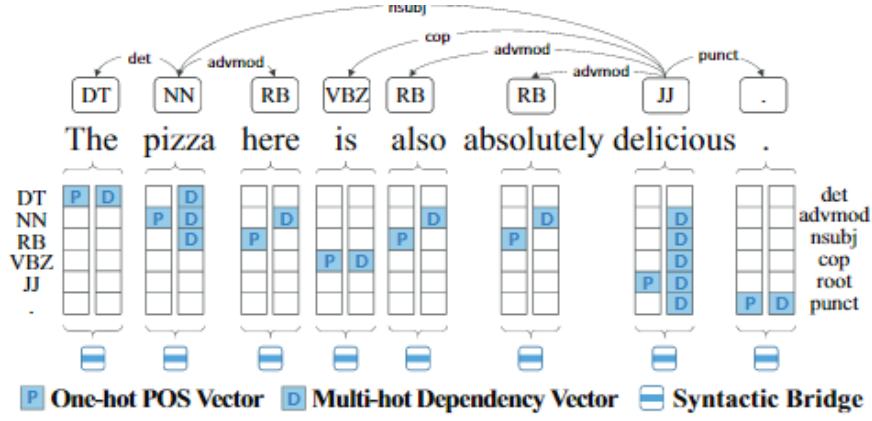


图 3: 句法桥构建示意图

在自然语言中，语言的表达方式是丰富而灵活的。相比之下，句法结构是有限的，并且是跨域通用的。基于此，复现的论文中提出基于源词和目标词的句法角色 (POS 标签和依赖关系) 而不是词汇项，在源词之间建立连接。并且，对每一个单词 x_i ，复现论文中使用 one-hot 向量 $b_{pos} \in R^{N_{pos}}$ 和 multi-hot 向量 $b_{dep} \in R^{N_{dep}}$ 来表示它的 POS 标签和依赖关系^[6]。其中， N_{pos} 和 N_{dep} 分别表示标签和关系类型的数量。对 b_{dep} ，复现论文合并所有与 x_i 有关的关系，而不管方向^[7]。为了扩大学习能力，我们将 b_{pos} 和 b_{dep} 投影到具有可学习的权重矩阵的相同维数上，并将它们连接起来形成语法桥：

$$b_{syn} = (w_{pos} \times b_{pos}) \oplus (w_{dep} \times b_{dep}) \quad (1)$$

3.3 语义桥

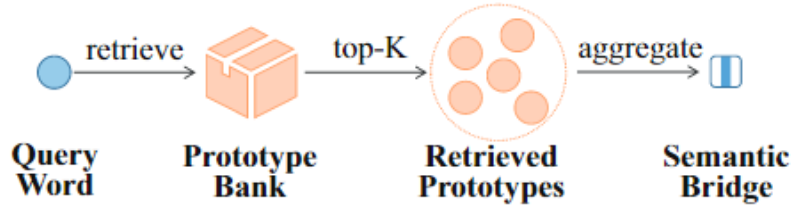


图 4: 语义桥构造示意图

与以往通过情感词或上下文构建属性词之间联系的方式不同，我们旨在建立源域和目标域属性词之间的直接联系。例如，例如，为了将知识从 D^S 中的 Pizza 转移到 D^T 中的 keyboard，将在 D^U 中引入一些补充的属性词，如 disk、OS、mouse，并直接提高其与 keyboard 的语义相关性。并且称这些补充术语为原型 (prototypes)，并将检索它们来构建语义桥。由于不同领域之间的属性词，差异比较大，因此，这些原型未必能够给适应领域自适应任务。

为了解决这个问题，复现论文设计了一个语法增强的相似性度量来检索可转移的语义原型。在开始之前，我们通过词频，过滤 D^U 中的单词。仅保留出现次数大于 r 的单词。我们将这些未标注的单词作为原型候选集并且构造一个原型库 \tilde{V} 。对于一个 query word (来自 D^S)，我们想找到一个在目标域中扮演类似句法角色的原型术语 $\tilde{v} \in \tilde{V}$ 。具体来说就是总结了 \tilde{v} 的全局用法，通过合并它在 D^S 中的 POS 和依赖嵌入，而后对 D^T 和 D^U ，采取同样的方式。这样，具有相同原型的源词和目标词就可以直接相互关联。

4 复现细节

基于上述构建的句法桥和语义桥，复现论文提出一种端到端的序列标注方法，用于 ATE 任务。论文构建的桥，有两个属性：1) 桥梁是域不变的，应该被保留下来。2) 桥梁可以帮助从 e_i 中提取域不变信息。因此，我们提出使用可传输桥 $b_{syn,i}$ 和 $b_{beq,i}$ 增强单词的词嵌入 e_i ，而后使用一个门控操作来融合桥梁。

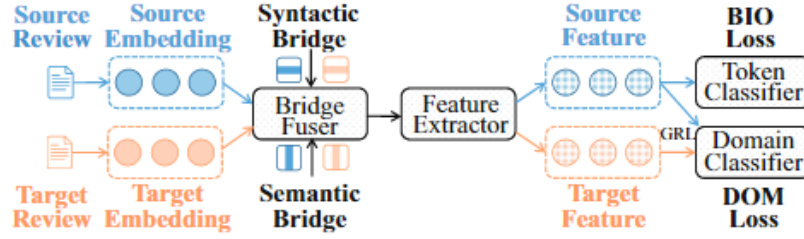


图 5: 基于桥的序列标记的训练

4.1 与已有开源代码对比

论文并没有将语法桥和句法桥结合在一起。所以说，可以考虑将句法桥和语义桥结合起来，由于句法桥和语义桥包含可传输的句法和语义信息中。具体来说，端对端的序列标准方法实现流程：

首先计算一个有维度的门 $g_{syn,i}$ ：

$$g_{syn,i} = \sigma(W_{syn}(e_i \oplus b_{syn,i})) \quad (2)$$

而使用一个包含 L 个堆叠卷积层的 ReLU 激活的 CNN 编码器来提取高级特征 $f_i = R^{d_f}$,

$$f_i^{l+1} = ReLU(f_{i-c_i:i+c}^L * K^l + b^l) \quad (3)$$

为了识别属性词和情感词，我们将最后一层的特征表示 f_i^L 输入分类器：

$$\tilde{y}_i == softmax \times (W_A \times f_i^L) \quad (4)$$

除了 BIO 标记外，复现论文还通过域对抗训练进一步增强了基于桥的特征的域不变性。具体来说，首先将 f_i^L 聚合为一个全局表示法 $f_g^{[8]}$ ：

$$f_g = MaxPool(f_{1:n}^L) \quad (5)$$

将梯度反转层 (GRL) 添加到具有尺度系数 λ 的 f_g 中，并训练一个域分类器来区分 f_g 所属的域。在训练中，只有来自 D^S 的样本具有相应的 BIO 标签 y_S 用于标记分类。目标是最小化识别方面术语的标记损失 loss：

$$\mathcal{L}_{BIO} = - \prod_{\sum_{D^S} \sum_{i=1}^l} (\tilde{y}_i, y_i) \quad (6)$$

利用 D^U 和 D^S 中的样本对域分类器进行训练，并最小化 loss。

4.2 实验环境搭建

配置好相应的 pytorch 环境，安装相应的 python 库，numpy、sklearn、pandas 等。

4.3 创新点

由于大多数方面术语都是特定于领域的，因此它们不能直接转移。在复现论文中，提出了一种新颖的主动域适应方法。作者目标是通过积极补充可转移知识来转移方面术语。为此，通过将句法角色

识别为枢纽而不是枢纽的链接来构建句法桥梁。还通过检索可转移的语义原型来构建语义桥梁，基于桥的序列标记设计了一个轻量级的端到端标记器。

但是，文章并没有将语法桥和句法桥结合在一起。所以说，可以考虑将句法桥和语义桥结合起来，由于句法桥和语义桥包含可传输的句法和语义信息，因此将它们组合起来以获得比任何一个单独模型更好的性能是很直观的。考虑应用一个非常简单的组合操作，对于嵌入 e_i 的词 x_i ，我们首先得到它的句法和语义桥 $b_{syn,i}$ 和 $b_{sem,i}$ ，并将它们合并成一个组合桥。

5 实验结果分析

5.1 消融实验

本论文进行了一系列消融研究以验证我们方法的有效性。结果如图 6 所示。

Index	Model	Variant	AVG.
1	BaseTagger	$-\mathcal{L}_{DOM}$	1.94
2		CNN \rightarrow BiLSTM	8.47
3	SynBridge	$-b_{pos}$	1.68
4		$-b_{dep}$	1.49
5		$b_{dep}\rightarrow$ Tree-LSTM	3.97
6		$b_{dep}\rightarrow$ GCN	4.21
7	SemBridge	$-c(e, \tilde{e})$	1.82
8		$-c(b_{pos}, \tilde{b}_{pos})$	2.30
9		$-c(b_{dep}, \tilde{b}_{dep})$	2.52

图 6: 消融实验结果

结果 1-2 表示 CNN 和域对抗训练都有助于整体良好的性能。结果 3-6 显示了 SynBridge 中 POS 和依赖嵌入的有效性。具体来说，在 5-6 中，我们用常用的 Tree-LSTM 和 GCN 替换我们提出的依赖结构来对依赖树进行建模，并发现性能显著下降。结果 7-9 显示了所有三种类型的相似性对于在 SemBridge 中检索原型的重要性。

5.2 参数实验

论文的方法中有三个关键超参数：GRL λ 的比例系数、频率阈值 τ 和原型 K 的数量。我们在 10⁻⁴ - 1.0 范围内改变 λ ，在 1 - 10 范围内改变 τ/K 以研究它们的影响并在图 7 中显示结果。

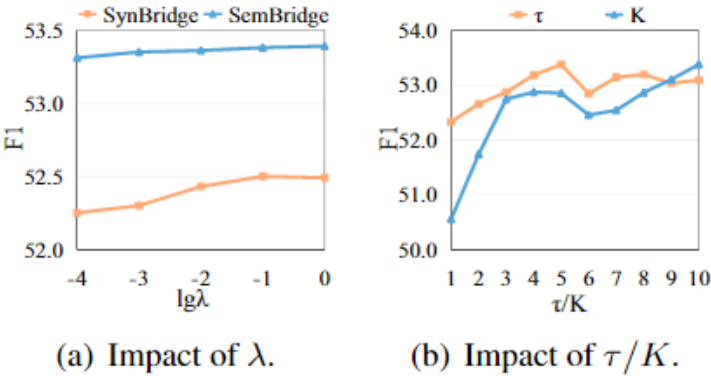


图 7: 超参数 λ 、 τ 和 K 的影响

这一结果表明，简单地强迫非枢轴点转移知识并不适合领域适应。在图 6(b) 中， τ 用于平衡多样性和准确性。低 τ 意味着原型是多样的，但其中一些是长尾词，对减少域差异贡献不大。相反，high

只保留频繁的原型，一些有意义的原型被过滤掉。因此，中间 $\tau=5$ 是一个合适的选择。对于 K ，当引入更多原型时，曲线通常向上。这种趋势是合理的，因为更多的原型等于更多的目标信息。

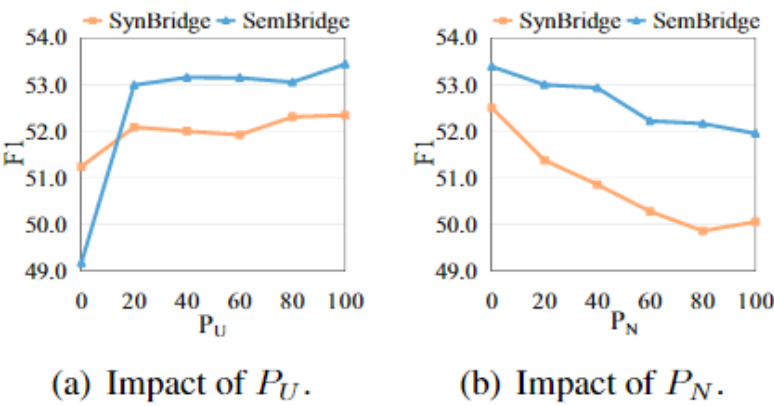


图 8: P_N 和 P_U 的影响

在图 8 中，我们进一步分析了未标记数据 P_U 的百分比和解析噪声 P_N 的百分比的影响。对于 P_U ，当引入更多未标记的目标数据时，性能通常更好。此外，约 20%~40% 的未标记数据足以实现令人满意的性能。对于 P_N ，我们手动干扰解析结果以观察我们方法的鲁棒性。显然，在解析时引入噪声后，性能开始下降，但幅度并不大。抵制解析错误有两个原因。首先，除了句法角色之外，我们还在检索原型时结合了嵌入相似性。其次，门机制可以进一步过滤无用的句法信息并保持单词表示的质量

5.3 计算成本分析

	Parameter	Runtime
TIMN	0.8M	132s
BERT-Cross	109M	84s
BaseTagger	1.3M	11s
SynBridge/SemBridge	1.4M	12s

图 9: 每种方法的计算成本

在实践中，对于任何传输对，句法语义桥的一次性构建可以在 30 秒内完成。因此，我们关注 SynBridge/SemBridge 的端到端训练成本。我们在传输对 $R \rightarrow L$ 上运行了五种性能最好的方法，并在图 9 中给出了每种方法的可训练参数数量和每个时期的运行时间。我们可以得出结论，我们提出的方法保持了相当低的计算成本。

6 总结与展望

在本文中，复现了一种用于方面项提取的新的主动域自适应方法。与之前通过将方面术语与枢轴关联来进行被动域适应的研究不同，我们通过为它们构建句法和语义桥梁来积极增强术语的可迁移性。然后，我们为基于桥的序列标记设计了一个轻量级的端到端标记器。

参考文献

[1] WANG W, PAN S J, DAHLMEIER D, et al. Coupled Multi-Layer Attentions for Co-Extraction of Aspect and Opinion Terms.[J]. Proceedings of the ... AAAI Conference on Artificial Intelligence, 2017.

[2] CHEN Z, QIAN T. Transfer Capsule Network for Aspect Level Sentiment Classification.[J]. Meeting of

the Association for Computational Linguistics, 2019.

- [3] LIU P, JOTY S, MENG H. Fine-grained Opinion Mining with Recurrent Neural Networks and Word Embeddings[J]. Empirical Methods in Natural Language Processing, 2015.
- [4] WU Y, ZHANG Q, HUANG X, et al. Phrase Dependency Parsing for Opinion Mining[J]. Empirical Methods in Natural Language Processing, 2009.
- [5] DING Y, YU J, JIANG J. Recurrent neural networks with auxiliary labels for cross-domain opinion target extraction[J]. Proceedings of the ... AAAI Conference on Artificial Intelligence, 2017.
- [6] XU H, LIU B, SHU L, et al. BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis[J]. Cornell University - arXiv, 2019.
- [7] WANG W, PAN S J. Recursive Neural Structural Correspondence Network for Cross-domain Aspect and Opinion Co-Extraction[J]. Meeting of the Association for Computational Linguistics, 2018.
- [8] GANIN Y, LEMPITSKY V. Unsupervised Domain Adaptation by Backpropagation[J]. arXiv: Machine Learning, 2014.