BiSyn-GAT+: Bi-Syntax Aware Graph Attention Network for Aspect-based Sentiment Analysis

期刊会议	ACL(Findings)
一作	Shuo Liang
通讯	Wei Wei
单位	华中科技大学
发表时间	2022
阅读时间	@July 10, 2022
状态	In progress

BiSyn-GAT+: Bi-Syntax Aware Graph Attention Network for Aspect-based Sentiment Analysis

Shuo Liang[1] , **Wei Wei**[2,✉], **Xian-Ling Mao**[3], **Fei Wang**[4], **Zhiyong He**[5]
[1,2] Cognitive Computing and Intelligent Information Processing (CCIIP) Laboratory,
School of Computer Science and Technology, Huazhong University of Science and Technology
[3] School of Computer Science and Technology, Beijing Institute of Technology
[4] Institute of Computig Technology, Chinese Academy of Sciences
[5] Naval University of Engineering
[1] shuoliang@hust.edu.cn, [2] weiw@hust.edu.cn, [3] maoxl@bit.edu.cn,
[4] wangfei@ict.ac.cn, [5] moonmon_pub@outlook.com

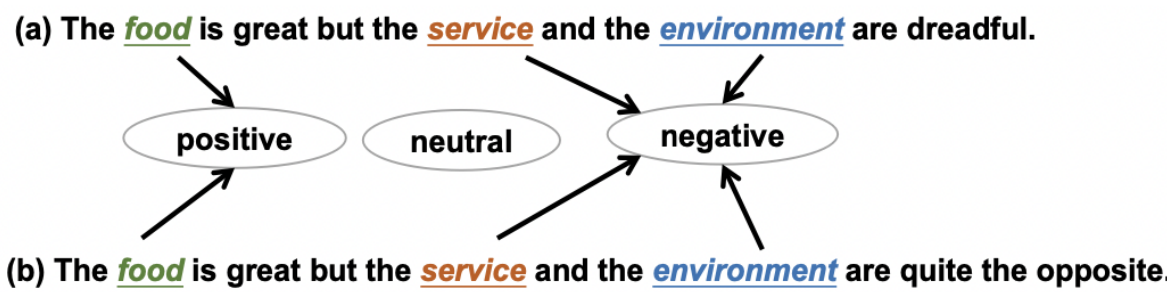
摘要

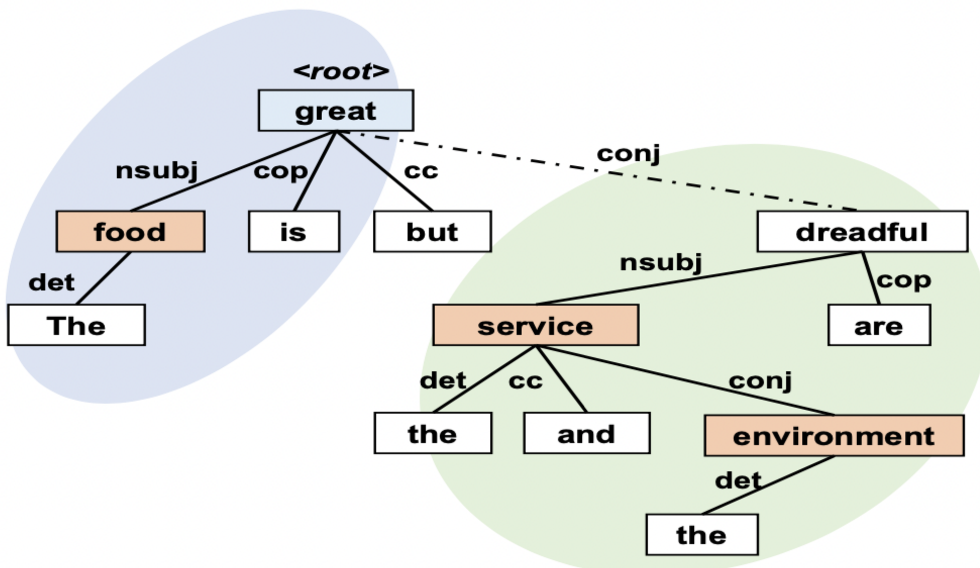
[Background] Aspect-based sentiment analysis (ABSA) is a fine-grained sentiment analysis task that aims to align aspects and corresponding sentiments for aspect-

specific sentiment polarity inference. It is challenging because a sentence may contain multiple aspects or complicated (e.g., conditional, coordinating, or adversative) relations. Recently, exploiting dependency syntax information with graph neural networks has been the most popular trend. Despite its success, methods that heavily rely on the dependency tree pose challenges in accurately modeling the alignment of the aspects and their words indicative of sentiment, since the dependency tree may provide noisy signals of unrelated associations (e.g., the “conj” relation between “great” and “dreadful” in Figure 2). **[Aim]** In this paper, to alleviate this problem, we propose a Bi-Syntax aware Graph Attention Network (BiSyn-GAT+). **[Method]** Specifically, BiSyn-GAT+ fully exploits the syntax information (e.g., phrase segmentation and hierarchical structure) of the constituent tree of a sentence to model the sentiment-aware context of every single aspect (called intra-context) and the sentiment relations across aspects (called inter-context) for learning. **[Result]** Experiments on four benchmark datasets demonstrate that BiSyn-GAT+ outperforms the state-of-the-art methods consistently.

引言

- **Identify a research area**
    - Aspect-based sentiment analysis (ABSA) aims to identify the sentiment polarity towards a given aspect in the sentence.
- **Review literature on the topic**
    - RNNs/CNNs+注意力+位置
    - GNNs+依存树
- **Point out a gap in current knowledge on that topic**
    - 位置假设不一定能使方面和上下文匹配
    - 依存树的依存关系连接无关上下文
    - 在一些句子中，上述两种方法都不能将方面及其对应上下文对齐（intra-aspect）
    - 无法对多个方面的复杂关系建模（inter-aspect）





- **Explain what you hope to do to fill the gap**
    - 成分树通常包含精确的短语分割和层次结构，有助于对齐方面及其上下文
    - 短语分割可以将复杂句子划分成多个子句
    - 层次结构可以区分aspect之间的关系
    - 提出**Bi-Syn**tax aware **G**raph **At**tention Network(**BiSyn-GAT+**)，利用依存树和成分树中的语法信息，分别用于建模方面上下文和方面之间的关系

相关工作

根据有无使用语法信息划分，语法信息指依存树或成分树

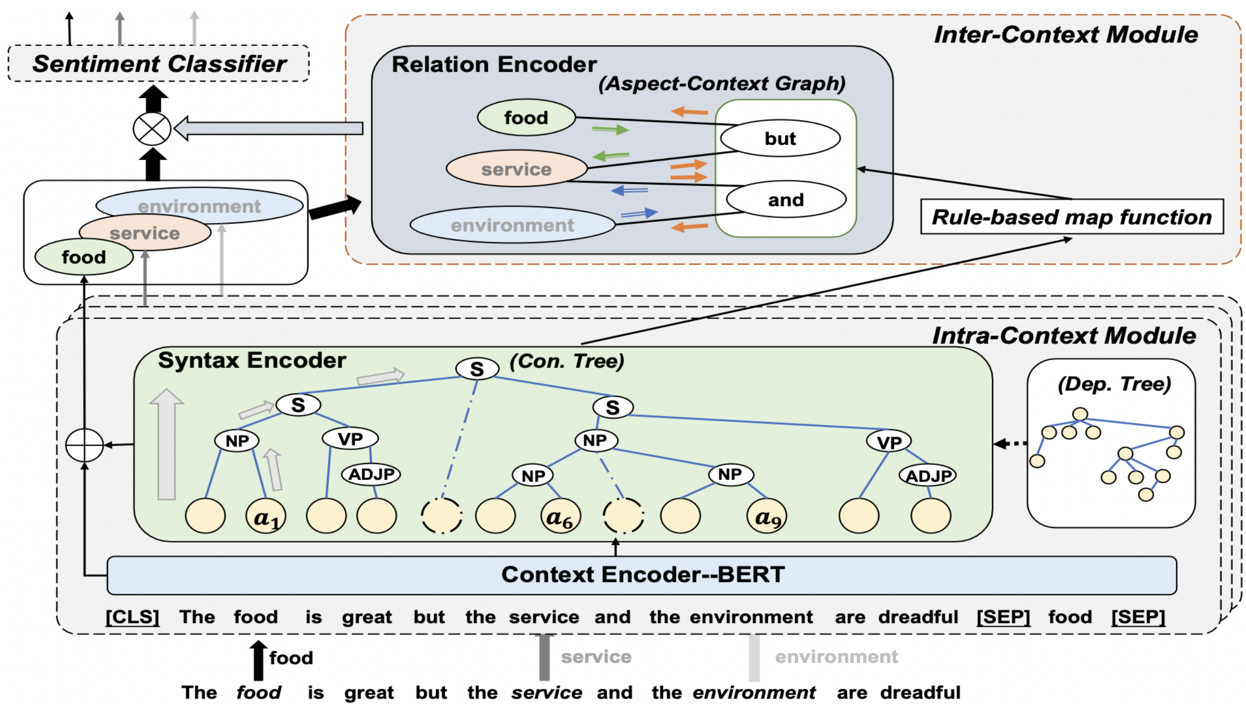
1. syntax-free methods
2. syntax-based methods

3. 对多个方面建模的工作

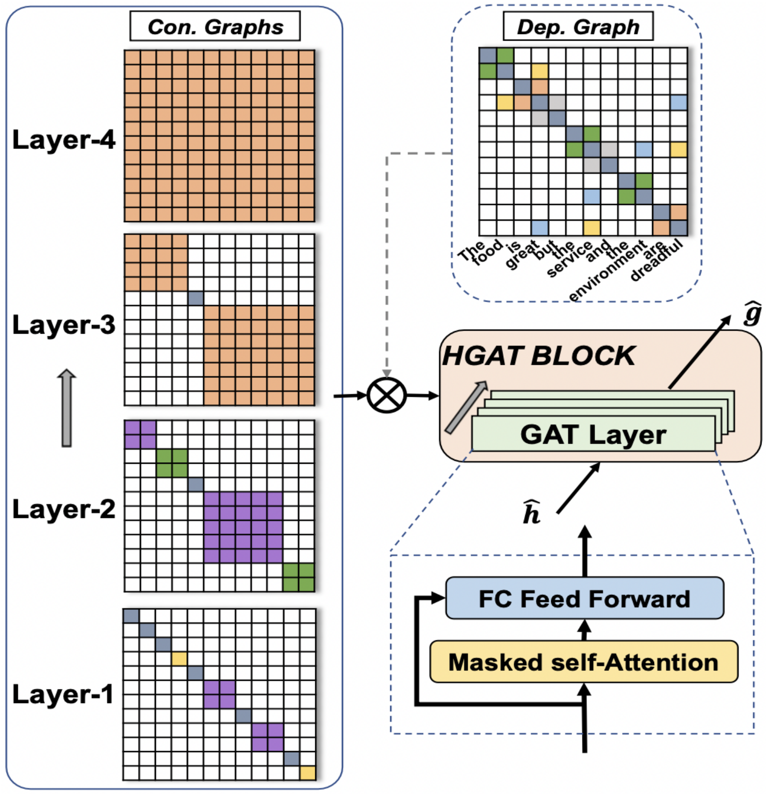
4. 使用成分树应用在ACSA任务上的SCAN

方法

整体模型



方面上下文内模块



上下文编码器

采取BERT-SPC这篇工作的方法构造输入：

$$BERT\_seq_t = [\mathrm{CLS}] + \{w_i\} + [\mathrm{SEP}] + a_t + [\mathrm{SEP}]$$

得到隐藏层：

$$h^t = \left\{ h_0^t, h_1^t, \ldots, h_{n'}^t, \ldots, h_{n'+2+m_t'}^t \right\}$$

将Bert的子词平均池化得到单词的表示：

$$\hat{\mathbf{h}}_i^t = \frac{1}{|BertT(w_i)|} \sum_{k \in BertT(w_i)} h_k^t$$

语法编码器

根据成分树的不同层得到各自的邻接矩阵：

$$\mathbf{CA}_{i,j}^{l} = \begin{cases} 1 & \text{if } w_i, w_j \text{ in same phrase of } \left\{ph_u^l\right\} \\ 0 & \text{otherwise} \end{cases}$$

提出了堆叠的层次图注意力块，Bert编码的隐藏层计算n层图注意力网络，不同数据集n设置在1到3之间。自底向上逐层聚合特征：

$$\mathbf{g}_i^{t,l} = \|_{z=1}^{Z} \sigma \left(\sum_{j \in \mathcal{N}^{t,l}(i)} \alpha_{ij}^{lz} \mathbf{W}_g^{lz} \hat{\mathbf{g}}_j^{t,l-1} \right)$$

$$\alpha_{ij}^{lz} = \frac{\exp\left(f\left(\hat{\mathbf{g}}_i^{t,l-1}, \hat{\mathbf{g}}_j^{t,l-1}\right)\right)}{\sum_{j' \in \mathcal{N}^l(i)} \exp\left(f\left(\hat{\mathbf{g}}_i^{t,l-1}, \hat{\mathbf{g}}_{j'}^{t,l-1}\right)\right)}$$

把更新前后的向量表示相加作全连接：

$$\hat{\mathbf{g}}_i^{t,l} = FC(\mathbf{g}_i^{t,l} + \hat{\mathbf{g}}_i^{t,l-1})$$

这里使用了残差网络的思想可能是考虑到成分树的深度很大，GAT的层数较多，避免梯度消失。

融合依存信息（可选）

根据依存树得到邻接矩阵：

$$\mathbf{DA}_{i,j} = \begin{cases} 1 & \text{if } w_i, w_j \text{ link directly in Dep.Tree} \\ 0 & \text{otherwise} \end{cases}$$

三种融合策略更新邻接矩阵：

1.按位点乘

$$\mathbf{FA} = \mathbf{CA} \cdot \mathbf{DA}$$

两个单词既要有依存关系，又要在同一成分内

2.按位相加

$$\mathbf{FA} = \mathbf{CA} + \mathbf{DA}$$

同一成分具有依存关系的单词得到更高的权重

3.条件按位相加

$$\mathbf{FA} = \mathbf{CA} \bigoplus \mathbf{DA}$$

先删除子句之间的依存关系，再执行按位相加

最后，该模块将上下文编码器和语法编码器的输出相加作为该每个单词的表示，后面拼接的一项为[CLS]向量：

$$\mathbf{v}_t^{as} = \left[\hat{\mathbf{h}}_t^t + \hat{\mathbf{g}}_t^t ; h_0^t \right]$$

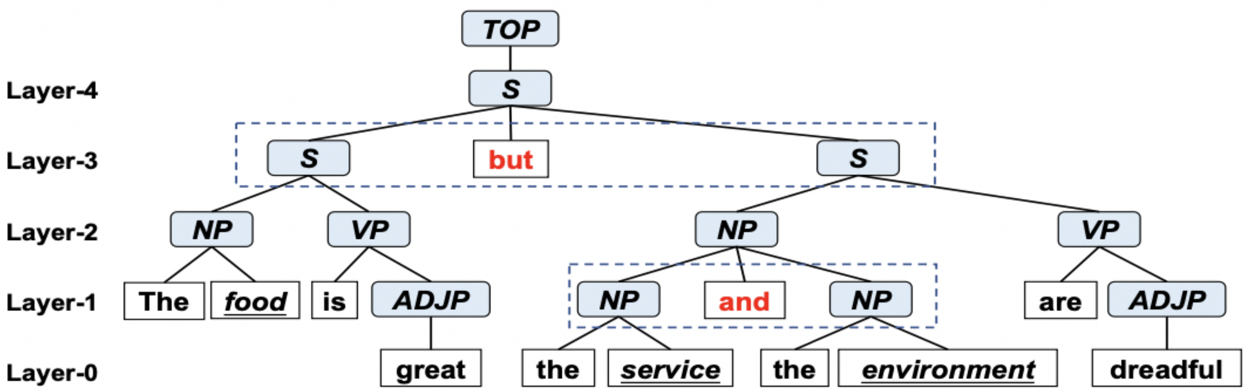
方面上下文间模块

方面分割

连词表示了方面间的关系，使用基于规则的映射函数分割两个aspect，流程如下：

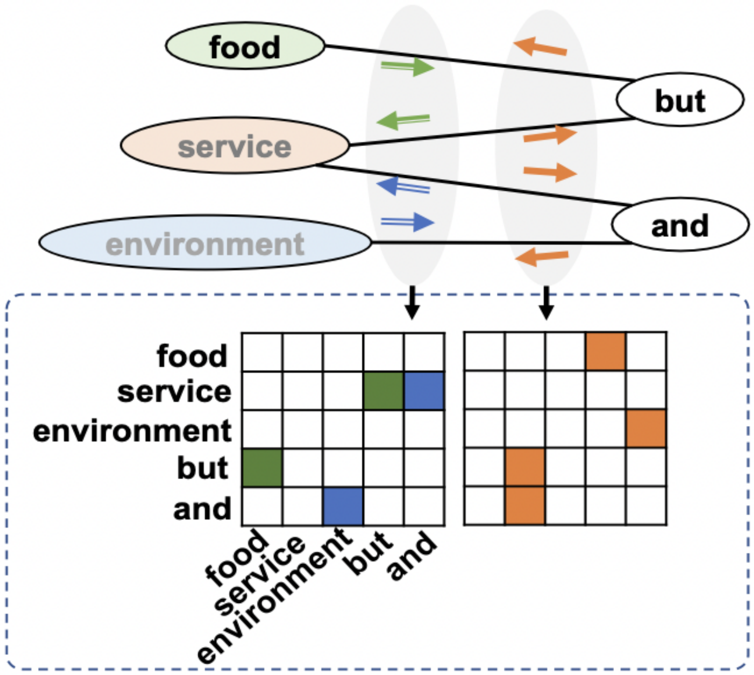
1. 找出两个方面的最小公共祖先 lowest common ancestor (LCA)
    a. 该成分包含两个方面的信息，和最少的无关信息
2. 对于LCA的子成分，如果方面A和方面B所在的成分之间存在成分（inner branch），返回该成分
3. 如果不满足第2点，返回两个方面之间的单词

$$PS(a_i, a_j) = \begin{cases} \{w_k\}, & \text{if } |Br(a_i, a_j)| = 0 \\ Br(a_i, a_j), & otherwise \end{cases}$$



关系编码器

作者认为一个aspect的影响范围应该是连续的，而aspect的相互影响随着距离的增加而衰减。为了减少计算，仅建模相邻aspect的关系



关系建模是有向的，设计了两个方面-上下文图分别建模两个方向。第一个学习奇数索引方面到相邻偶数索引方面的影响，第二个则是学习偶数对奇数的影响。两个图都是使用上面提到的层次图注意力块进行计算，使用上下文内模块得到的单词的表示作为输入，获得所有其他方面的表示

情感分类器

将上下文内模块得到的当前aspect的表示和上下文间模块得到的其他方面的表示相加，经过全连接作最终的预测，损失使用交叉熵：

$$\mathbf{o_t} = \mathbf{v}_t^{as} + \mathbf{v}_t^{aa}$$

$$\mathbf{p}(\mathbf{t}) = softmax(\mathbf{W_p o_t} + \mathbf{b_p})$$

对于单aspect的句子，其他方面的表示设置为零向量

实验&结论

实验设置

数据集：Lap14、Rest14、MAMS、Twitter

成分树解析器：SuPar中的CRF constituency parser

依存树解析器：SuPar中的deep Biaffine Parser

ADAM学习率2e-5

L2参数1e-5

实验结果

BiSyn-GAT：没有对多方面关系建模

BiSyn-GAT+：完整模型，对多方面关系建模

Category	Model	Restaurant		Laptop		MAMS		Twitter	
		Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)
w/o Syn.	BERT-SPC	84.46	76.98	78.99	75.03	82.82	81.90	73.55	72.14
	AEN-BERT	83.12	73.76	79.93	76.31	-	-	74.71	73.13
w/ Syn.	R-GAT	86.60	81.35	78.21	74.07	-	-	76.15	74.88
	RGAT+	86.68	80.92	80.94	78.20	84.52	83.74	76.28	75.25
	DGEDT	86.30	80.00	79.80	75.60	-	-	77.90	75.40
	DualGCN	87.13	81.16	81.80	78.10	-	-	77.40	76.02
	SDGCN	83.57	76.47	81.35	78.34	-	-	-	-
	InterGCN	87.12	81.02	82.87	79.32	-	-	-	-
Ours	BiSyn-GAT	87.49	81.63	82.44	79.15	84.90	84.43	**77.99**	**76.80**
	BiSyn-GAT+	**87.94**	**82.43**	**82.91**	**79.38**	**85.85**	**85.49**		

结论

1. SOTA，在Rest和MAMS中，F1提升1.27和1.75（Ours vs baselines）
2. 有语法信息的模型性能更好（w/ Syn vs w/o Syn）
3. 比只用依存信息的模型更有优势（Con. vs Dep.）
4. 关系建模能够提高性能（BiSyn-GAT+ vs BiSyn-GAT）

消融实验

w/o AA：去掉关系建模

w/ AA：保留关系建模

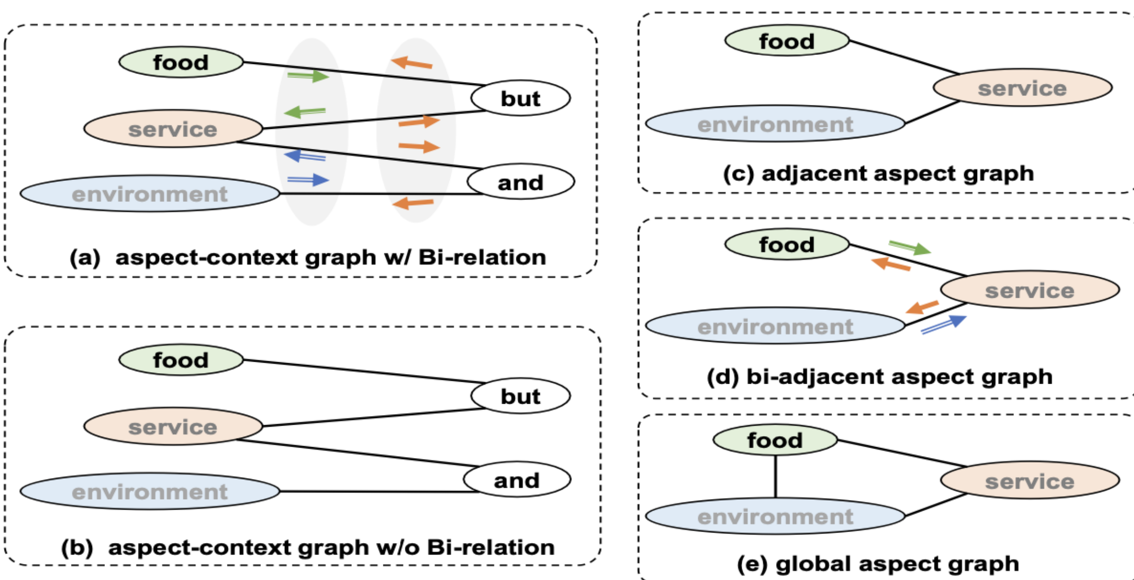
×, +, ⊕：三种融合依存信息的策略

Category	Ablation	Dataset							
		Restaurant		Laptop		MAMS		Twitter	
		Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)
w/o AA	w/o syn. & dep.(BERT+)	84.99	78.51	79.11	75.76	82.71	82.22	75.48	74.54
	w/o con.	86.42	80.10	80.22	76.42	83.38	82.90	76.51	75.29
	w/o dep.	86.60	81.51	81.80	78.48	84.58	84.09	76.81	75.86
	con.×dep.	86.86	80.82	80.85	77.27	84.21	83.76	76.51	75.37
	con.+dep.	86.86	81.59	82.12	78.93	84.73	84.14	77.40	76.39
	con.⊕dep. (BiSyn-GAT)	87.49	81.63	82.44	79.15	84.90	84.43	**77.99**	**76.80**
w/ AA	con.+dep.	87.76	82.18	82.75	79.16	85.48	85.05	-	-
	con.⊕dep. (BiSyn-GAT+)	**87.94**	**82.43**	**82.91**	**79.38**	**85.85**	**85.49**	-	-

结论

1. 语法信息、成分信息都有助于ABSA
2. 一些依存关系会带来噪声（w/o con. vs con.×dep.）
    a. 前者使用所有的依存关系，后者删除非同一成分的依存关系
3. 成分信息融合依存信息效果最好，但con.×dep.效果较差的原因可能是因为删除的边太多导致邻接矩阵稀疏
4. 关系建模也是有帮助的，特别是句子中包含多个方面的数据集

方面-上下文图



左边是作者提出的方面-上下文图，节点是方面以及连接两个方面的上下文

右边是受到其他工作启发提出的方面图，节点是方面

Model		Dataset			
		Restaurant		MAMS	
		Acc.(%)	F1.(%)	Acc.(%)	F1.(%)
BiSyn-GAT		87.49	81.63	84.90	84.43
aspect-context graph	w/ Bi-relation	**87.94**	**82.43**	**85.85**	**85.49**
	w/o Bi-relation	87.85	82.27	85.10	84.69
aspect graph	adjacent	87.49	81.69	85.10	84.61
	Bi-adjacent	87.40	81.53	85.18	84.74
	global	87.49	81.70	85.32	84.88

结论

1. 对于方面-上下文图，有向连接和分割方面的上下文都有助于建模方面间的关系
2. 对于方面图，有向连接并没有取得最好结果，说明上下文的重要性
3. 邻接和全连接的结果差别不大，可能是原因是大部分句子是2个方面的句子

解析器

Model	Parser	Restaurant		MAMS	
		Acc.(%)	F1.(%)	Acc.(%)	F1.(%)
Base		84.99	78.51	82.71	82.22
w/o dep.	Stanford Parser	86.51	81.34	84.51	84.06
	SuPar	86.60	81.51	84.58	84.09
BiSyn-GAT	Stanford Parser	86.66	81.56	84.88	84.31
	SuPar	87.49	81.63	84.90	84.43
BiSyn-GAT+	Stanford Parser	87.84	82.39	85.78	85.40
	SuPar	87.94	82.43	85.85	85.49

案例分析

Sentences	Aspects	BiSyn-GAT	BiSyn-GAT+
it doesn't look like much on the **outside**$_{\mathbf{neg}}$, but the minute you walk inside, it's a whole other **atmosphere**$_{\mathbf{pos}}$.	outside	neu ✗	neg ✓
	atmosphere	pos ✓	pos ✓
while the **service**$_{\mathbf{neg}}$ and **setting**$_{\mathbf{neg}}$ were average , the **food**$_{\mathbf{pos}}$ was excellent.	service	neg ✓	neg ✓
	setting	neu ✗	neg ✓
	food	pos ✓	pos ✓
food was average, the **appetizers**$_{\mathbf{pos}}$ were better than the **main courses**$_{\mathbf{neu}}$.	appetizers	pos ✓	pos ✓
	main courses	pos ✗	neu ✓
i have no complaints about the **wait**$_{\mathbf{pos}}$ or the **service**$_{\mathbf{pos}}$ but the **pizza**$_{\mathbf{neg}}$ was bit at all something to write home about.	wait	neu ✗	pos ✓
	service	neg ✗	pos ✓
	pizza	neg ✓	neg ✓

好在哪里？

1. 提出了成分树的使用方法，一是将成分树的节点分层作为GAT的邻接矩阵，二是分割两个方面获取上下文表示方面间的关系
2. 实验丰富，在多个数据集结果显著。包括Baselines对比、消融实验、不同关系建模方式的对比、解析器的对比
3. 附录补充各种信息，数据集中对于成分树的深度的统计信息、多方面的分布情况、基于规则的上下文的缺陷

个人思考

1. 对成分树每一层用GAT，不同句子的层数不同，而论文设置为3层GAT，多于3的成分树只用3层，少于3的成分树也是3层GAT。相当于在自底向上第3层，作者认为成分树已经很好地将方面及其上下文划分出来。但是这个3是通过统计数据集的信息才确定出来的（见附录），如果能找出每个句子最好的划分情况，是不是更好地学习intra-aspect的信息？
2. 多方面关系建模是基于规则+成分树的方法做的，主要思路就是找两个方面的最小父节点，然后将两个方面之间的兄弟姐妹节点视为连接两个方面的上下文，即这个上下文表示了两个方面的关系。作者在附录提到了基于规则的不足：只能在两个方面之间找。可以尝试把关系直接建模成边，比如并列（表示情感相同）、转折（表示情感相反），然后预测边的类型？

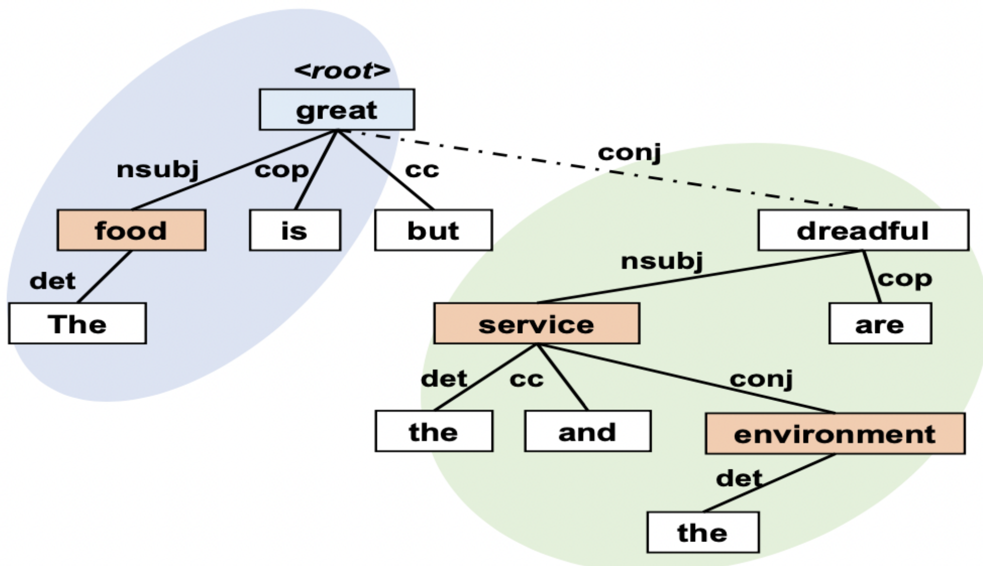
总结

1个思路

使用成分树对齐方面和相关上下文

2个图表

1. 方面-上下文对齐



2. 根据模型特征细分

Category	Model	Dataset: Restaurant Acc.(%)	Restaurant F1.(%)	Laptop Acc.(%)	Laptop F1.(%)	MAMS Acc.(%)	MAMS F1.(%)	Twitter Acc.(%)	Twitter F1.(%)
w/o Syn.	BERT-SPC	84.46	76.98	78.99	75.03	82.82	81.90	73.55	72.14
	AEN-BERT	83.12	73.76	79.93	76.31	-	-	74.71	73.13
w/ Syn.	R-GAT	86.60	81.35	78.21	74.07	-	-	76.15	74.88
	RGAT+	86.68	80.92	80.94	78.20	84.52	83.74	76.28	75.25
	DGEDT	86.30	80.00	79.80	75.60	-	-	<u>77.90</u>	75.40
	DualGCN	87.13	81.16	81.80	78.10	-	-	77.40	<u>76.02</u>
	SDGCN	83.57	76.47	81.35	78.34	-	-	-	-
	InterGCN	87.12	81.02	<u>82.87</u>	<u>79.32</u>	-	-	-	-
Ours	BiSyn-GAT	<u>87.49</u>	<u>81.63</u>	82.44	79.15	<u>84.90</u>	<u>84.43</u>	**77.99**	**76.80**
	BiSyn-GAT+	**87.94**	**82.43**	**82.91**	**79.38**	**85.85**	**85.49**		

5个句式

1. However, the assumption might not be valid as exemplified in Figure 1 (a),

2. We notice that the influence range of one aspect should be continuous, and the mutual influence of aspects at- tenuates with distance.
3. ***BiSyn-GAT+*** shows consistent improvement compared to ***BiSyn-GAT***, which means modeling aspect-aspect relations can improve performance, especially when more multi-aspect sentences are available, *e.g.,* 0.8 and 1.06 F1 improvements on Restaurant and MAMS.
4. Moreover, it shows superiority in the alignments between aspects and corresponding words indicative of sentiment.
5. It also verifies that words within the same phrases of Con.Tree are essential for aligning aspects and corresponding opinions.

10个单词

alleviate	减轻，缓和
valid	有效的
mitigate	缓解
incapable	无法
discourage	不鼓励，阻碍
continuous	连续的
attenuate	衰减
superiority	优势
inferior	不及
verifie	验证