

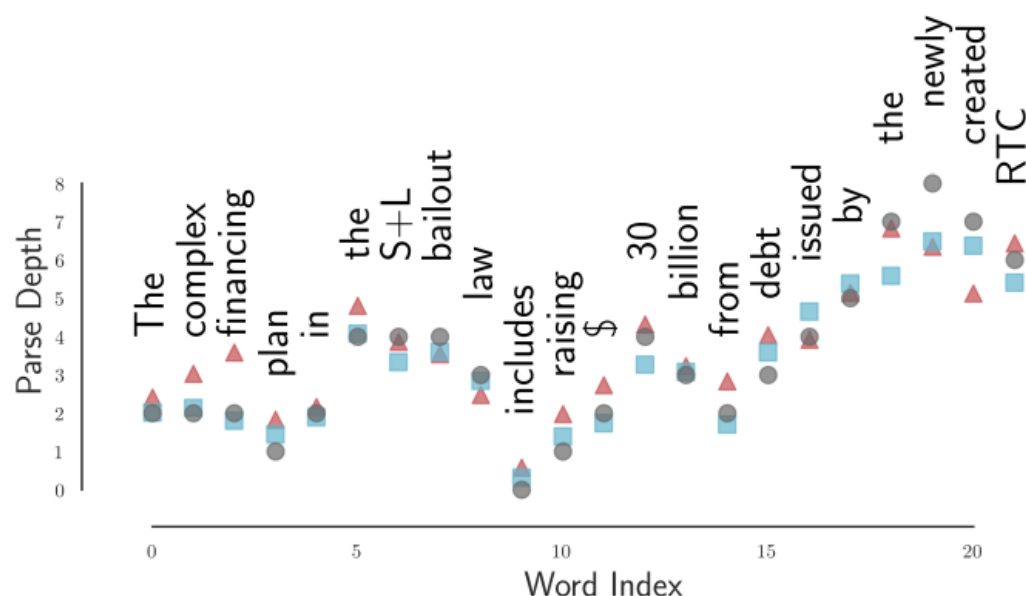
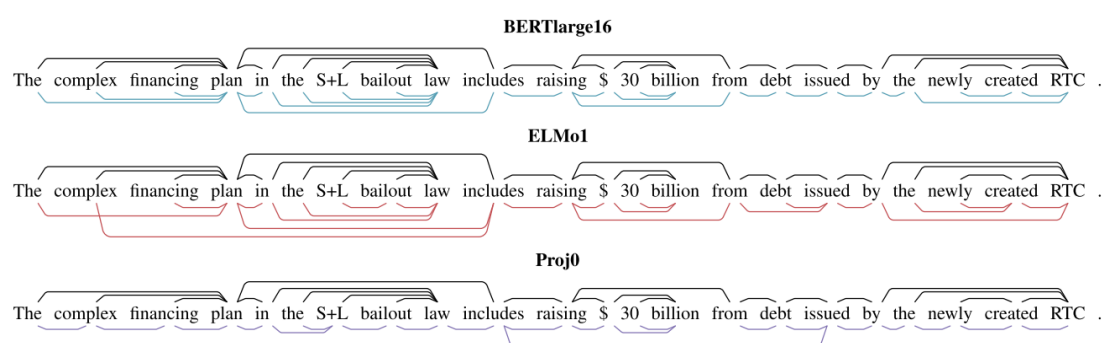
# 发现语法树

- Word2Vec 的词向量满足矢量相加
- BERT或ELMO 给出的词向量编码了上下文信息，不再满足矢量相加
- <https://arxiv.org/pdf/1906.02715.pdf> 【1】和<https://nlp.stanford.edu/pubs/hewitt2019structural.pdf> 【2】两篇文章发现上下文相关的词向量学到了语法的树结构

## A Structural Probe for Finding Syntax in Word Representations

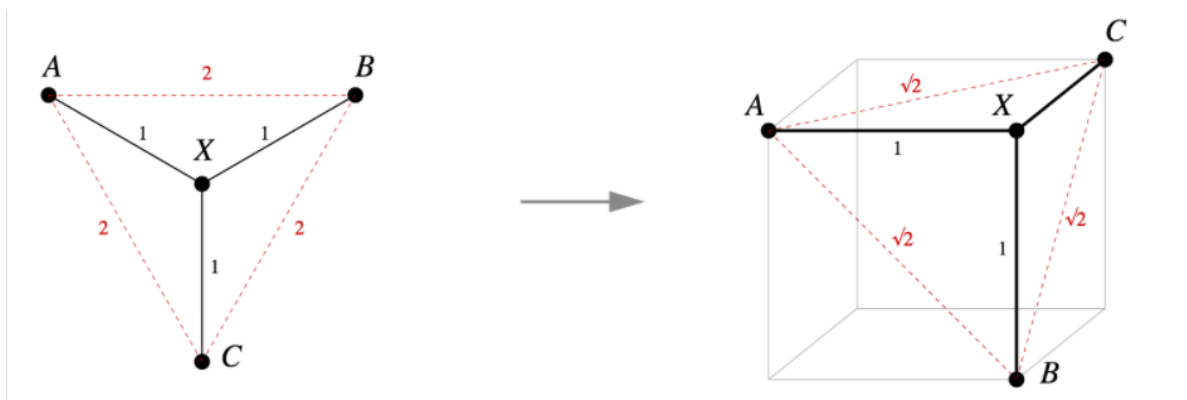
【2】这篇文章发明了一种结构探针“structural probe”，计算词向量在高维语义空间（比如1024维）中的距离，并将其映射到语法树中词与词之间的距离。从下图看出，BERT 词向量几乎完美学到了语句的语法树结构。

黑线为标准语法树，下方蓝线为BERT模型词向量学到的语法树

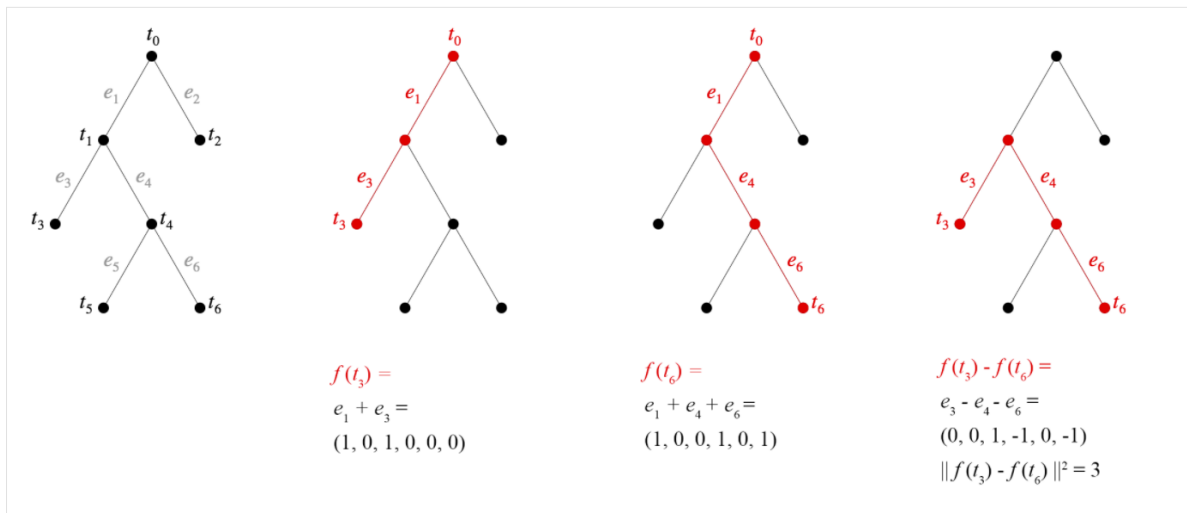


## 语法树中词与词之间的距离 Tree Distance

- 语法树中词与词之间的距离满足勾股定理，即 Squared L2 distance
- 下图右边树的节点A和B之间距离计算公式： $|AB| = \sqrt{|AX|^2 + |XB|^2}$ 。其中，语法树上的 4 个节点是3维立方体上的顶点。用三维坐标可以表示为， $A = (1, 0, 0), X = (0, 0, 0), B = (0, 1, 0), C = (0, 0, 1)$ 。
- <https://pair-code.github.io/interpretability/bert-tree/> 【3】中给出证明，任何 N 个节点的树，都可以嵌入到 N-1 维的超立方体



- 对于更复杂的语法树，【1】【3】给出了树距离的一般计算公式，即每条边都对应一个新的与之前所有边都垂直的维度



## 高维空间词向量之间的距离

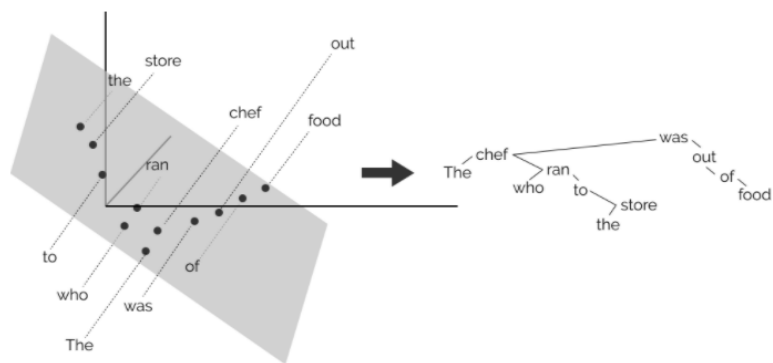
知道了计算语法树中两个词  $w_i$  与  $w_j$  之间的树距离  $d_{\text{tree}}(w_i, w_j)$ 。接下来计算两个词的词向量  $h_i$  与  $h_j$  在高维隐空间的距离。

只要词向量  $h_i$  与  $h_j$  之间的距离表现出勾股定理的模式(别名欧几里德距离，L2距离)，则可说明 BERT 词向量学到了语法树结构。

计算两个1024维向量之间的距离，最简单的公式是内积距离： $d = (h_i - h_j)^T (h_i - h_j)$ 。但这并不是唯一的距离计算公式。一般来说，只要满足三个条件，都是可行的距离定义，

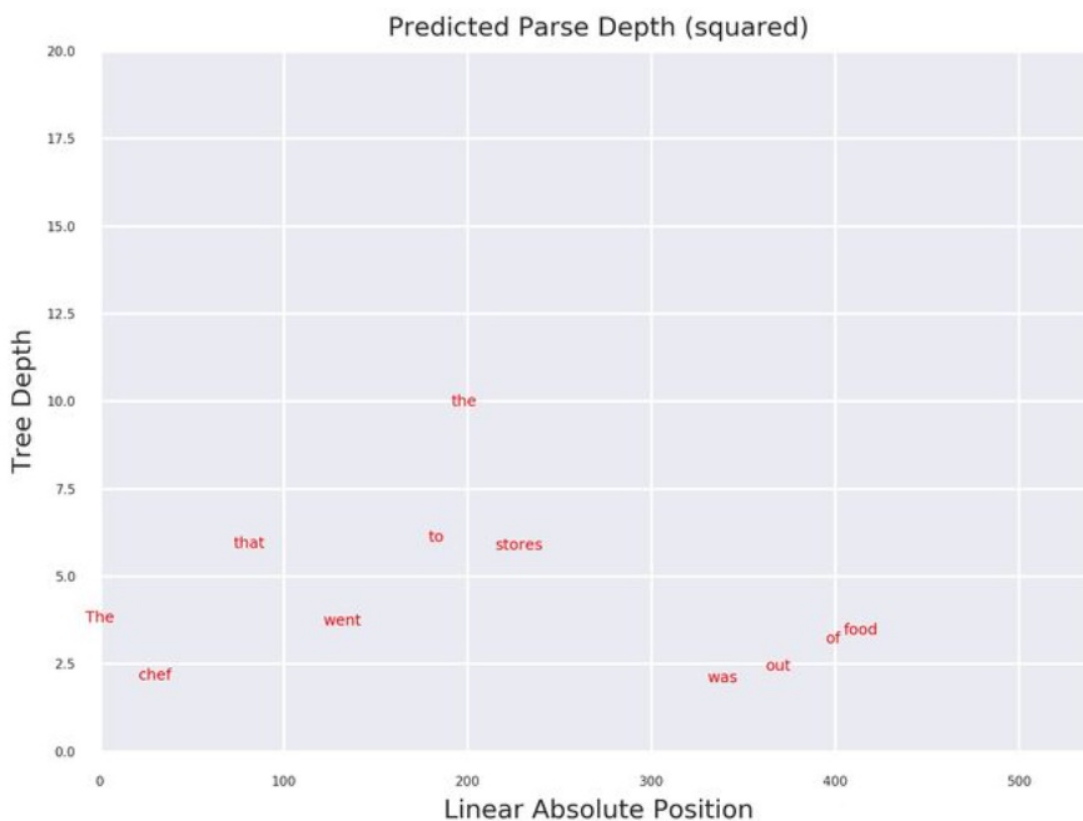
- 半正定  $d(x, y) \geq 0$
- 交换对称  $d(x, y) = d(y, x)$
- 三角不等式  $d(x, y) + d(y, z) \geq d(x, z)$

## 【2】 github



## structural-probes

The chef who ran to the stores is out of food



## 总结

- 以上文章从一个角度研究 BERT 词向量究竟学到了什么。结果是，语法树结构编码在 BERT 词向量中。
- 但BERT 词向量是1024维空间的向量，语法树只是这个巨大空间的低维投影。
- 如果将BERT词向量向其他方向投影，定义新的距离测度，揭示 BERT 词向量学到的其他内容？

# 【Paper-2021.naacl-main.146】 Does syntax matter? A strong baseline for Aspect-based Sentiment Analysis with RoBERTa

引入句法信息和图神经网络是近期细粒度情感分析（Aspect-based Sentiment Analysis, ABSA）任务研究的热门。

然而，不同的依存树(图)结构在ABSA模型中有怎样的效果并不确切。这篇论文针对这个问题设计了大量的实验。实验结果表明：

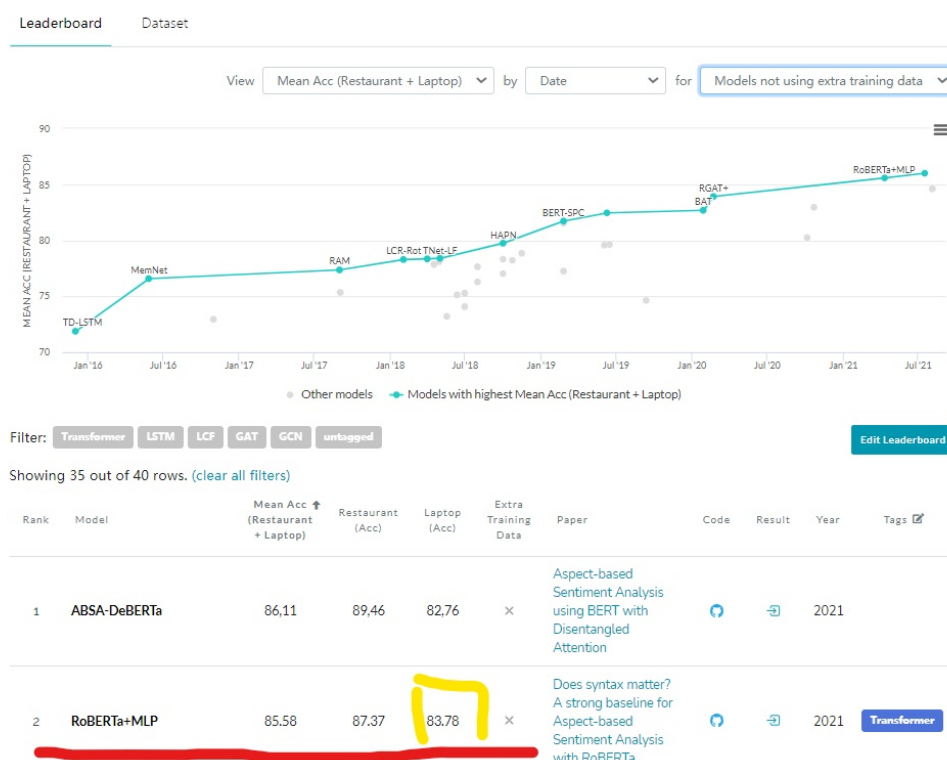
1. **不同的树结构对ALSC任务的性能影响有限。**论文在四种语言的六个数据集上进行了实验。其中，表现最好的树结构（从预训练模型中诱导出的树结构）相比句法依存树在英文数据集上有约1%的提升。
2. **从任务微调后的预训练模型中诱导的依存树可能比语言学家定义的句法依存树更适应任务本身。**在分析环节中，作者提出了两个有趣的评估指标，指标的结果都指明依存树在微调过程中发生了趋同性的改变。
3. **直接基于预训练模型进行微调**，就可以利用到模型中隐式蕴含的依存树。作者据此提出了简单有效的基于RoBERTa的baseline 模型。
4. 在paperwithcode的ABSA排行榜上，RoBERTa+MLP的性能结果在**不包含额外训练数据**的模型中取得**第1（第2）**的成绩。

Browse > Natural Language Processing > Aspect-Based Sentiment Analysis > SemEval 2014 Task 4 Sub Task 2 dataset

## Aspect-Based Sentiment Analysis on SemEval 2014 Task 4 Sub Task 2



# Aspect-Based Sentiment Analysis on SemEval 2014 Task 4 Sub Task 2



## 任务介绍

- Aspect-based Sentiment Analysis (ABSA)旨在针对句子中的包含的aspect term进行情感分析。
- 其中，aspect term是句子的文本子序列。
- ABSA包含Aspect Extraction (AE) 以及 Aspect-Level Sentiment Classification (ALSC) 两个子任务。
- 两个子任务并非完全割裂。有许多研究工作尝试同时解决整体两个子任务，也有丰富的研究工作只聚焦在其中某一个子任务上。
- 这篇论文聚焦在ALSC任务上。

## 出发点

- ALSC的一个热点研究方向是将句法依存树的树(图)结构类型的信息引入到ALSC模型中。涌现了许多基于(句法依存)树的ALSC模型大多数是通过GNN来糅合句法依存树。
- PTMs可以被探测 (probe) 出具有树(图)结构的信息。
- 研究者希望句法依存树的引入能够“指导”模型更好地结合aspect与其对应的opinion信息，从而提升情感分类的性能

## 提出问题

- 当结合不同的基于(句法依存)树的ALSC模型时，PTMs中诱导 (induce)得到的树，与依存句法分析器得到的句法依存树相比，是否有性能的差异？
- 当PTMs在ALSC任务上微调 (fine-tuning) 时，内部蕴含的树(图)结构是否会向ALSC任务适应调整？

# 实验

- 论文选取了三个典型的基于(句法依存)树的ALSC模型，**ASGCN**，**PWCN**以及**RGAT**，分别对应了三种利用(句法依存)树的方法: 基于树的结构，基于树的距离以及同时基于树的结构与距离。
- 例如，ASGCN模型只利用句法依存树给定的节点之间的依存关系，也即某节点与哪些节点连接。而PWCN利用树的距离，也即其他节点到某节点的树上的距离。
- **对于如何从PTMs中诱导树，论文选择了Perturbed Masking方法**
- 作者从预训练的BERT和RoBERTa中诱导出树，统称为“PTMs Induced Tree”，分别称为“BERT Induced Tree”和“RoBERTa Induced Tree”；
- 从在ALSC任务微调后的BERT和RoBERTa中诱导出树，统称为“FT-PTMs Induced Tree”，分别称为“FT- BERT Induced Tree”和“ FT-RoBERTa Induced Tree”。
- 加上依存句法分析器得到的句法依存树，称为“Dep.”，以及“Left-chain”和“Right-chain”。
- 一共实验了七种树结构。这七种树结构依次与ASGCN，PWCN，RGAT结合进行实验

# 分析

1. FT-RoBERTa Induced Tree能取得最好的结果，与Dep.相比有约1%的性能提升。
2. PTMs Induced Tree与Left-chain和Right-chain的性能都比较接近。而与Dep.相比，结果相当或者更低。这也回答了背景中提出的问题一。
3. 模型结合模型微调后诱导得到的树，与结合预训练模型直接得到的树相比，会得到性能的提升。

为了解释性能上的差异，论文提出了两个评估指标。

## 1. Neighboring Connections

1. 该指标在一个层面指明了FT-PTMs Induced Tree的树结构往往**更加多样**。
2. 而PTMs Induced Tree则有相当占比的Neighboring Connection。
3. 这也提供了FT-PTMs Induced Tree对PTMs Induced Tree的提升的可能来源，同时也是FT-RoBERTa甚至能获得比Dep.更好性能的原因

## 2. Aspects-sentiment Distance

1. 就是不同类型的树结构中aspect与opinion/sentiment words之间的距离
2. 数值越小表示aspect与opinion在树上的距离更短

总之，在ALSC上的微调使得PTMs更加sentiment word-oriented，也更适应ALSC任务。

# 总结

- 论文主要针对ALSC模型中的句法依存树进行了分析实验，指出**从任务适应的预训练模型中诱导的依存树可能比语言学家定义的句法依存树更适应模型本身**。
- 实验中指出使用FT-Roberta会取得1%的性能提升。然而与复杂的诱导树的步骤相比仍然需要权衡。
- **作者不推荐在ALSC模型中简单堆砌使用GNN以糅合句法依存树**。
- 基于RoBERTa的MLP模型与SOTA相比，已经极具竞争力。

# 复现

---

- Rest14
  - Acc.: 87.67(87.37) (max 87.52)
  - F1 : 81.51(80.96) (max 81.61)

```
In Epoch:16/Step:872, got best dev performance:  
AccuracyMetric: acc=0.876786  
ClassifyFPreRecMetric: f=0.815147, pre=0.837241, rec=0.798862
```

- Laptop14
  - Acc.: 83.85(83.78) (max 84.16)
  - F1 : 81.16(80.73) (max 81.18)

```
In Epoch:8/Step:298, got best dev performance:  
AccuracyMetric: acc=0.838558  
ClassifyFPreRecMetric: f=0.81164, pre=0.805035, rec=0.824178
```