Relational Graph Attention Network for Aspect-based Sentiment Analysis

Kai Wang[1], Weizhou Shen[1], Yunyi Yang[1], Xiaojun Quan[1]*, Rui Wang[2]

[1]School of Data and Computer Science, Sun Yat-sen University, China
[2]Alibaba Group, China
{wangk73, shenwzh3, yangyy37}@mail2.sysu.edu.cn
quanxj3@mail.sysu.edu.cn, masi.wr@alibaba-inc.com

Abstract

Aspect-based sentiment analysis aims to determine the sentiment polarity towards a specific aspect in online reviews. Most recent efforts adopt attention-based neural network models to implicitly connect aspects with opinion words. However, due to the complexity of language and the existence of multiple aspects in a single sentence, these models often confuse the connections. In this paper, we address this problem by means of effective encoding of syntax information. Firstly, we define a unified aspect-oriented dependency tree structure rooted at a target aspect by reshaping and pruning an ordinary dependency parse tree. Then, we propose a relational graph attention network (R-GAT) to encode the new tree structure for sentiment prediction. Extensive experiments are conducted on the SemEval 2014 and Twitter datasets, and the experimental results confirm that the connections between aspects and opinion words can be better established with our approach, and the performance of the graph attention network (GAT) is significantly improved as a consequence.

1 Introduction

Aspect-based sentiment analysis (ABSA) aims at fine-grained sentiment analysis of online affective texts such as product reviews. Specifically, its objective is to determine the sentiment polarities towards one or more aspects appearing in a single sentence. An example of this task is, given a review `great food but the service was dreadful`, to determine the polarities towards the aspects `food` and `service`. Since the two aspects express quite opposite sentiments, just assigning a sentence-level sentiment polarity is inappropriate. In this regard, ABSA can provide better insights into user reviews compared with sentence-level sentiment analysis.

Intuitively, connecting aspects with their respective opinion words lies at the heart of this task. Most recent efforts (Wang et al., 2016b; Li et al., 2017; Ma et al., 2017; Fan et al., 2018) resort to assorted attention mechanisms to achieve this goal and have reported appealing results. However, due to the complexity of language morphology and syntax, these mechanisms fail occasionally. We illustrate this problem with a real review `So delicious was the noodles but terrible vegetables`, in which the opinion word `terrible` is closer to the aspect `noodles` than `delicious`, and there could be `terrible noodles` appearing in some other reviews which makes these two words closely associated. Therefore, the attention mechanisms could attend to `terrible` with a high weight when evaluating the aspect `noodles`.

Some other efforts explicitly leverage the syntactic structure of a sentence to establish the connections. Among them, early attempts rely on handcrafted syntactic rules (Qiu et al., 2011; Liu et al., 2013), though they are subject to the quantity and quality of the rules. Dependency-based parse trees are then used to provide more comprehensive syntactic information. For this purpose, a whole dependency tree can be encoded from leaves to root by a recursive neural network (RNN) (Lakkaraju et al., 2014; Dong et al., 2014; Nguyen and Shirai, 2015; Wang et al., 2016a), or the internal node distance can be computed and used for attention weight decay (He et al., 2018a). Recently, graph neural networks (GNNs) are explored to learn representations from the dependency trees (Zhang et al., 2019; Sun et al., 2019b; Huang and Carley, 2019). The shortcomings of these approaches should not be overlooked. First, the dependency relations, which may indicate the connections between aspects and opinion words, are ignored. Second, empirically, only a small part of the parse tree is related to this

---

*Corresponding author.

task and it is unnecessary to encode the whole tree (Zhang et al., 2018; He et al., 2018b). Finally, the encoding process is tree-dependent, making the batch operation inconvenient during optimization.

In this paper, we re-examine the syntax information and claim that revealing task-related syntactic structures is the key to address the above issues. We propose a novel aspect-oriented dependency tree structure constructed in three steps. Firstly, we obtain the dependency tree of a sentence using an ordinary parser. Secondly, we reshape the dependency tree to root it at a target aspect in question. Lastly, pruning of the tree is performed to retain only edges with direct dependency relations with the aspect. Such a unified tree structure not only enables us to focus on the connections between aspects and potential opinion words but also facilitates both batch and parallel operations. Then we propose a relational graph attention network (R-GAT) model to encode the new dependency trees. R-GAT generalizes graph attention network (GAT) to encode graphs with labeled edges. Extensive evaluations are conducted on the SemEval 2014 and Twitter datasets, and experimental results show that R-GAT significantly improves the performance of GAT. It also achieves superior performance to the baseline methods.

The contributions of this work include:

- We propose an aspect-oriented tree structure by reshaping and pruning ordinary dependency trees to focus on the target aspects.

- We propose a new GAT model to encode the dependency relations and to establish the connections between aspects and opinion words.

- The source code of this work is released for future research.[1]

2 Related Work

Most recent research work on aspect-based sentiment analysis (ABSA) utilizes attention-based neural models to examine words surrounding a target aspect. They can be considered an implicit approach to exploiting sentence structure, since opinion words usually appear not far from aspects. Such approaches have led to promising progress. Among them, Wang et al. (2016b) proposed to use an attention-based LSTM to identify important sentiment information relating to a target aspect.

Chen et al. (2017) introduced a multi-layer attention mechanism to capture long-distance opinion words for aspects. For a similar purpose, Tang et al. (2016) employed Memory Network with multi-hop attention and external memory. Fan et al. (2018) proposed a multi-grained attention network with both fine-grained and coarse-grained attentions. The pre-trained language model BERT (Devlin et al., 2018) has made successes in many classification tasks including ABSA. For example, Xu et al. (2019) used an additional corpus to post-train BERT and proved its effectiveness in both aspect extraction and ABSA. Sun et al. (2019a) converted ABSA to a sentence-pair classification task by constructing auxiliary sentences.

Some other efforts try to directly include the syntactic information in ABSA. Since aspects are generally assumed to lie at the heart of this task, establishing the syntactic connections between each target aspect and the other words are crucial. Qiu et al. (2011) manually defined some syntactic rules to identify the relations between aspects and potential opinion words. Liu et al. (2013) obtained partial alignment links with these syntactic rules and proposed a partially supervised word alignment model to extract opinion targets. Afterward, neural network models were explored for this task. Lakkaraju et al. (2014) used a recursive neural network (RNN) to hierarchically encode word representations and to jointly extract aspects and sentiments. In another work, Wang et al. (2016a) combined the recursive neural network with conditional random fields (CRF). Moreover, Dong et al. (2014) proposed an adaptive recursive neural network (AdaRNN) to adaptively propagate the sentiments of words to the target aspect via semantic composition over a dependency tree. Nguyen et al. (2015) further combined the dependency and constituent trees of a sentence with a phrase recursive neural network (PhraseRNN). In a simpler approach, He et al. (2018a) used the relative distance in a dependency tree for attention weight decay. They also showed that selectively focusing on a small subset of context words can lead to satisfactory results.

Recently, graph neural networks combined with dependency trees have shown appealing effectiveness in ABSA. Zhang et al. (2019) and Sun et al. (2019b) proposed to use graph convolutional networks (GCN) to learn node representations from a dependency tree and used them together with

---

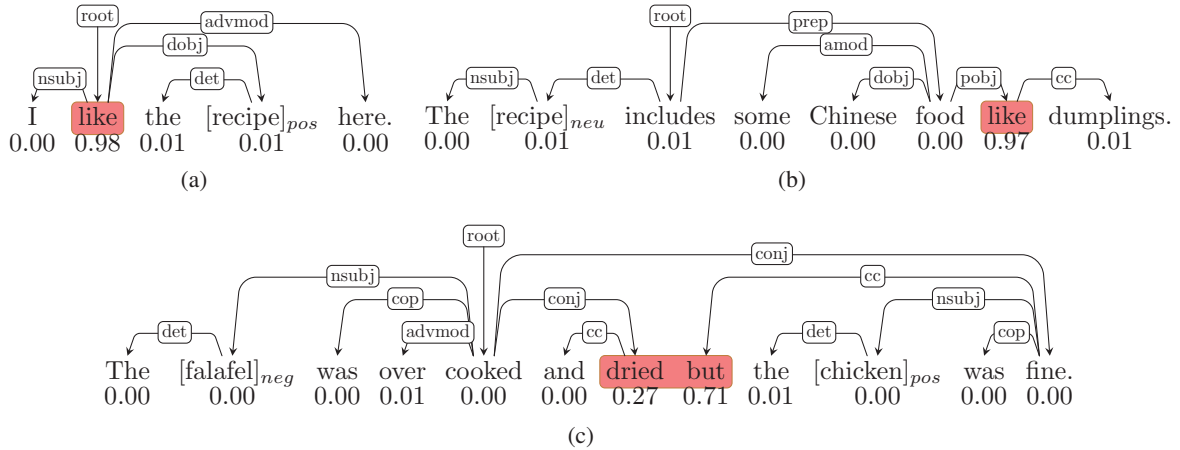[1]https://github.com/shenwzh3/RGAT-ABSA

Figure 1: Three examples from restaurant reviews to illustrate the relationships among aspect, attention, and syntax in ABSA. Labeled edges indicate dependency relations, and scores under each word represent attention weights assigned by the attention-equipped LSTM. Words with high attention weights are highlighted in red boxes, and words in brackets are the target aspects followed by their sentiment labels.

other features for sentiment classification. For a similar purpose, Huang and Carley (2019) used graph attention networks (GAT) to explicitly establish the dependency relationships between words. However, these approaches generally ignore the dependency relations which might identify the connections between aspects and opinion words.

3 Aspect-Oriented Dependency Tree

In this section, we elaborate on the details of constructing an aspect-oriented dependency tree.

3.1 Aspect, Attention and Syntax

The syntactic structure of a sentence can be uncovered by dependency parsing, a task to generate a dependency tree to represent the grammatical structure. The relationships between words can be denoted with directed edges and labels. We use three examples to illustrate the relationships among aspect, attention and syntax in ABSA, as shown in Figure 1. In the first example, the word `like` is used as a verb and it expresses a positive sentiment towards the aspect `recipe`, which is successfully attended by the attention-based LSTM model. However, when it is used as a preposition in the second example, the model still attends to it with a high weight, resulting in a wrong prediction. The third example shows a case where there are two aspects in a single sentence with different sentiment polarities. For the aspect `chicken`, the LSTM model mistakenly assigns high attention weights to the words `but` and `dried`, which leads

to another prediction mistake. These examples demonstrate the limitations of the attention-based model in this task. Such mistakes are likely to be avoided by introducing explicit syntactic relations between aspects and other words. For example, it might be different if the model noticed the direct dependency relationship between `chicken` and `fine` in the third example, rather than with `but`.

3.2 Aspect-Oriented Dependency Tree

The above analysis suggests that dependency relations with direct connections to an aspect may assist a model to focus more on related opinion words, and therefore should be more important than other relations. Also, as shown in Figure 1, a dependency tree contains abundant grammar information, and is usually not rooted at a target aspect. Nevertheless, the focus of ABSA is a target aspect rather than the root of the tree. Motivated by the above observations, we propose a novel aspect-oriented dependency tree structure by reshaping an original dependency tree to root it at a target aspect, followed by pruning of the tree so as to discard unnecessary relations.

Algorithm 1 describes the above process. For an input sentence, we first apply a dependency parser to obtain its dependency tree, where r_{ij} is the dependency relation from node i to j. Then, we build an aspect-oriented dependency tree in three steps. Firstly, we place the target aspect at the root, where multiple-word aspects are treated as entities. Secondly, we set the nodes with direct connections to the aspect as the children, for which the original
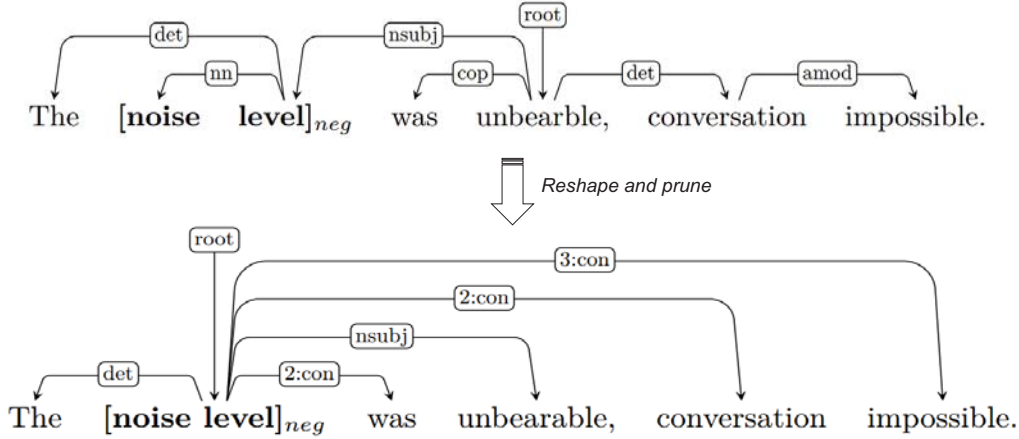
3231

Figure 2: Construction of an aspect-oriented dependency tree (bottom) from an ordinary dependency tree (top).

Algorithm 1 Aspect-Oriented Dependency Tree

Input: aspect $a = \{w_i^a, w_{i+1}^a, ...w_k^a\}$, sentence $s = \{w_1^s, w_2^s, ...w_n^s\}$, dependency tree T, and dependency relations r.

Output: aspect-oriented dependency tree $\hat{T}$.

1: Construct the root R for $\hat{T}$;
2: **for** i to k **do**
3:     **for** $j = 1$ to n **do**
4:         **if** $w_j^s \xrightarrow{r_{ji}} w_i^a$ **then**
5:             $w_j^s \xrightarrow{r_{ji}} R$
6:         **else if** $w_j^s \xleftarrow{r_{ij}} w_i^a$ **then**
7:             $w_j^s \xleftarrow{r_{ij}} R$
8:         **else**
9:             $n = distance(i, j)$
10:            $w_j^s \xrightarrow{n:con} R$
11:         **end if**
12:     **end for**
13: **end for**
14: **return** $\hat{T}$

dependency relations are retained. Thirdly, other dependency relations are discarded, and instead, we put a virtual relation n:con (n connected) from the aspect to each corresponding node, where n represents the distance between two nodes.[2] If the sentence contains more than one aspect, we construct a unique tree for each aspect. Figure 2 shows an aspect-oriented dependency tree constructed from the ordinary dependency tree. There are at least two advantages with such an aspect-oriented structure. First, each aspect has its own dependency tree and can be less influenced by unrelated nodes and relations. Second, if an aspect contains more than

one word, the dependency relations will be aggregated at the aspect, unlike in (Zhang et al., 2019; Sun et al., 2019b) which require extra pooling or attention operations.

The idea described above is partially inspired by previous findings (He et al., 2018a; Zhang et al., 2018; He et al., 2018b) that it could be sufficient to focus on a small subset of context words syntactically close to the target aspect. Our approach provides a direct way to model the context information. Such a unified tree structure not only enables our model to focus on the connections between aspects and opinion words but also facilitates both batch and parallel operations during training. The motivation we put a new relation n:con is that existing parsers may not always parse sentences correctly and may miss important connections to the target aspect. In this situation, the relation n:con enables the new tree to be more robust. We evaluate this new relation in the experiment and the results confirm this assumption.

4 Relational Graph Attention Network

To encode the new dependency trees for sentiment analysis, we propose a relational graph attention network (R-GAT) by extending the graph attention network (GAT) (Veličković et al., 2017) to encode graphs with labeled edges.

4.1 Graph Attention Network

Dependency tree can be represented by a graph $\mathcal{G}$ with n nodes, where each represents a word in the sentence. The edges of $\mathcal{G}$ denote the dependency between words. The neighborhood nodes of node i can be represented by $\mathcal{N}_i$. GAT iteratively updates each node representation (e.g., word embeddings)

---

[2]We set $n = \infty$ if the distance is longer than 4.

by aggregating neighborhood node representations using multi-head attention:

$$h_{att_i}^{l+1} = \|_{k=1}^{K} \sum_{j \in \mathcal{N}_i} \alpha_{ij}^{lk} W_k^l h_j^l \qquad (1)$$

$$\alpha_{ij}^{lk} = attention(i, j) \qquad (2)$$

where $h_{att_i}^{l+1}$ is the attention head of node i at layer $l+1$, $\|_{k=1}^{K} x_i$ denotes the concatenation of vectors from x_1 to x_k, α_{ij}^{lk} is a normalized attention coefficient computed by the k-th attention at layer l, W_k^l is an input transformation matrix. In this paper, we adopt dot-product attention for $attention(i, j)$.[3]

4.2 Relational Graph Attention Network

GAT aggregates the representations of neighborhood nodes along the dependency paths. However, this process fails to take dependency relations into consideration, which may lose some important dependency information. Intuitively, neighborhood nodes with different dependency relations should have different influences. We propose to extend the original GAT with additional relational heads. We use these relational heads as relation-wise gates to control information flow from neighborhood nodes. The overall architecture of this approach is shown in Figure 3. Specifically, we first map the dependency relations into vector representations, and then compute a relational head as:

$$h_{rel_i}^{l+1} = \|_{m=1}^{M} \sum_{j \in \mathcal{N}_i} \beta_{ij}^{lm} W_m^l h_j^l \qquad (3)$$

$$g_{ij}^{lm} = \sigma(relu(r_{ij} W_{m1} + b_{m1}) W_{m2} + b_{m2}) \qquad (4)$$

$$\beta_{ij}^{lm} = \frac{exp(g_{ij}^{lm})}{\sum_{j=1}^{\mathcal{N}_i} exp(g_{ij}^{lm})} \qquad (5)$$

where r_{ij} represents the relation embedding between nodes i and j. R-GAT contains K attentional heads and M relational heads. The final representation of each node is computed by:

$$x_i^{l+1} = h_{att_i}^{l+1} \| h_{rel_i}^{l+1} \qquad (6)$$

$$h_i^{l+1} = relu(W_{l+1} x_i^{l+1} + b_{l+1}) \qquad (7)$$

---

[3]Dot product has fewer parameters but similar performance with feedforward neural network used in (Veličković et al., 2017).
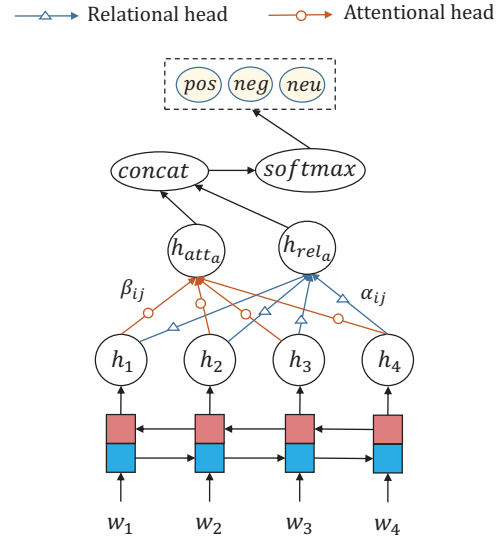

Figure 3: Structure of the proposed relational graph attention network (R-GAT), which includes two genres of multi-head attention mechanism, i.e., attentional head and relational head.

4.3 Model Training

We use BiLSTM to encode the word embeddings of tree nodes, and obtain its output hidden state h_i for the initial representation h_i^0 of leaf node i. Then, another BiLSTM is applied to encode the aspect words, and its average hidden state is used as the initial representation h_a^0 of this root. After applying R-GAT on an aspect-oriented tree, its root representation h_a^l is passed through a fully connected softmax layer and mapped to probabilities over the different sentiment polarities.

$$p(a) = softmax(W_p h_a^l + b_p) \qquad (8)$$

Finally, the standard cross-entropy loss is used as our objective function:

$$L(\theta) = - \sum_{(S,A) \in \mathcal{D}} \sum_{a \in A} \log p(a) \qquad (9)$$

where $\mathcal{D}$ contains all the sentence-aspects pairs, A represents the aspects appearing in sentence S, and θ contains all the trainable parameters.

5 Experiments

In this section, we first introduce the datasets used for evaluation and the baseline methods employed for comparison. Then, we report the experimental results conducted from different perspectives. Finally, error analysis and discussion are conducted with a few representative examples.

3233

Dataset	Positive		Neutral		Negative	
	Train	Test	Train	Test	Train	Test
Laptop	994	341	870	128	464	169
Restaurant	2164	728	807	196	637	196
Twitter	1561	173	3127	346	1560	173

Table 1: Statistics of the three datasets.

5.1 Datasets

Three public sentiment analysis datasets are used in our experiments, two of them are the Laptop and Restaurant review datasets from the SemEval 2014 Task (Maria Pontiki and Manandhar, 2014),[4] and the third is the Twitter dataset used by (Dong et al., 2014). Statistics of the three datasets can be found in Table 1.

5.1.1 Implementation Details

The Biaffine Parser (Dozat and Manning, 2016) is used for dependency parsing. The dimension of the dependency relation embeddings is set to 300. For R-GAT, we use the 300-dimensional word embeddings of GLoVe (Pennington et al., 2014). For R-GAT+BERT, we use the last hidden states of the pre-trained BERT for word representations and fine-tune them on our task. The PyTorch implementation of BERT [5] is used in the experiments. R-GAT is shown to prefer a high dropout rate in between [0.6, 0.8]. As for R-GAT+BERT, it works better with a low dropout rate of around 0.2. Our model is trained using the Adam optimizer (Kingma and Ba, 2014) with the default configuration.

5.2 Baseline Methods

A few mainstream models for aspect-based sentiment analysis are used for comparison, including:

- **Syntax-aware models:** LSTM+SynATT (He et al., 2018a), AdaRNN (Dong et al., 2014), PhraseRNN (Nguyen and Shirai, 2015), AS-GCN (Zhang et al., 2019), CDT (Sun et al., 2019b), GAT (Veličković et al., 2017) and TD-GAT (Huang and Carley, 2019).

- **Attention-based models:** ATAE-LSTM (Wang et al., 2016b) , IAN (Ma et al., 2017), RAM (Chen et al., 2017), MGAN (Fan et al., 2018), attention-equipped LSTM, and fine-tuned BERT (Devlin et al., 2018).

- **Other recent methods:** GCAE (Xue and Li, 2018), JCI (Wang et al., 2018) and TNET (Li

---

[4] http://alt.qcri.org/semeval2014/task4/.
[5] https://github.com/huggingface/transformers

---

et al., 2018).

- **Our methods:** R-GAT is our relational graph attention network. R-GAT+BERT is our R-GAT with the BiLSTM replaced by BERT, and the attentional heads of R-GAT will also be replaced by that of BERT.

5.3 Results and Analysis

5.3.1 Overall Performance

The overall performance of all the models are shown in Table 2, from which several observations can be noted. First, the R-GAT model outperforms most of the baseline models. Second, the performance of GAT can be significantly improved when incorporated with relational heads in our aspect-oriented dependency tree structure. It also outperforms the baseline models of ASGCN, and CDT, which also involve syntactic information in different ways. This proves that our R-GAT is better at encoding the syntactic information. Third, the basic BERT can already outperform all the existing ABSA models by significant margins, demonstrating the power of this large pre-trained model in this task. Nevertheless, after incorporating our R-GAT (R-GAT+BERT), this strong model sees further improvement and has achieved a new state of the art. These results have demonstrated the effectiveness of our R-GAT in capturing important syntactic structures for sentiment analysis.

5.3.2 Effect of Multiple Aspects

The appearance of multiple aspects in one single sentence is very typical for ABSA. To study the influence of multiple aspects, we pick out the reviews with more than one aspect in a sentence. Each aspect is represented with its averaged (GloVe) word embeddings, and the distance between any two aspects of a sentence is calculated using the Euclidean distance. If there are more than two aspects, the nearest Euclidean distance is used for each aspect. Then, we select three models (GAT, R-GAT, R-GAT+BERT) for sentiment prediction, and plot the aspect accuracy by different distance ranges in Figure 4. We can observe that the aspects with nearer distances tend to lead to lower accuracy scores, indicating that the aspects with high semantic similarity in a sentence may confuse the models. However, with our R-GAT, both GAT and BERT can be improved across different ranges, showing that our method can alleviate this problem to a certain extent.

Category	Method	Restaurant		Laptop		Twitter	
		Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
Syn.	LSTM+SynATT	80.45	71.26	72.57	69.13	-	-
	AdaRNN	-	-	-	-	66.30	65.90
	PhraseRNN	66.20	59.32	-	-	-	-
	ASGCN	80.77	72.02	75.55	71.05	72.15	70.40
	CDT	82.30	74.02	77.19	72.99	74.66	73.66
	GAT	78.21	67.17	73.04	68.11	71.67	70.13
	TD-GAT	80.35	76.13	74.13	72.01	72.68	71.15
Att.	ATAE-LSTM	77.20	-	68.70	-	-	-
	IAN	78.60	-	72.10	-	-	-
	RAM	80.23	70.80	74.49	71.35	69.36	67.30
	MGAN	81.25	71.94	75.39	72.47	72.54	70.81
	LSTM	79.10	69.00	71.22	65.75	69.51	67.98
	BERT	85.62	78.28	77.58	72.38	75.28	74.11
Others	GCAE	77.28	-	69.14	-	-	-
	JCI	-	68.84	-	67.23	-	-
	TNET	80.69	71.27	76.54	71.75	74.90	73.60
Ours	R-GAT	83.30	76.08	77.42	73.76	75.57	73.82
Ours	R-GAT+BERT	**86.60**	**81.35**	**78.21**	**74.07**	**76.15**	**74.88**

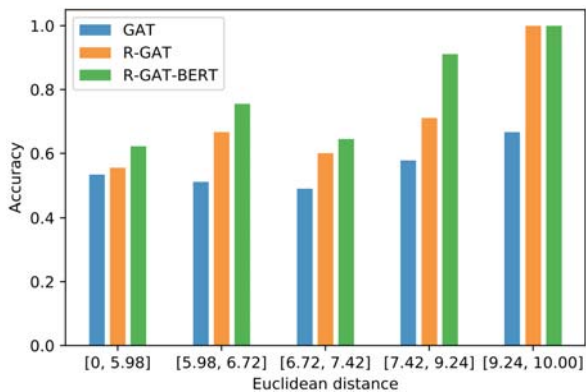Table 2: Overall performance of different methods on the three datasets.

Figure 4: Results of multiple aspects analysis, which shows that the aspects with nearer distances tend to lead to lower accuracy scores.

5.3.3 Effect of Different Parsers

Dependency parsing plays a critical role in our method. To evaluate the impact of different parsers, we conduct a study based on the R-GAT model using two well-known dependency parsers: Stanford Parser (Chen and Manning, 2014) and Biaffine Parser (Dozat and Manning, 2016).[6] Table 3 shows the performance of the two parsers in UAS and LAS metrics, followed by their performance for aspect-based sentiment analysis. From the table,

---

Parser	Performance		Dataset		
	UAS	LAS	Restaurant	Laptop	Twitter
Stanford	94.10	91.49	0.8133	0.7539	0.7283
Biaffine	**95.74**	**94.08**	**0.8330**	**0.7742**	**0.7557**

Table 3: Results of R-GAT based on two different parsers, where UAS and LAS are metrics to evaluate the parsers and higher scores mean better performance.

Tree	Method	Restaurant	Laptop	Twitter
Ordinary	GAT	78.21	73.04	71.67
	R-GAT	79.91	72.72	71.76
Reshaped	GAT	78.57	72.10	71.82
	R-GAT	83.30	77.42	75.57
	R-GAT$_{-n:con}$	81.16	73.66	70.95

Table 4: Results of ablation study, where "Ordinary" means using ordinary dependency trees, "Reshaped" denotes using the aspect-oriented trees, and "*-n:con" denote the aspect-oriented tree without using n:con.

we can find that the better Biaffine parser results in higher sentiment classification accuracies. Moreover, it further implies that while existing parsers can capture most of the syntactic structures correctly, our method has the potential to be further improved with the advances of parsing techniques.

5.3.4 Ablation Study

We further conduct an ablation study to evaluate the influence of the aspect-oriented dependency tree

Category	(%)	Example
Neutral	46	No green beans, no egg, no anchovy dressing, no [nicoise olives]$_{neu}$, no red onion.
Comprehension	32	It took about 2 1/2 hours to be served our 2 [courses]$_{neg}$.
Advice	6	Try the [rose roll]$_{pos}$ (not on menu).
Double negation	6	But [dinner]$_{pos}$ here is never disappointing, even if the prices are a bit over the top.
Neutral	50	Entrees include classics like lasagna, [fettuccine alfredo]$_{neu}$ and chicken parmigiana.
Comprehension	31	We requested they re-slice the [sushi]$_{pos}$, and it was returned to us in small cheese-like cubes.
Advice	5	Gave a [mojito]$_{pos}$ and sit in the back patio.
Double negation	3	And these are not small, wimpy fast food type [burgers]$_{pos}$ - these are real, full sized patties

Table 5: Results of error analysis of R-GAT and R-GAT+BERT on 100 misclassified examples from the Restaurant dataset. The reasons are classified into four categories, for which a sample is given. The upper table corresponds to the results of R-GAT and the lower one corresponds to R-GAT+BERT.

structure and the relational heads. We present the results on ordinary dependency trees for comparison. From table 4, we can observe that R-GAT is improved by using the new tree structure on all three datasets, while GAT is only improved on the Restaurant and Twitter datasets. Furthermore, after removing the virtual relation n:con, the performance of R-GAT drops considerably. We manually examined the misclassified samples and found that most of them can be attributed to poor parsing results where aspects and their opinion words are incorrectly connected. This study validates that adding the n:con relation can effectively alleviate the parsing problem and allows our model to be robust. In this paper, the maximal number of n is set to 4 according to empirical tests. Other values of n are also explored but the results are not any better. This may suggest that words with too long dependency distances from the target aspect are unlikely to be useful for this task.

5.3.5 Error Analysis

To analyze the limitations of current ABSA models including ours, we randomly select 100 misclassified examples by two models (R-GAT and R-GAT+BERT) from the Restaurant dataset. After looking into these bad cases, we find the reasons behind can be classified into four categories. As shown in Table 5, the primary reason is due to the misleading neutral reviews, most of which include an opinion modifier (words) towards the target aspect with a direct dependency connection. The second category is due to the difficulty in comprehension, which may demand deep language understanding techniques such as natural language inference. The third category is caused by the advice which only recommend or disrecommend people to try, with no obvious clues in the sentences

indicating the sentiments. The fourth category is caused by double negation expression, which is also difficult for current models. Through the error analysis, we can note that although current models have achieved appealing progress, there are still some complicated sentences beyond their capabilities. There ought to be more advanced natural language processing techniques and learning algorithms developed to further address them.

6 Conclusion

In this paper, we have proposed an effective approach to encoding comprehensive syntax information for aspect-based sentiment analysis. We first defined a novel aspect-oriented dependency tree structure by reshaping and pruning an ordinary dependency parse tree to root it at a target aspect. We then demonstrated how to encode the new dependency trees with our relational graph attention network (R-GAT) for sentiment classification. Experimental results on three public datasets showed that the connections between aspects and opinion words can be better established with R-GAT, and the performance of GAT and BERT are significantly improved as a result. We also conducted an ablation study to validate the role of the new tree structure and the relational heads. Finally, an error analysis was performed on incorrectly-predicted examples, leading to some insights into this task.

Acknowledgments

References

Danqi Chen and Christopher Manning. 2014. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 740–750.

Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. Recurrent attention network on memory for aspect sentiment analysis. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 452–461.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. 2014. Adaptive recursive neural network for target-dependent twitter sentiment classification. In *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers)*, volume 2, pages 49–54.

Timothy Dozat and Christopher D Manning. 2016. Deep biaffine attention for neural dependency parsing. *arXiv preprint arXiv:1611.01734*.

Feifan Fan, Yansong Feng, and Dongyan Zhao. 2018. Multi-grained attention network for aspect-level sentiment classification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3433–3442.

Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson Liu, Matthew Peters, Michael Schmitz, and Luke Zettlemoyer. 2018. Allennlp: A deep semantic natural language processing platform. *arXiv preprint arXiv:1803.07640*.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2018a. Effective attention modeling for aspect-level sentiment classification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1121–1131.

Shexia He, Zuchao Li, Hai Zhao, and Hongxiao Bai. 2018b. Syntax for semantic role labeling, to be, or not to be. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 2061–2071. Association for Computational Linguistics.

Binxuan Huang and Kathleen Carley. 2019. Syntax-aware aspect level sentiment classification with graph attention networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5472–5480, Hong Kong, China. Association for Computational Linguistics.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Himabindu Lakkaraju, Richard Socher, and Chris Manning. 2014. Aspect specific sentiment analysis using hierarchical deep learning. In *NIPS Workshop on deep learning and representation learning*.

Cheng Li, Xiaoxiao Guo, and Qiaozhu Mei. 2017. Deep memory networks for attitude identification. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 671–680. ACM.

Xin Li, Lidong Bing, Wai Lam, and Bei Shi. 2018. Transformation networks for target-oriented sentiment classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 946–956.

Kang Liu, Heng Li Xu, Yang Liu, and Jun Zhao. 2013. Opinion target extraction using partially-supervised word alignment model. In *Twenty-Third International Joint Conference on Artificial Intelligence*.

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60.

John Pavlopoulos Harris Papageorgiou Ion Androutsopoulos Maria Pontiki, Dimitris Galanis and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35.

Thien Hai Nguyen and Kiyoaki Shirai. 2015. Phrasernn: Phrase recursive neural network for aspect-based sentiment analysis. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2509–2514.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1):9–27.

Chi Sun, Luyao Huang, and Xipeng Qiu. 2019a. Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proceedings of the*

2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 380–385.

Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. 2019b. Aspect-level sentiment analysis via convolution over dependency tree. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5683–5692, Hong Kong, China. Association for Computational Linguistics.

Duyu Tang, Bing Qin, and Ting Liu. 2016. Aspect level sentiment classification with deep memory network. *arXiv preprint arXiv:1605.08900*.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Shuai Wang, Sahisnu Mazumder, Bing Liu, Mianwei Zhou, and Yi Chang. 2018. Target-sensitive memory networks for aspect sentiment classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 957–967.

Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016a. Recursive neural conditional random fields for aspect-based sentiment analysis. *arXiv preprint arXiv:1603.06679*.

Yequan Wang, Minlie Huang, Li Zhao, et al. 2016b. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 606–615.

Hu Xu, Bing Liu, Lei Shu, and S Yu Philip. 2019. Bert post-training for review reading comprehension and aspect-based sentiment analysis. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2324–2335.

Wei Xue and Tao Li. 2018. Aspect based sentiment analysis with gated convolutional networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2514–2523.

Chen Zhang, Qiuchi Li, and Dawei Song. 2019. Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4560–4570, Hong Kong, China. Association for Computational Linguistics.

Yuhao Zhang, Peng Qi, and Christopher D. Manning. 2018. Graph convolution over pruned dependency trees improves relation extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 2205–2215. Association for Computational Linguistics.