

# Aspect-Oriented Opinion Alignment Network for Aspect-Based Sentiment Classification

Xueyi Liu<sup>a</sup>, Rui Hou<sup>a</sup>, Yanglei Gan<sup>a</sup>, Da Luo<sup>a</sup>, Changlin Li<sup>a</sup>, Xiaojun Shi<sup>b</sup> and Qiao Liu<sup>a,\*</sup>

<sup>a</sup>University of Electronic Science and Technology of China

<sup>b</sup>China Academy of Electronics and Information Technology

**Abstract.** Aspect-based sentiment classification is a crucial problem in fine-grained sentiment analysis, which aims to predict the sentiment polarity of the given aspect according to its context. Previous works have made remarkable progress in leveraging attention mechanism to extract opinion words for different aspects. However, a persistent challenge is the effective management of semantic mismatches, which stem from attention mechanisms that fall short in adequately aligning opinion words with their corresponding aspect in multi-aspect sentences. To address this issue, we propose a novel Aspect-oriented Opinion Alignment Network (AOAN) to capture the contextual association between opinion words and the corresponding aspect. Specifically, we first introduce a neighboring span enhanced module which highlights various compositions of neighboring words and given aspects. In addition, we design a multi-perspective attention mechanism that align relevant opinion information with respect to the given aspect. Extensive experiments on three benchmark datasets demonstrate that our model achieves state-of-the-art results. The source code is available at <https://github.com/AONE-NLP/ABSA-AOAN>.

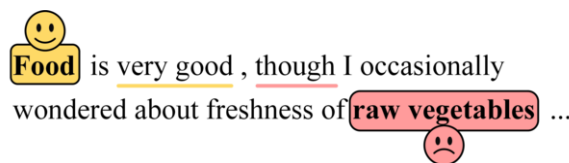
## 1 Introduction

The main purpose of aspect-based sentiment classification (ABSC) is to judge the sentiment polarity (positive, negative, neutral) [1, 2] of aspect words in sentences expressing opinions. ABSC is an entity-level oriented and fine-grained challenge for sentiment analysis. To illustrate, consider the following sample sentence taken from the SemEval 2014 restaurant dataset:

*Food is very good, though I occasionally wondered about freshness of raw vegetables in side orders.*

In this sentence, the aspect are "food" and "raw vegetables", and the expected sentiment polarities of these aspects are intended to be positive and negative. Identifying the sentiment polarity of aspect is crucial for applications [3, 4] such as product review analysis, where understanding customers' opinions on specific aspects of a product can provide valuable insights for businesses.

To solve the aspect-based sentiment classification (ABSC) task, it is crucial to establish the semantic relationship between an aspect and its corresponding opinion words. Various recurrent neural networks (RNNs) [5, 6] have been proposed to learn representations directly from the left or/and right context with regard to the given aspect. However, prior studies have faced difficulties in accurately establishing semantic associations between aspect and long-distance context



**Figure 1.** An example sentence contains two aspects but with opposite sentiment polarities from the restaurant reviews.

on sequential modeling. Therefore, attention mechanism have been widely adopted in ABSC tasks to model the correlations between aspect and context [7, 8, 9]. Unlike RNN-based models, attention mechanisms possess global modeling capability, which allows them to capture long-distance dependencies between aspect and context. However, attention mechanisms may not be effective when dealing with sentences containing multiple aspects. For instance, the given aspect "raw vegetables" may be associated with the opinion words "very good" and "though" simultaneously. In such cases, attention mechanisms may struggle to align opinion words with their corresponding aspect, resulting in semantic mismatch [8].

To address this issue, various works introduce position information (e.g. fixed size window) [10, 11] and proximity strategy (e.g. position weight decay) [12, 13], which have proved that context words in closer proximity are more likely to be the actual opinion words of the aspect. However, these approaches may not encompass or emphasize all relevant opinion words, limiting the model's ability to fully comprehend the contextual meaning. As shown in Figure 1, the aspect "food" has the opinion words "very good" that are relatively close to the aspect and easy to capture its sentiment by the aforementioned methods. However, accurately calculating the sentiment of the aspect "raw vegetables" requires the capture of a more comprehensive set of semantic information, including the opinion word "though", which is much farther away from the aspect. Thus, the way prior works take position information into consideration may fall short in scenarios where the opinion words are distant from the aspect or convey complex semantic information. Therefore, the challenge remains on *how to utilize attention mechanisms to accurately capture and match appropriate opinion words with respect to the given aspect*.

Drawing on the insights from recent studies on the critical role of different semantic compositionality which can improve expressive ability and provide syntactic structure for natural language understanding [14], as well as the demonstrated effectiveness of span-level information in aspect-based sentiment classification (ABSC) [15, 16], we present a novel Aspect-oriented Opinion Alignment

\* Corresponding Author. Email: qiliu@uestc.edu.

Network (AOAN) for aspect-based sentiment classification. Our proposed model addresses the issue of semantic mismatch by introducing a neighboring span enhanced module to highlights a variety of neighboring words compositions with respect to aspect. These compositions are used to emphasize different ranges of aspect neighboring words, providing flexibility to the contextual association between the aspect and neighboring words. To capture more comprehensive relevant opinion words based on different compositions of neighboring words, we then propose a multi-perspective attention module that utilizes abstract understanding representations to model multi-perspective sentiment representations of each aspect. This parallel attention mechanism improves the accuracy and comprehensiveness of capturing the relevant opinion words regarding the given aspect. Finally, the multi-perspective sentiment representations are combined by a global average pooling layer, which aggregates the information from all neighboring spans, providing a comprehensive representation of the overall sentiment expressed by the given aspect. The main contributions can be summarized as follows:

- We propose an Aspect-oriented Opinion Alignment Network (AOAN), a novel framework designed to mitigate the semantic mismatch problem. This method is capable of exploiting contextual association of different neighboring spans and guarantees proper alignment between opinion words and the given aspect.
- Our propose AOAN employs a neighboring span enhanced module that highlights various compositions of neighboring words and given aspects, enabling the capture of more evident information related to a given aspect. Additionally, a multi-perspective attention module is designed to align comprehensive opinion information with the given aspect in a parallel way.
- We conduct extensive experiments on three benchmark datasets to evaluate the effectiveness of our approach. Experimental results demonstrate that our model outperforms the state-of-the-art methods, which confirms the efficacy of our proposed approach.

## 2 Related Work

Aspect-based sentiment classification (ABSC) is a fine-grained sentiment analysis task that focuses on extracting sentiment polarity towards a specific aspect within a given context. Early ABSC methods [17, 18] relied on handcrafted features and failed to capture the intrinsic semantic associations between the given aspect and context.

Recently, various neural network-based approaches, such as Convolutional Neural Networks (CNNs) [13, 19], Recurrent Neural Networks (RNNs) [6, 20], and Memory Networks [7], have been proposed to model the semantic relation between the aspect and context in an implicit way. For instance, Tang et al. [6] introduced two LSTM-based models, namely TD-LSTM and TC-LSTM, which segmented the sentence into three parts: the preceding context words, the aspect, and the following context words. However, RNN-based and its variants methods face challenges in capturing long-distance contextual sentiment features when the aspect is far away from the opinion words, due to the limitation of sequential modeling.

With this in mind, researchers have deployed attention mechanisms for the aspect-based sentiment classification (ABSC) task to capture long-distance semantic features through global modeling. Tang et al. [7] proposed a Deep Memory Network (MemNet) that utilizes an attention mechanism to explicitly capture the relevance of each contextual word with respect to the aspect and infer the sentiment polarity. However, the inherent defects of attention mechanisms cannot differentiate the correlations of contextual words with respect to the given aspect, leading to a semantic mismatch problem.

To tackle the aforementioned issue, some works improved attention mechanisms by modeling a global perspective of sentence-level information [8] or interaction between aspect and context [21, 22]. Liu et al. [8] proposed a content attention-based aspect-based sentiment classification model (Cabasc) that captures crucial information about given aspects from a global perspective. Ma et al. [21] introduced an interactive attention network (IAN) that generates the representations of target and context interactively. Huang et al. [22] proposed an attention over attention networks (AOA) that learns attentions from both aspect-to-text and text-to-aspect, suggesting that opinion words are highly correlated with the aspect. However, they neglected the fact that the position information is also crucial for identifying the sentiment of the aspect.

Taking this into consideration, some researchers have introduced various position information and proximity strategies to improve the effectiveness of aspect-based sentiment classification (ABSC) models. Gu et al. [10] proposed a position-aware bidirectional attention network (PBAN) that gives more attention to neighboring words of the aspect than words with long distances, while Zhou et al. [11] proposed a position-aware hierarchical transfer (PAHT) model that utilizes position information from multiple levels. Chen et al. [12] adopted a proximity strategy that assumes a closer opinion word is more likely to be the actual modifier of the target and designed a recurrent attention network (RAM) to counter irrelevant information using weight decay mechanisms. However, these approaches may not encompass or emphasize all relevant opinion words, limiting the model's ability to fully comprehend the contextual meaning.

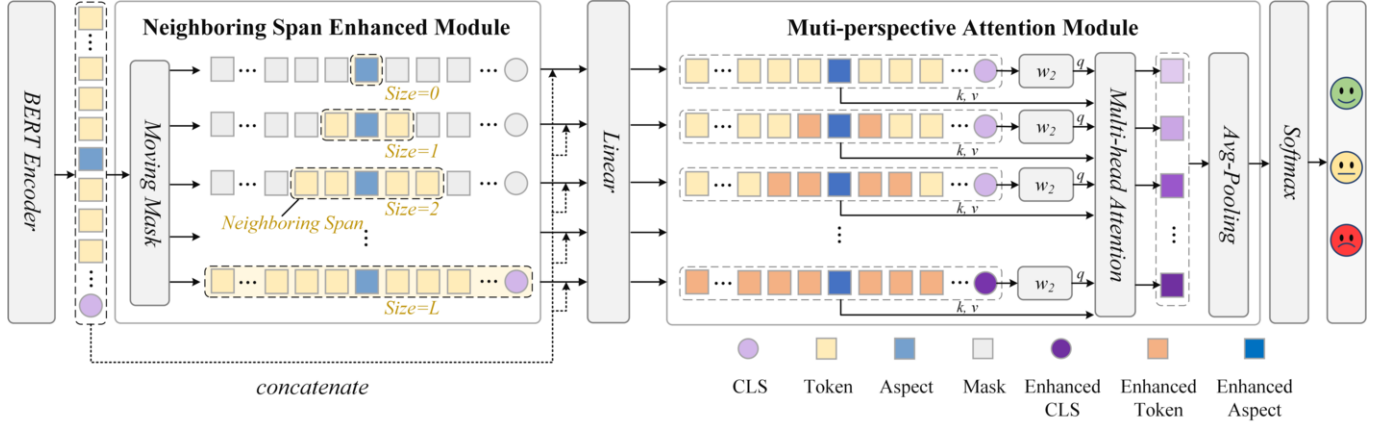
Another trend of research has explored the use of graph neural networks (GNNs) for modeling syntactic structures of sentence based on dependency trees. For instance, Zhang et al. [23] introduced aspect-specific graph convolutional networks (ASGCN) to handle aspect-level sentiment classification tasks. Tian et al. [24] proposed a type-aware graph convolutional network (T-GCN) that utilizes an attentive layer ensemble to learn contextual information from different GCN layers. Li et al. [25] proposed a dual graph convolutional network (DualGCN) that simultaneously took syntax structures and semantic correlations into consideration. Although syntactic-based methods have achieved promising results, the imperfect parsing performance and randomness of input sentences inevitably introduce noise through the dependency tree.

Prior works [10, 12, 13] which take position information into consideration may fall short in scenarios where the opinion words are distant from the aspect or convey complex semantic information. In this paper, we do not rely on syntactic information and focus on the different compositions of aspect neighboring words, which provide comprehensive insights into the sentiment expressed.

## 3 Methodology

### 3.1 Overview

We describe our model AOAN in this section, which has two main modules shown in Figure 2: a neighboring span enhanced module and a multi-perspective attention module. The neighboring span enhanced module highlights different compositions of neighboring words through multiple neighboring spans. The multi-perspective attention module captures relevant opinion words regarding the given aspect via multi-perspective sentiment representations. We will discuss each component in detail in the following sub-sections.



**Figure 2.** The overall architecture of AOAN, which is composed primarily of a neighboring span enhanced module highlights neighboring word spans of varying ranges, while a multi-perspective attention module captures relevant opinion words with a comprehensive view.

### 3.2 Task Definition

The aim of our model is to predict the sentiment polarity of a given sentence towards a given aspect, based on the contextual information in the sentence. Specifically, let  $S = \{w_1, \dots, w_n\}$  represent a sentence comprising  $n$  words, and let  $A = \{w_{a+1}, \dots, w_{a+m}\}$  denote the aspect mentioned in the sentence, consisting of  $m$  words. Our goal is to accurately predict the sentiment polarity of the sentence  $S$  towards the aspect  $A$  from the set  $\{positive, neutral, negative\}$ .

### 3.3 BERT Encoder

To obtain contextual representations of sentences, we employ the pre-trained BERT language model [26], which maps sentences to input vector  $H \in \mathbb{R}^{(n+1) \times d}$  including a class token and  $n$  word tokens, where  $d$  is the embedding dimension of the input vectors and  $n$  is the length of input sentence. To capture aspect-aware context representations [25, 27] of the sentence, we construct a sentence-aspect pair as "[CLS] Sentence [SEP] Aspect [SEP]" to serve as the input  $S$  of BERT encoder. This process yields aspect-aware context representations  $H = \{h_{CLS}, h_1, \dots, h_n\}$  of the sentence, which is the input of the subsequent modules and enable our model to effectively extract and comprehend the semantic meaning of the input.

$$H = BERT(S) \quad (1)$$

### 3.4 Neighboring Span Enhanced Module

Taking into account the observation that words surrounding a given aspect tend to carry favourable information for sentiment polarity prediction [10, 11, 13], we construct multiple cohesive units, referred to as *neighboring spans*, to model the tokens that precede and succeed the aspect, including the aspect itself. These neighboring spans encompass various ranges of tokens surrounding the aspect, each within a specific range of positions. Specifically, we employ a moving mask mechanism that first identifies the position of each token relative to the aspect. We then calculate the relative distance  $d_i$  of each token with regard to the aspect, denoted as:

$$d_i = \begin{cases} P_{a+1} - P_i & i < a + 1 \\ 0 & a + 1 \leq i \leq a + m \\ P_i - (P_{a+1} + m) & i > a + m \end{cases} \quad (2)$$

Here,  $P_i$  denotes the position of  $i$ -th token,  $i$  ranges from 0 to  $n$ , where  $n$  is the length of the input sentence.  $P_{a+1}$  denotes the start position of the aspect, and  $m$  represents the length of the aspect.

To model various composition of contextual neighboring words, we construct multiple neighboring spans based on the span size threshold  $L$ , which can range from 0 to  $L$ . This approach allows us to provide  $L + 1$  neighboring spans. The mask vector  $V_i^l$  for each context word in a span of size  $l \in [0, L]$  is defined as:

$$V_i^l = \begin{cases} E & d_i \leq l \\ O & d_i > l \end{cases} \quad (3)$$

where  $E \in \mathbb{R}^d$  is the ones vector and  $O \in \mathbb{R}^d$  is the zeros vector.  $V_i^l$  denotes the mask vector of the  $i$ -th token for the span size  $l$ . We then obtain the  $l$ -th neighboring span representation  $H_{span}^l \in \mathbb{R}^{(n+1) \times d}$ , where the span size is  $l$  as follows:

$$H_{span}^l = M^l \cdot H \quad (4)$$

where the mask matrix  $M^l = \{V_0^l, V_1^l, \dots, V_n^l\} \in \mathbb{R}^{(n+1) \times d}$  is composed of  $n + 1$  mask vectors  $V_i^l$  and represents a span of size  $l$  for the input sentence.

The neighboring span enhanced representation  $H_{enhanced}^l \in \mathbb{R}^{(n+1) \times d}$  is then obtained by the concatenation of neighboring span representation and the original context representations  $H$ , and then mapped back to the unified  $d$ -dimensional semantic space via a shared linear transformation, denoted as:

$$H_{enhance}^l = W_1(H_{span}^l \odot H) + b_1 \quad (5)$$

where  $W_1 \in \mathbb{R}^{d \times 2d}$  and  $b_1 \in \mathbb{R}^d$  are the weight and bias matrix in the linear transformation layer, and  $\odot$  denotes concatenate operation.

### 3.5 Multi-perspective Attention Module

In order to precisely capture opinion words and mitigate the problem of semantic mismatch, a multi-perspective attention module is utilized in this study. This module allows for parallel capturing sentiment information in neighboring span enhanced representations  $H_{enhance}^l$  from various perspectives. Unlike other state-of-the-art approaches [7, 21, 28], we utilize the abstract understanding representations  $C^l \in \mathbb{R}^d$  transformed from  $h_{CLS}^l$ , as the query vector

instead of the given aspect as the class token learns an abstract understanding of all the tokens in a sentence, which gives a comprehensive view. This ensures a more concentrated attention distribution in neighboring span, as it is believed that neighboring words may contain more valuable information for sentiment classification:

$$C^l = f(W_2 \cdot h_{CLS}^l + b_2) \quad (6)$$

$$Q^l = W_Q \cdot C^l + b_Q \quad (7)$$

$$K^l = W_K \cdot H_{enhance}^l + b_K \quad (8)$$

$$V^l = W_V \cdot H_{enhance}^l + b_V \quad (9)$$

where  $W_2 \in \mathbb{R}^{d \times d}$  and  $b_2 \in \mathbb{R}^d$  are the weight and bias matrix in a linear transformation layer, respectively. The activation function  $Tanh$  is denoted as  $f(\cdot)$ .  $W_Q \in \mathbb{R}^{d \times d}$ ,  $W_K \in \mathbb{R}^{d \times d}$  and  $W_V \in \mathbb{R}^{d \times d}$  are the weight matrices,  $b_Q \in \mathbb{R}^d$ ,  $b_K \in \mathbb{R}^d$  and  $b_V \in \mathbb{R}^d$  are the bias matrices in three linear transformation layers.

The multi-head scaled dot-product attention is calculated as:

$$y_a^l = W_h(head_1^l \odot head_2^l \odot \dots \odot head_h^l) \quad (10)$$

$$head_i^l = Attention(Q_i^l, K_i^l, V_i^l) \quad (11)$$

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (12)$$

where  $\sqrt{d_k}$  is a scaling factor to avoid large values in the dot product between the query and key vectors, which can lead to instability in the softmax function.  $Q_i^l$ ,  $K_i^l$  and  $V_i^l$  denotes the query, key and value of the  $i$ -th head, respectively. The dimensions of  $Q_i^l$ ,  $K_i^l$  and  $V_i^l$  are  $\frac{d}{h}$  and  $h$  refers to the total number of heads. Furthermore,  $W_h \in \mathbb{R}^{d \times d}$  represents a weight matrix to enable linear transformation.

Each attention output vector  $y_a^l \in \mathbb{R}^d$  in  $[y_a^0, \dots, y_a^L]$  are subjected to a linear transformation with shared parameters to produce three-dimensional sentiment representations  $y_s^l \in \mathbb{R}^3$ , defined as:

$$y_s^l = W_3 \cdot y_a^l + b_3 \quad (13)$$

where  $W_3 \in \mathbb{R}^{3 \times d}$  and  $b_3 \in \mathbb{R}^3$  are the weight and bias matrix in the linear transformation layer, respectively.

To integrate the sentiment representations obtained from the diverse compositions of neighboring spans, we utilize a global average pooling layer, which effectively combines the multi-perspective sentiment representations  $Y_s = \{y_s^0, \dots, y_s^L\} \in \mathbb{R}^{(L+1) \times 3}$ , denoted as:

$$y_c = AvgPool1d(Y_s) \quad (14)$$

where  $y_c \in \mathbb{R}^3$  is the integrated representation and the  $AvgPool1d(\cdot)$  is a 1-d average pooling layer with kernel size  $L+1$ .

### 3.6 Sentiment Classifier

The integrated representation  $y_c$  can be utilized with a softmax function to predict sentiment polarity distribution:

$$\hat{y} = Softmax(W_o y_c + b_o) \quad (15)$$

**Table 1.** Statistics of the three experimental datasets.

Datasets	Positive		Negative		Neural	
	Train	Test	Train	Test	Train	Test
Laptop	944	341	870	128	464	169
Restaurant	2164	728	807	196	637	196
Twitter	1561	173	1560	173	3127	346

where  $\hat{y} \in \mathbb{R}^3$  represents the predicted sentiment polarity distribution,  $W_o \in \mathbb{R}^{3 \times 3}$  and  $b_o \in \mathbb{R}^3$  are learned parameters. To train the entire network, a loss function is minimized as follows:

$$L(\hat{y}, y) = - \sum_{i=1}^N \sum_{j=1}^C y_i^j \log(\hat{y}_i^j) + \lambda \left( \sum_{\theta \in \Theta} \theta^2 \right) \quad (16)$$

where  $y_i^j$  is the ground truth sentiment polarity,  $C$  is the number of sentiment polarity categories,  $\hat{y}_i^j$  represents the predicted sentiment probabilities,  $\theta$  represents each parameter to be regularized,  $\Theta$  is a collection of all parameters,  $\lambda$  denotes the weight coefficient for  $L_2$  regularization.

## 4 Experiments

### 4.1 Datasets

We perform experiments on three public standard benchmark datasets. The Restaurant and Laptop datasets from SemEval2014 task 4 [2], consist of reviews on the restaurant and laptop domains, respectively. The Twitter dataset is constructed from Twitter posts [29]. All of these datasets include instances with three sentiment polarities: positive, neutral and negative. Moreover, each sentence in these datasets is annotated with corresponding aspects and their polarities. Statistics for the three datasets are presented in Table 1.

### 4.2 Implementation Details

For our experiments, we utilize 768-dimensional word embeddings initialized with *bert-base-uncased English version*<sup>1</sup> as the input. We employ the Adam optimizer with a learning rate of  $2 \times 10^{-5}$ . To address overfitting, we incorporate  $L_2$  regularization with a regularization weight of  $\lambda_2$  set to  $1 \times 10^{-5}$ . Hyper-parameters  $L$  are swept over  $\{10, 10, 0\}$  on the Laptop, Restaurant and Twitter datasets, respectively. The number of head  $h$  in multi-head attention is set to 12. All the model weights are initialized from a uniform distribution. The model is trained in 20 epochs with a batch size of 16, and the maximum sequence length during training is set to 100. Furthermore, the model is trained on a single NVIDIA A100 Tensor Core GPU and an Intel Xeon Bronze 3104 CPU. We also conduct the T-test on multiple random seeds to ensure the robustness of our results.

### 4.3 Baseline Models

To comprehensively evaluate our model, we compare it with state-of-the-art baselines. The models are briefly described as follows:

- **ATAE-LSTM** [20]: Wang et al. combine the attention mechanism and LSTM to model the connection between aspect and context.
- **MemNet** [7]: Tang et al. repeatedly use the attention mechanism to extract contextual features of aspect. The last attention output is used to predict sentiment polarity.

<sup>1</sup> <https://huggingface.co/bert-base-cased>



**Table 2.** Comparison of experimental results on three public datasets. Results that indicate a statistically significant improvement with a p-value of less than 0.05 under the bootstrap paired t-test are marked with an "\*".

Model	Laptop		Restaurant		Twitter	
	Accuracy(%)	F1(%)	Accuracy(%)	F1(%)	Accuracy(%)	F1(%)
ATAE-LSTM	68.70	-	77.20	-	-	-
MemNet	72.37	-	80.95	-	-	-
RAM	69.59	64.61	78.30	68.46	70.52	67.95
IAN	72.10	-	78.60	-	-	-
Cabasc	72.57	67.06	79.37	69.46	70.95	69.00
PBAN	74.12	-	81.16	-	-	-
BERT-SPC	78.99	75.03	84.46	76.98	73.55	72.14
AEN-BERT	79.93	76.31	83.12	73.76	74.71	73.13
BERT4GCN	77.49	73.01	84.75	77.11	74.73	73.76
T-GCN+BERT	80.88	77.03	86.16	79.95	76.45	75.25
DualGCN+BERT	81.80	78.10	87.13	81.16	77.40	76.02
AOIARN	81.50	78.20	83.40	75.70	75.00	73.80
SSEGCN+BERT	81.01	77.69	<b>87.31</b>	81.09	77.40	76.02
<b>AOAN</b>	<b>82.45*</b>	<b>79.09*</b>	86.43*	<b>81.19*</b>	<b>77.46*</b>	<b>76.54*</b>

- **RAM** [12]: Chen et al. improve MemNet by representing memory with BiLSTM. At the same time, GRU is introduced to process the features of multi-layer attention mechanism.
- **IAN** [21]: Ma et al. generate aspect and context representations through two LSTMs respectively, and learn aspect and context representations interactively.
- **Cabasc** [8]: Liu et al. utilize two types of attention mechanisms to learn information of a given aspect from a global perspective and sequential correlations between the words and the aspect.
- **PBAN** [10]: Gu et al. utilize the relative positions between aspects and sentiment words to attend to informative parts of the input.
- **BERT-SPC** [26]: The input of the pre-trained BERT model is "[CLS] + context + [SEP] + aspect + [SEP]". Finally, the sentiment polarity of aspect is predicted by the classifier "[CLS]".
- **AEN-BERT** [30]: Song et al. design an attention encoding network to extract features related to aspects in context, and improved performance by the pre-training BERT.
- **BERT4GCN** [31]: Zhang et al. integrate the grammatical sequential features from the PLM of BERT and the syntactic knowledge from dependency graphs.
- **T-GCN** [24]: Tian et al. utilize dependency types to distinguish different relations in the graph and uses attentive layer ensemble to learn the contextual information from different GCN layers.
- **DualGCN** [25]: Li et al. design a SynGCN module and a SemGCN module with orthogonal and differential regularizers.
- **AOIARN** [32]: Yang et al. adopts an interactive attention between aspect embeddings learned from GCN and opinion semantic embeddings learned from three bi-LSTMs.
- **SSEGCN** [27]: Zhang et al. model both semantic correlations and syntactic structures of a sentence for ABSC.

#### 4.4 Main Results

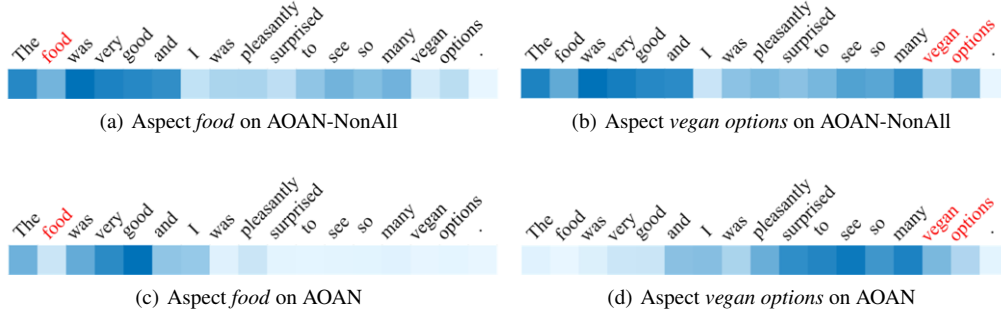
The assessment of Aspect-Based Sentiment Classification (ABSC) models is typically done using accuracy and macro-averaged F1-score. In Table 2, we present a comparison of our model with existing approaches. Our proposed AOAN model achieves state-of-the-art performance on the Laptop, Restaurant and Twitter datasets with F1-scores of 79.09%, 81.19%, and 76.54%, respectively. Although our approach falls slightly short of satisfactory accuracy on

**Table 3.** Ablation Experiment Result on SemEval2014 datasets.

Model	Laptop		Restaurant	
	Accuracy(%)	F1(%)	Accuracy(%)	F1(%)
AOAN-NonAll	79.31	75.94	85.45	78.30
AOAN-NonSpan	79.00	75.58	85.54	78.47
AOAN-Aspect	80.41	75.97	86.16	78.91
AOAN-Maxpool	75.86	71.15	81.25	71.15
AOAN-Single(0)	80.09	77.01	85.09	77.70
(1)	80.25	77.36	85.54	79.14
(2)	79.15	74.92	85.09	77.75
(3)	79.62	76.14	85.35	77.69
(4)	79.94	76.63	85.36	77.24
(5)	79.62	76.30	85.36	78.34
(6)	<b>81.19</b>	<b>78.37</b>	85.09	78.37
(7)	80.41	77.75	85.36	77.48
(8)	80.09	76.49	85.18	77.62
(9)	81.03	78.73	<b>86.34</b>	<b>80.84</b>
(10)	80.25	77.66	84.82	77.31
<b>AOAN</b>	<b>82.45</b>	<b>79.09</b>	<b>86.43</b>	<b>81.19</b>

the Restaurant dataset, it still performs well. These results indicate that our AOAN model can effectively capture accurate and comprehensive multi-perspective sentiment representations based on different neighboring span compositions to solve semantic mismatch.

Compared to semantic-based models, such as RAM, PBAN, AEN-BERT and AOIARN, our AOAN model outperforms them with F1-score improvements of 0.89%, 4.21%, and 2.74% on three datasets, respectively. Compared to RAM and PBAN which based on the position information and proximity strategy, one possible reason why our model performs better is that we use the BERT as the encoder which is effectively to extract contextual semantics. Another possible reason is that our model combines different neighboring word compositions and utilizes abstract understanding representations to capture important information from a global view while ignoring unimportant information makes it effective in aligning opinion words with their corresponding aspect. Furthermore, our model shows improved performance of 0.99%, 0.03%, and 0.52% in F1-score on three datasets when compared to syntactic-based models such as BERT4GCN, T-GCN, DualGCN, and SSEGCN, without the need to incorporate any syntactic knowledge. In summary, the evaluation of our model shows its effectiveness and superiority in ABSC task, out-



**Figure 3.** The attention scores of AOAN-NonAll and AOAN model on Restaurant dataset example.

performing existing approaches without the need for complex syntactic structures.

#### 4.5 Ablation Study

To investigate the effectiveness of the different modules in our AOAN model, we conducted extensive ablation studies on the SemEval2014 datasets, evaluating the impact of each module individually. The results of our experiments are presented in Table 3. We evaluated four different variants of our model: AOAN-NonAll, AOAN-NonSpan, AOAN-Aspect, AOAN-Maxpool and AOAN-Single(L). The AOAN-Single(L) model tests the model's ability to align opinion words with their corresponding aspect. Descriptions of each variant are detailed below:

**AOAN-NonAll** refers to a model that uses the output embeddings  $H$  from the BERT encoder and the average vector of aspect representations  $\{h_{a+1}, \dots, h_{a+m}\}$  as the input to an attention mechanism, which produces the final classification.

**AOAN-NonSpan** removes the neighboring span enhanced module and instead uses the output embeddings  $H$  from the BERT encoder as the input to the multi-perspective attention module.

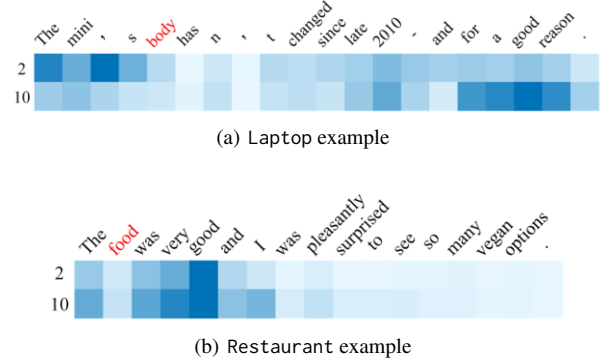
**AOAN-Aspect** replaces the abstract understanding representation  $C^l$  of AOAN with the average vector of aspect representations  $\{h_{a+1}^l, \dots, h_{a+m}^l\}$  in multiple neighboring span enhanced representations as the query for the multi-head attention mechanism.

**AOAN-Maxpool** replaces the global average pooling layer of AOAN with a max pooling layer to synthesize multi-perspective sentiment representations  $Y_s$ .

**AOAN-Single(L)** only uses the result of a single fixed span with size  $L$ , rather than combining the results of  $L + 1$  neighboring spans.

The experimental results demonstrate the contribution of each component to the overall effectiveness of the AOAN model. Specifically, utilizing aspect words to capture opinion words in the sentence (AOAN-NonAll) leads to a significant decrease in performance due to the lack of modules which align the opinion words with the given aspects. On the other hand, the effect of AOAN-NonSpan proves that the neighboring span enhanced module is effective to highlight the importance of neighboring words, resulting in improved performance. The comparison between AOAN-Aspect and AOAN shows that utilizing abstract understanding representations is a crucial clue to capture accurate and comprehensive opinion words.

Furthermore, the significant decrease in F1-score of 7.94% and 10.04% on Laptop and Restaurant datasets for AOAN-Maxpool suggests that each of the multi-perspective representations should be evenly utilized in the final classification. The lower performance of AOAN-Single(0) to AOAN-Single(10) in comparison to AOAN further supports the idea that combining different compositions can pro-



**Figure 4.** The attention scores of AOAN model on Restaurant and Laptop datasets examples.

vide flexibility to the contextual association between the aspect and neighboring words. Overall, the results of the ablation experiments provide evidence for the effectiveness of each component of AOAN in improving the overall performance of the ABSC task.

#### 4.6 Visualization

To provide a clear understanding of the attention mechanism employed in our AOAN model, we present visualizations of the attention weights in two sample sentences in Figures 3 and 4. The attention weights are computed as the average of all heads, and the intensity of the color represents the impact of the word on the sentence. To better illustrate, Figure 3 shows the average of attention scores for each size of neighboring span, while Figure 4 shows attention scores for different size of neighboring span in our AOAN model.

The sentence "The food was very good and I was pleasantly surprised to see so many vegan options." is taken from the restaurant dataset with aspects "food" and "vegan options". Figure 3(a) and 3(b) show the attention scores of the AOAN-NonAll model, which uses aspect to capture opinion words. For the aspect "food", unrelated words like "see" and "many" are given high weights, introducing noise. Similarly, for the "vegan options" aspect, the irrelevant words "was very good" receive the highest weights, leading to semantic mismatch. In contrast, our AOAN model (Figure 3(c) and 3(d)), with neighboring spans and multi-perspective attention, focuses attention around the aspect, aligning opinion words accurately.

In Figure 4, we show attention scores in span size 2 and 10 of the above Restaurant sentence and another example sentence, "The mini's body hasn't changed since late 2010 - and for a good reason." with the aspect "body" taken from the Laptop dataset. In Figure 4(a),

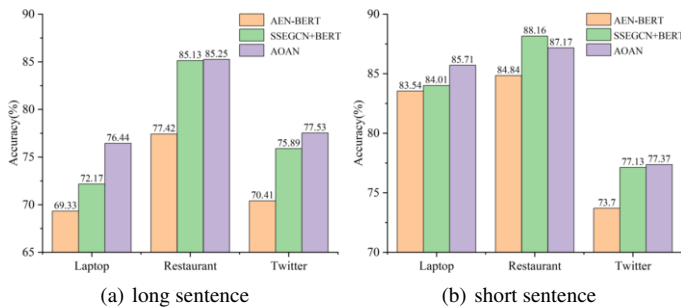
**Table 4.** Case studies of our AOAN model compared with MemNet, AEN-BERT and SSEGCN+BERT.

#	Sentence	MemNet	AEN-BERT	SSEGCN	AOAN
1	Not as fast as I would have expect for an <b>i5</b> .	P <del>X</del>	N <del>✓</del>	N <del>✓</del>	N <del>✓</del>
2	This newer netbook has no <b>hard drive</b> or <b>network lights</b> .	(N <del>X</del> ,N <del>X</del> )	(N <del>X</del> ,N <del>X</del> )	(N <del>X</del> ,N <del>X</del> )	(O <del>✓</del> ,O <del>✓</del> )
3	Yes, they're a bit more expensive then typical, but then again, so is their <b>food</b> .	O <del>X</del>	N <del>X</del>	N <del>X</del>	P <del>✓</del>
4	Most of the <b>sandwiches</b> are made with <b>soy mayonaise</b> which is actually pretty good.	(P <del>✓</del> ,P <del>✓</del> )	(O <del>X</del> ,O <del>X</del> )	(O <del>X</del> ,P <del>✓</del> )	(P <del>✓</del> ,P <del>✓</del> )
5	But <b>dinner</b> here is never disappointing, even if the <b>prices</b> are a bit over the top.	(O <del>X</del> ,N <del>✓</del> )	(P <del>✓</del> ,P <del>X</del> )	(P <del>✓</del> ,N <del>✓</del> )	(P <del>✓</del> ,N <del>✓</del> )

the span size of 2 fails to capture the long-distance context, while the relevant opinion words "for a good reason" are aligned by the span size of 10. However, in Figure 4(b), the span size of 10 introduce more noises from irrelevant words such as "I" and "pleasantly" compared to span size of 2. The attention scores in different span size in our AOAN model reveal that the model can capture opinion words from different perspectives, which provide flexibility to the contextual association between the aspect and neighboring words. Therefore, the AOAN model can capture more comprehensive opinion words based on different compositions of neighboring words.

#### 4.7 Impact of Sentence Length

To investigate the effectiveness of our proposed AOAN model in mitigating the problem of semantic mismatch, we considered the fact that longer sentences tend to contain more irrelevant words, thus exacerbating the issue. To this end, we defined a sentence whose length is greater than or equal to the average length as a long sentence and a sentence whose length is less than the average length as a short sentence. We then divided the test datasets of Laptop, Restaurant, and Twitter into short and long sentences based on their respective average lengths [21, 23, 26]. We evaluated the accuracy of AEN-BERT which outperforms other semantic-based baselines, SSEGCN+BERT which outperforms other syntactic-based baselines, and our AOAN on short and long sentences.

**Figure 5.** Accuracies of our AOAN model on sentence of different length compared with AEN-BERT and SSEGCN+BERT.

As depicted in Figure 5(a) and 5(b), compared to AEN-BERT and SSEGCN-BERT, our model achieves better performance overall. Furthermore, our AOAN model achieves more significant improvement on long sentence compared with short sentence. Experimental results indicate that our proposed AOAN model is effective in mitigating the problem of semantic mismatch in long sentences and across varying sentence lengths, thereby improving the overall performance of sentiment analysis tasks.

#### 4.8 Case Study

Table 4 illustrates several case studies analyzed using different models, denoted by symbols P, N, and O for positive, negative, and neutral sentiment, respectively. Aspect words are highlighted in red for easy identification. In the first sentence, MemNet failed to consider the negative meaning of the word "not" and made an incorrect prediction. In contrast, our AOAN model predicted neutral sentiment for the aspects "hard drive" and "network lights" in the second sentence, demonstrating its effectiveness in capturing overall sentiment. The third sentence, with no explicit sentiment expression, was correctly predicted by our AOAN model.

These examples highlight the importance of learning the integral semantics of a sentence for the ABSC task, and our AOAN successfully utilizes abstract understanding representations to capture comprehensive opinion information. In the last two sentences with different aspects, our AOAN accurately predicted the sentiment, showcasing its ability to align opinion words with their corresponding aspect and address the problem of semantic mismatch.

### 5 Conclusions

In this paper, we propose an AOAN to address the semantic mismatch problem in ABSC task. Specifically, we first design a neighboring span enhanced module, which highlights a variety of neighboring words compositions with respect to aspect for better capturing the accurate relevant opinion words. Then, we design a multi-perspective attention module to utilize abstract understanding to parallel model the multi-perspective sentiment representations that improves the accuracy and comprehensiveness of the capturing the relevant opinion words regarding the given aspect. Experimental results on three datasets achieve the state-of-the-art performance which demonstrate the effectiveness of our approach.

### Acknowledgements

We would like to thank the anonymous reviewers for their helpful discussion and feedback. This work was supported by the National Science Foundation of China (U19B2028, U22B2061) and the National Science and Technology Major Project of the Ministry of Science and Technology of China (2022YFB4300603).

### References

- [1] Bing Liu, 'Sentiment analysis and opinion mining', *Synthesis lectures on human language technologies*, 5(1), 1–167, (2012).
- [2] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammed AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al., 'Semeval-2016 task 5: Aspect based sentiment analysis', in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pp. 19–30. Association for Computational Linguistics, (2016).

- [3] Ashima Yadav and Dinesh Kumar Vishwakarma, 'Sentiment analysis using deep learning architectures: a review', *Artificial Intelligence Review*, **53**(6), 4335–4385, (2020).
- [4] Hai Ha Do, Penatiyana WC Prasad, Angelika Maag, and Abeer Al-sadoon, 'Deep learning for aspect-based sentiment analysis: a comparative review', *Expert systems with applications*, **118**, 272–299, (2019).
- [5] Thien Hai Nguyen and Kiyooki Shirai, 'Phrasernn: Phrase recursive neural network for aspect-based sentiment analysis', in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2509–2514, (2015).
- [6] Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu, 'Effective LSTMs for target-dependent sentiment classification', in *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers*, pp. 3298–3307. The COLING 2016 Organizing Committee, (2016).
- [7] Duyu Tang, Bing Qin, and Ting Liu, 'Aspect level sentiment classification with deep memory network', in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 214–224, (2016).
- [8] Qiao Liu, Haibin Zhang, Yifu Zeng, Ziqi Huang, and Zufeng Wu, 'Content attention model for aspect based sentiment analysis', in *Proceedings of the 2018 World Wide Web conference*, pp. 1023–1032, (2018).
- [9] Feifan Fan, Yansong Feng, and Dongyan Zhao, 'Multi-grained attention network for aspect-level sentiment classification', in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3433–3442, (2018).
- [10] Shuqin Gu, Lipeng Zhang, Yuexian Hou, and Yin Song, 'A position-aware bidirectional attention network for aspect-level sentiment analysis', in *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 774–784, (2018).
- [11] Jie Zhou, Qin Chen, Jimmy Xiangji Huang, Qinmin Vivian Hu, and Liang He, 'Position-aware hierarchical transfer model for aspect-level sentiment classification', *Information Sciences*, **513**, 1–16, (2020).
- [12] Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang, 'Recurrent attention network on memory for aspect sentiment analysis', in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 452–461, (2017).
- [13] Xin Li, Lidong Bing, Wai Lam, and Bei Shi, 'Transformation networks for target-oriented sentiment classification', in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 946–956, (2018).
- [14] Richard Socher, Brody Huval, Christopher D Manning, and Andrew Y Ng, 'Semantic compositionality through recursive matrix-vector spaces', in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 1201–1211, (2012).
- [15] Yan Zhou, Longtao Huang, Tao Guo, Jizhong Han, and Songlin Hu, 'A span-based joint model for opinion target extraction and target sentiment classification', in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pp. 5485–5491, (2019).
- [16] Wei Chen, Peiqin Lin, Wanqi Zhang, Jinglong Du, and Zhongshi He, 'Hierarchical interactive network for joint aspect extraction and sentiment classification', *Knowledge-Based Systems*, **256**, 109825, (2022).
- [17] Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao, 'Target-dependent twitter sentiment classification', in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: human language technologies*, pp. 151–160, (2011).
- [18] Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad, 'Nrc-canada-2014: Detecting aspects and sentiment in customer reviews', in *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, pp. 437–442, (2014).
- [19] Yoon Kim, 'Convolutional neural networks for sentence classification', in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1746–1751, (2014).
- [20] Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao, 'Attention-based lstm for aspect-level sentiment classification', in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 606–615, (2016).
- [21] Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang, 'Interactive attention networks for aspect-level sentiment classification', in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 4068–4074, (2017).
- [22] Binxuan Huang, Yanglan Ou, and Kathleen M Carley, 'Aspect level sentiment classification with attention-over-attention neural networks', in *Proceedings of Social, Cultural, and Behavioral Modeling: 11th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, pp. 197–206. Springer, (2018).
- [23] Chen Zhang, Qiuchi Li, and Dawei Song, '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*, pp. 4568–4578, (2019).
- [24] Yuanhe Tian, Guimin Chen, and Yan Song, 'Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble', in *Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: human language technologies*, pp. 2910–2922, (2021).
- [25] Ruifan Li, Hao Chen, Fangxiang Feng, Zhanyu Ma, Xiaojie Wang, and Eduard Hovy, 'Dual graph convolutional networks for aspect-based sentiment analysis', in *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)*, pp. 6319–6329, (2021).
- [26] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, 'Bert: Pre-training of deep bidirectional transformers for language understanding', 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)*, pp. 4171–4186, (2019).
- [27] Zheng Zhang, Zili Zhou, and Yanna Wang, 'Ssegcn: Syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis', in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4916–4925, (2022).
- [28] Xiaodi Wang, Mingwei Tang, Tian Yang, and Zhen Wang, 'A novel network with multiple attention mechanisms for aspect-level sentiment analysis', *Knowledge-Based Systems*, **227**, 107196, (2021).
- [29] Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu, '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)*, pp. 49–54, (2014).
- [30] Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao, 'Attentional encoder network for targeted sentiment classification', *arXiv preprint arXiv:1902.09314*, (2019).
- [31] Zeguan Xiao, Jiarun Wu, Qingliang Chen, and Congjian Deng, 'Bert4gcn: Using bert intermediate layers to augment gcn for aspect-based sentiment classification', in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 9193–9200, (2021).
- [32] Baiyu Yang, Donghong Han, Rui Zhou, Di Gao, and Gang Wu, 'Aspect opinion routing network with interactive attention for aspect-based sentiment classification', *Information Sciences*, **616**, 52–65, (2022).