

Direct parsing to sentiment graphs

David Samuel,¹ Jeremy Barnes,² Robin Kurtz,³ Stephan Oepen,¹ Liljavrelid¹ and Erik Velldal¹

摘要

This paper demonstrates how a graph-based semantic parser can be applied to the task of structured sentiment analysis, directly predicting sentiment graphs from text. We advance the state of the art on 4 out of 5 standard benchmark sets. We release the source code, models and predictions.

1 引言

论文的主要研究任务是结构化情感分析（SSA）。根据近年来的研究，依赖解析方法在 SSA 任务上取得了一定的效果，但是这些方法也存在一定局限性，文章据此提出了一种新的情感解析器——直接从文本预测情感图，不再依赖启发式有损转换到中间依赖项的表示。最终在来自四种不同语言的五个数据集上对模型进行了评估，发现在情感分析上达到了更优的效果。

2 相关工作

2.1 文本解析

2.1.1 解析算法

根据 Dozat 和 Manning(2017) 的基于 LSTM 的句法解析器，句法依赖注释集中在句子的表面或功能结构上。^[1]在此基础上，对句法解析器训练和生成图结构的研究进行了扩展，提出语义依赖注释，旨在使用图形结构表示来捕获与句子含义更密切相关的词间关系。^[2]由此产生的系统本身就达到了最先进的性能，比以前更复杂的最先进系统高出 0.6%，标记为 F1。添加语言上更丰富的输入表示将使性能更优，能够以 1.9% 的标签 F1 击败以前的系统。

2.1.2 依赖解析方法

结构化情感分析试图从文本中提取完整的意见元组，但随着时间的推移，该任务已细分为越来越小的子任务，例如，目标提取或目标极性分类。这种划分已经适得其反，通过提出了一个新的统一框架可以补救这种情况。将结构化情感问题转化为依赖图解析，其中节点是情感持有者、目标和表达的跨度，弧是它们之间的关系，并对四种语言（英语、挪威语、巴斯克语和加泰罗尼亚语）的五个数据集进行了实验，并表明这种方法比最先进的基线有很大的改进。分析表明，使用句法依赖信息改进情绪图可进一步改善结果。^[3]

稀疏模糊注意机制^[4]：注意评分者在语义和语法依赖分析等解析任务中取得了成功。然而，在建模为解析的任务中，如结构化情感分析，“依赖边”非常稀疏，这阻碍了解析器的性能。因此，有研究者提出了一个具有池化层的稀疏模糊注意力评分器，它提高了解析器的性能，并为结构化情感分析设置了最新的技术。并且进一步探讨了基于二阶解析的结构化情感分析的解析建模，引入了一种新的稀疏二阶边缘构建过程，从而显著提高了解析性能。

2.2 基于图的解析器

Samuel and Straka 提出了 PERIN, 这是一种用于句子到图形语义解析的新的排列不变方法。PERIN 是一种多功能、跨框架和语言独立的架构, 用于语义结构的通用建模。这个系统参与了 CoNLL 2020 共享任务, 即跨框架含义表示解析 (MRP 2020), 在该任务中, 它在四种语言的五个不同框架 (AMR、DRG、EDS、PTG 和 UCCA) 上进行了评估。PERIN 是共同任务的获胜者之一。^[5]复现所选择的这篇论文也是基于 PERIN 结构进行了一定的优化, 具体优化工作将在 3.3 节进行具体阐述。

3 本文方法

3.1 本文方法概述

本文复现的工作主要任务是 SSA (结构化情感分析), 也就是定位句子中的所有观点元组, 具体示例如图 1 所示:

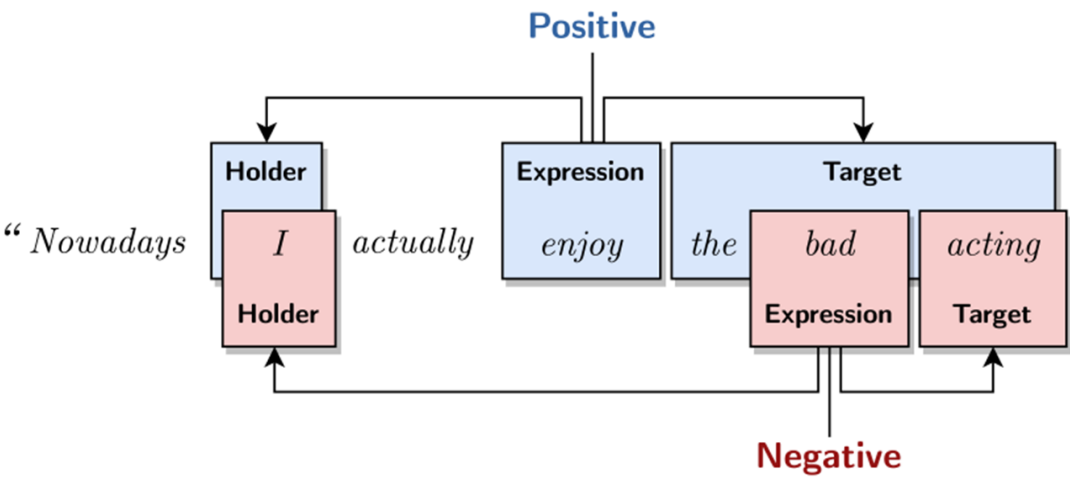


图 1: 两个对立观点嵌套

一个观点包含 a) 极性表达, b) 可选的表达者, c) 可选的情感目标, d) 积极的, 消极的或中性的极性, 其中 I 是表达者, enjoy/bad 是极性表达, acting 是情感目标, 根据极性表达可以判断出积极、消极或者中性的极性, enjoy 就属于积极的极性, bad 属于消极的极性。在图 1 的例子中, 存在两个嵌套的观点, 按照 2.1.2 的依赖解析方法, 很难确定消除两个目标的歧义, 确定 bad 是仅用于表达积极意见目标的一部分, 所以本文去掉了中间的依赖解析表示, 选择直接从文本预测情绪图。具体实现结构见 3.3 节

3.2 数据集

关于数据集, 使用的是 2022 版开源的 MPQA (用英文注释新闻专线文本)、DSU (注释了网络大学的英文评论)、NoReCf (挪威语专业评论的多领域数据集)、EU 和 CA (分别包含巴斯克语和加泰罗尼亚语的酒店评论), 数据集的具体参数 (包括用于训练、测试和验证的句子数量和不同情感的句子数量) 如图 2 所示:

		sentences	holders	targets	exps.	+	neu	-
NoReC	train	8634	898	6778	8448	5684	—	2756
	dev	1531	120	1152	1432	988	—	443
	test	1272	110	993	1235	875	—	358
CA	train	1174	169	1695	1981	1272	—	708
	dev	168	15	211	258	151	—	107
	test	336	52	430	518	313	—	204
EU	train	1064	205	1285	1684	1406	—	278
	dev	152	33	153	204	168	—	36
	test	305	58	337	440	375	—	65
MPQA	train	5873	1431	1487	1715	671	337	698
	dev	2063	414	503	581	223	126	216
	test	2112	434	462	518	159	82	223
DSU	train	2253	65	836	836	349	104	383
	dev	232	9	104	104	31	16	57
	test	318	12	142	142	59	12	71

图 2: 数据集参数

图 3展示了数据集包含嵌套观点的句子数和这些句子占整个数据集的百分比:

	holders		targets		expressions	
	#	%	#	%	#	%
NoReC	95	1.5	1187	14.1	1075	9.3
EU	30	2.2	79	4.5	16	0.7
CA	43	2.9	28	1.2	23	0.9
MPQA	48	2.2	250	9.3	145	5.6
DSU	0	0.0	10	1.1	7	0.5

图 3: 数据集嵌套计数

这里选用了跨语言的数据集，是根据 Dozat 和 Manning(2018) 提出的语义依赖注释研究,^[2]语言输入更丰富的情况下可以使效果更优。

3.3 模型结构

在 Samuel and Straka (2020) 提出的 PERIN 模型^[5]的基础上主要进行了两个方面的改变：只使用 PERIN 完整功能的一个子集——它不需要使用“相对标签规则”和模型节点属性或边缘属性，也就是删除了令牌、引理和 POS 嵌入；用 XLM-Roberta-base 替换原来的 mBERT 作为预训练模型，Roberta 全称是 A Robustly Optimized BERT Pretraining Approach，顾名思义，也就是 Bert 模型的优化，xlm 是 Cross-lingual Language Model Pretraining，也就是跨语言预训练模型。

模型的主要结构如图 4:

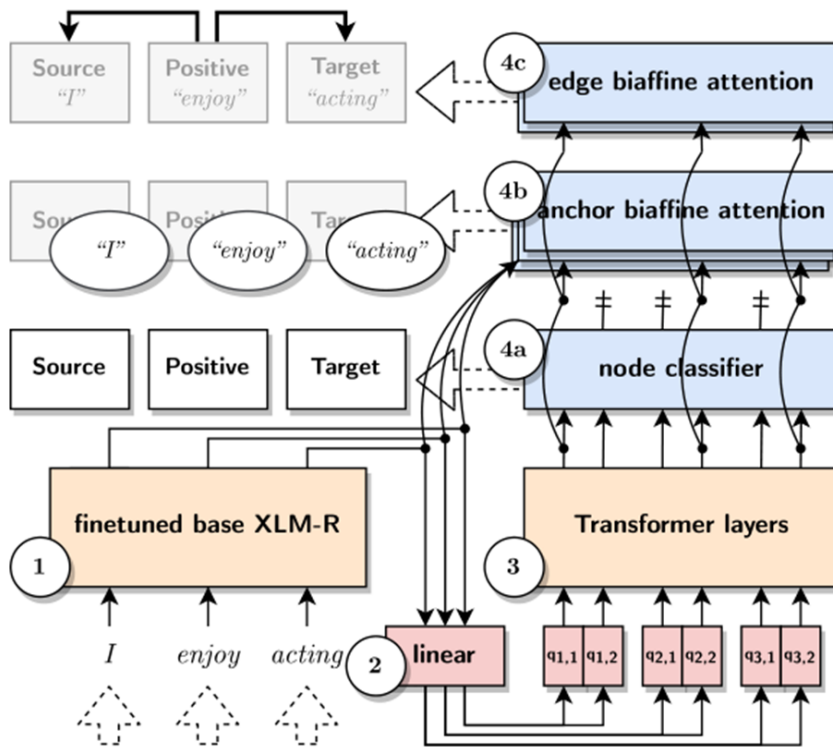


图 4: 数据集参数

当文本被输入模型时，主要经过以下四步：

- 模型选用 XLM-Roberta-base 作为预训练模型，将预训练模型的上下文嵌入与学习到的字符级嵌入相结合；
- 生成查询，每个令牌通过线性转换映射到潜在查询；
- 查询被进一步处理，一个没有位置嵌入的 Transformer(编码器) 层堆栈 (Vaswani et al, 2017)^[6] 可选地建模查询间依赖性；
- 分类头在节点上选择和标记查询，建立从节点到标记的锚定，并预测节点到节点的边缘。它们被放入 a) 节点，b) 锚点和 c) 边分类头。查询和节点间的匹配采用的是匈牙利算法，也就是在所有查询和节点之间创建一个加权二部图，目标是找到最可能的匹配，使用匈牙利算法可以在多项式时间内高效地完成。最后，就可以将每个节点分配给一个查询。

而模型训练中几个比较重要的参数如下：

Loss function : CrossEntropyLoss

parameter:

epochs: 200

n_layers: 3

decoder_learning_rate: 1.0e-4 graph encoding

encoder_learning_rate: 6.0e-6 initial encoder learning

3.4 图编码方式

文本被处理后，需要选择不同的图编码方式构建情绪图，文章一共提出了 3 种图编码方式，见图

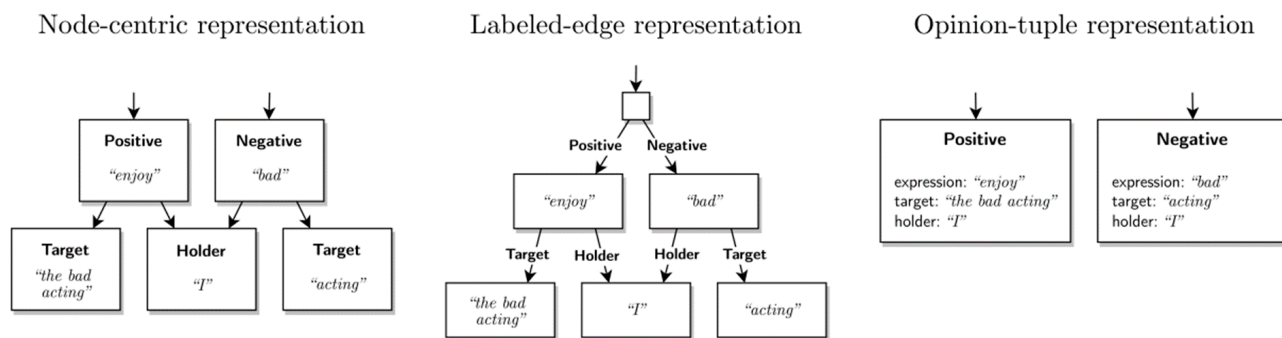


图 5: 图编码方式

三种图编码方式对应的处理流程如下：

以节点为中心编码：给节点加标签，定位情绪，增加情绪指向持有者的有向边

标记边编码：节点根据 label 分为两类，定位结点对不同的边加标签并分类

意见元组编码：将结构化的情绪信息表示为观点四元组序列 (target, sentiment, opinion holder, time)。这种编码是最受限制的，具有最低的自由度。

4 复现细节

4.1 与已有开源代码对比

原代码链接：https://github.com/jerbarnes/direct_parsing_to_sent_graph

代码架构见图 6：

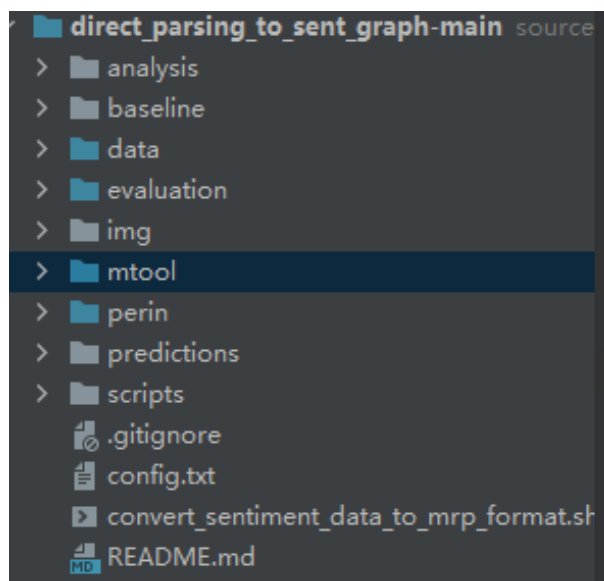


图 6: 代码架构

其中/analysis 用于分析嵌套情况；/data 主要是数据集；/mtool 是一些代码运行需要的工具，以及其安装代码；/evaluation 是模型效果评估，复现工作部分的重点就是/perin 文件夹下的模型训练部分和情绪推理部分。

4.2 实验环境搭建

代码运行环境：Pycharm+ 服务器

代码配置：Python 3.8

4.3 使用说明

在 inference.py 添加数据集，便可利用模型推理出数据集对应的情绪态度并输出情感图。

4.4 创新点

主要的创新增量是修改预训练模型 XLM-Roberta-base 为 XLM-Roberta-large，以上两个预训练模型主要差别在 transformer 层。具体优化效果可以见 5.2 节

5 实验结果分析

5.1 实验结果展示

原始数据集情况可见图 7：



图 7: 原始数据集

模型经过训练后，推理出的情绪图如图 8：



图 8: 实验结果

5.2 模型效果评估

模型评估得分如图 9:

label_edged_norec	seq_norec
'source/f1': 0.5257137816821045,	'source/f1': 0.586956015418668,
'target/f1': 0.5489923434271615,	'target/f1': 0.585333703856009,
'expression/f1': 0.5969490326642386,	'expression/f1': 0.6272580321400014,
'sentiment_tuple/unlabeled_f1': 0.4639579576025104,	'sentiment_tuple/unlabeled_f1': 0.4936582876010412,
'sentiment_tuple/precision': 0.458320086162361,	'sentiment_tuple/precision': 0.4402762949662679,
'sentiment_tuple/recall': 0.3923878306500741,	'sentiment_tuple/recall': 0.4516776940740899,
'sentiment_tuple/f1': 0.42279899081332983	'sentiment_tuple/f1': 0.44590412534575224
node_norec	label_edged_mpqa
'source/f1': 0.5384610327863594,	'source/f1': 0.5545166460857727,
'target/f1': 0.548205008678681,	'target/f1': 0.5561645655404243,
'expression/f1': 0.5857749419157302,	'expression/f1': 0.523891287696951,
'sentiment_tuple/unlabeled_f1': 0.45890021448769563,	'sentiment_tuple/unlabeled_f1': 0.3284070766468985,
'sentiment_tuple/precision': 0.43809650545810996,	'sentiment_tuple/precision': 0.36132148013105225,
'sentiment_tuple/recall': 0.38482126803376404,	'sentiment_tuple/recall': 0.2629593776953006,
'sentiment_tuple/f1': 0.409734382175707	'sentiment_tuple/f1': 0.30439143014580394

图 9: 模型评估结果

在选定 norec 数据集的基础上, 分别对三种图编码方式得到的情绪图进行了评估, 同时在标记边编码的图编码方式下分别比较了 norec 和 mpqa 的结果。

5.3 优化效果评估

选定标记边编码做为图编码方式, mpqa 作为数据集, 优化前后的评估效果可以见图 10

```
'source/f1': 0.5545166460857727,
'target/f1': 0.5561645655404243,
'expression/f1': 0.523891287696951,
'sentiment_tuple/unlabeled_f1': 0.3284070766468985,
'sentiment_tuple/precision': 0.36132148013105225,
'sentiment_tuple/recall': 0.2629593776953006,
'sentiment_tuple/f1': 0.30439143014580394

'source/f1': 0.6393239941327142,
'target/f1': 0.6455579972396308,
'expression/f1': 0.5620306806468587,
'sentiment_tuple/unlabeled_f1': 0.4369385075003583,
'sentiment_tuple/precision': 0.47480713059399077,
'sentiment_tuple/recall': 0.33171125484391134,
'sentiment_tuple/f1': 0.3905648573966605
```

图 10: 模型优化结果

6 总结与展望

目前实现的内容存在以下的不足之处:

1. 评估应该运行 5 次取平均, 实际只运行了一次, 结果代表性不够强;
2. 没有运行 baseline 做对比
3. 在结果展示上应该寻求可读性更强的方式

展望:

1. 关于对人的情绪的解析, 可以尝试达到更高的情绪粒度
2. 可以尝试基于获取到的情绪抓取关键词, 例如酒店评论, 可以跟据关键词对酒店环境进行一定的改善。

参考文献

- [1] DOZAT T, MANNING C D. Deep Biaffine Attention for Neural Dependency Parsing[J/OL]. CoRR, 2016, abs/1611.01734. arXiv: 1611.01734. <http://arxiv.org/abs/1611.01734>.
- [2] DOZAT T, MANNING C D. Simpler but More Accurate Semantic Dependency Parsing[C/OL]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Melbourne, Australia: Association for Computational Linguistics, 2018: 484-490. <https://aclanthology.org/P18-2077>. DOI: 10.18653/v1/P18-2077.
- [3] BARNES J, KURTZ R, OEPEN S, et al. Structured Sentiment Analysis as Dependency Graph Parsing [C/OL]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Online: Association for Computational Linguistics, 2021: 3387-3402. <https://aclanthology.org/2021.acl-long.263>. DOI: 10.18653/v1/2021.acl-long.263.
- [4] PENG L, LI Z, ZHAO H. Sparse Fuzzy Attention for Structured Sentiment Analysis[EB/OL]. arXiv. 2021. <https://arxiv.org/abs/2109.06719>.
- [5] SAMUEL D, STRAKA M. ÚFAL at MRP 2020: Permutation-invariant Semantic Parsing in PERIN [C/OL]// Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing. Online: Association for Computational Linguistics, 2020: 53-64. <https://aclanthology.org/2020.conll-shared.5>. DOI: 10.18653/v1/2020.conll-shared.5.
- [6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need[EB/OL]. arXiv. 2017. <https://arxiv.org/abs/1706.03762>.