

“跨领域句法分析评测” 参赛邀请函

苏州大学自然语言处理团队

2021 年 3 月

近年来，随着深度学习技术的发展，尤其是预训练语言模型 ELMo/BERT 的提出，封闭领域规范文本上的句法分析性能已经达到了比较高的水平。但是，在跨领域文本，尤其是不规范的网络文本上，句法分析的性能会急剧下降。同时，如何在深度学习模型中利用结构化的句法信息也是 NLP 领域的重要研究问题之一，但是受到真实文本上句法性能的限制。因此，领域移植问题已经成为句法分析研究的重要挑战。为了更充分地支持句法分析领域移植研究，我们自 2017 年起标注了多领域句法数据集，并命名为汉语开放依存树库 (CODT)，目前包含 16 万句，涵盖 10 多个领域或来源的文本。2019 年，我们依托 NLPC-2019 会议，组织了第一届跨领域句法分析评测 (Peng et al., 2019)，针对“单源零样本”和“单源少样本”领域移植场景，发布了 CODT1.0 数据集，包含 4 万句、对应 4 个领域。

今年，我们依托 CCL-2021 会议，组织第二届跨领域句法分析评测，针对“多源零样本”领域移植场景，发布 CODT2.0 数据集，包含 7 万句，对应 6 个领域。其中，通过错误检测和人工校正，CODT2.0 与 CODT1.0 重合的数据的质量进一步提升。

此外，受中国中文信息学会资助，本次评测还设置了 2 万元奖金，以及获奖证书。

更多详情请查看评测网站：<http://hlt.suda.edu.cn/index.php/CCL2021>。欢迎大家参与。

组织者：李正华（苏州大学）

联系人：周明月（苏大研究生、z641295453@163.com）

李帅克（苏大研究生、skli20@stu.suda.edu.cn）

一、数据设置

在本次评测任务中，我们提供约 29K 句平衡语料 (BC)、9K 句产品博客 (PB)、10K 句产品评论 (PC)、8K 句财经资讯 (FIN)、8K 句法律资讯 (LEG) 和 3K 句网络小说 (ZX) 六个领域的的数据。其中 BC 作为基础数据，只作为源领域。后五个领域既可以作为源领域，也可以作为目标领域。另外，针对后五个领域，我们还提供了大规模的无标注数据。

标注规范：我们充分参考了哈工大依存树库标注规范和 UD 标注规范，针对汉语的特点，同时考虑标注一致性和可计算性，制定了一个详细的标注规范，包含 21 个依存标签 (郭丽娟等, 2018)。最新的标注规范请查看：[依存句法数据标注规范.pdf](#)

标注过程：所有的数据均采用严格的双人标注，不一致时由资深标注员讨论确定答案。另外，为了尽量降低人工标注的代价，大部分数据的标注均采用了基于局部标注的主动学习方法 (Jiang et al., 2018)。对于训练数据集，我们使用 TreeCRF-BERT 模型把数据补全为高质量的完整句法树 (Zhang et al., 2020)；而开发集、测试集都均为局部标注数据，因此评价时需要做小的调整，只考虑人工标注的依存弧。表 1 中记录了本次测评使用数据的分布情况。

表 1: 数据分布情况 (句子数)

	BC	PB	PC	FIN	LEG	ZX
Train	26.1K	5K	6.8K	5K	5K	1.5K
Dev	1K	1.3K	1.3K	1K	1K	0.5K
Test	2K	2.5K	2.6K	1.8K	1.9K	1K
Unlabeled	0	291K	349K	118K	150K	33K

二、任务设置

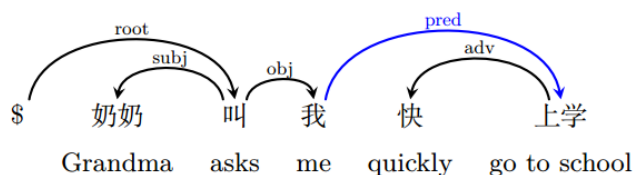


图 1: 依存句法分析示例。其中依存弧从核心词指向修饰词, “pred” 表示 “谓词” 关系。

任务描述: 输入一个自然语句, 依存句法分析任务是预测最优的依存句法树, 如图 1 所示。本评测包含 2 个子任务:

子任务一 (Closed): 多源零样本领域移植 (封闭)

子任务二 (Open): 多源零样本领域移植 (开放)

含义如下:

- **多源的含义:** 可以使用其他领域的训练数据作为源领域数据。例如, 当目标领域为 ZX 时, 可以使用其他五个领域 (包括 BC) 的训练数据, 即 BC/PB/PC/FIN/LEG-Train。

- **零样本的含义:** 不能使用对应领域的训练数据。例如, 当目标领域为 ZX 时, 那么 ZX-Train 不可作为训练数据使用。

- **Closed 赛道的要求:**

- 1) 只能使用我们提供的数据和信息, 包括: 所有数据我们都提供分词、自动词性; 我们会提供 pre-trained word embedding (在 Chinese Gigaword 3 和所有领域 Train/Unlabeled 数据上训练 word2vec 得到)。
- 2) 不可以使用其他开源工具得到新的信息, 如词性、词义等特征。
- 3) 不可以使用其他的资源, 如词典、句法语义树库等。
- 4) 不可以使用大规模无标注数据上训练的预训练语言模型 (任何开源的通用语言模型), 但是可以通过预训练语言模型的技术来使用我们提供的无标注数据。即可以使用在我们提供的有标注和无标注数据上从头开始训练预训练模型。
- 5) 不可以使用 NLPCC-2019 CODT1.0 的数据。

- **Open 赛道的要求:**

- 1) 可以使用任何可通过某种公开方式获取（包括购买）的数据、资源或工具。但是建议参赛者在系统报告中明确说明使用的外部资源以及这些外部资源对性能的影响。

- **注意事项**

- 1) **开发集的使用**：不允许将 Dev 加入到训练数据中。Dev 数据只可以用来调参和选择模型。
- 2) **无标注数据（Unlabeled）的使用**：不限制。同时，6 个领域的 Train 中的句子，只要不使用句法标注信息，也都可以作为无标注数据使用。
- 3) **预训练语言模型微调**：我们基于 Peng et al. (2019)中描述的预训练语言模型微调的方法，将 Google 发布的 bert-base-chinese 在所有无标注数据（6 个领域的 Train，5 个领域的 Unlabeled 数据）进行多轮微调（3 轮），即继续使用语言模型损失函数继续训练。这个模型仅限于 Open 赛道上的系统使用。我们做这个事情的目的是方便大家试试这个方向，避免大家浪费资源，重复训练。

如果您对规则有疑问，请随时联系我们确认。如果违反规则，发现后，成绩会直接取消。

三、结果提交和评价

提交结果时，请将测试文件打包（zip/tar.gz）发送给周明月同学（联系方式如下）。压缩包内请按照如下目录/文件命名和组织形式（没有参加的子任务忽略即可），方便我们后续处理。请参考 Dev 数据及 Readme，确保每个文件严格采用 CoNLL 格式。

```
subtask1-closed/PC-Test.out.conll
subtask1-closed/PB-Test.out.conll
subtask1-closed/ZX-Test.out.conll
...
subtask2-open/PC-Test.out.conll
subtask2-open/PB-Test.out.conll
subtask2-open/ZX-Test.out.conll
...
```

我们使用 labeled attachment score（LAS，即核心词及其依存关系标签都预测正确的词语百分比）。对于任何一个子任务，所有参赛队伍必须提交 5 个目标领域的测试结果，从而得到 5 个 LAS 值。我们直接对 5 个 LAS 求平均，确定最终排名。

注意，开发和测试集合大部分句子是局部标注的，评价时需要做简单的调整，只考虑有人工标注数据的依存弧（head 不为-1）。您也可以参考我们在网站上提供的 evaluate.py 训练集中，我们还提供了自动补全的依存弧的概率，对应第 CoNLL 格式的 9 列（从 0 开始算起）。如果是人工标注的依存弧，那么对应的概率为 2.0。

参考文献

Yu Zhang, Zhenghua Li, Min Zhang. 2020. Efficient Second-Order TreeCRF for Neural Dependency Parsing. Proceedings of ACL, pp. 3295-3305. Online. 2020.([pdf](#))

Xue Peng, Zhenghua Li, Min Zhang, Rui Wang, Yue Zhang, Luo Si. 2019. Overview of the NLPCC 2019 Shared Task: Cross-Domain Dependency Parsing Proceedings of NLPCC-2019, pp. 760-771. Dunhuang, China, 9-14 Oct. 2019.([pdf](#))

郭丽娟, 李正华, 彭雪, 张民. 2018. 适应多领域多来源文本的汉语依存句法数据标注规范. 中文信息学报. 2018, 32(10):28-35-52 ([pdf](#))

Xinzhou Jiang, Bo Zhang, Zhenghua Li, Min Zhang, Sheng Li, Luo Si. 2018. Supervised Treebank Conversion: Data and Approaches. Proceedings of ACL, pp. 2706-2716. Melbourne, Australia. 15-20 Jul. 2018 ([pdf](#))

Yue Zhang, Zhenghua Li, Jun Lang, Qingrong Xia, Min Zhang. 2017. Dependency Parsing with Partial Annotations: An Empirical Comparison. Proceedings of IJCNLP, pp. 49-58. Taiwan, Nov. 27 - Dec. 1, 2017 ([pdf](#))