*Article*

# Integration of Multi-Branch GCNs Enhancing Aspect Sentiment Triplet Extraction

Xuefeng Shi [1,2,3], Min Hu [1,2,3,*], Jiawen Deng [4], Fuji Ren [5,*], Piao Shi [1,2,3] and Jiaoyun Yang [1,2,4]

1   School of Computer and Information, Hefei University of Technology, Hefei 230601, China
2   Key Laboratory of Knowledge Engineering with Big Data, Hefei University of Technology, Hefei 230601, China
3   Key Laboratory of Affective Computing and Advanced Intelligent Machine, Hefei University of Technology, Hefei 230601, China
4   Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
5   School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
*   Correspondence: jsjxhumin@hfut.edu.cn (M.H.); renfuji@uestc.edu.cn (F.R.)

**Abstract:** Aspect Sentiment Triplet Extraction (ASTE) is a complex and challenging task in Natural Language Processing (NLP). It aims to extract the triplet of aspect term, opinion term, and their associated sentiment polarity, which is a more fine-grained study in Aspect Based Sentiment Analysis. Furthermore, there have been a large number of approaches being proposed to handle this relevant task. However, existing methods for ASTE suffer from powerless interactions between different sources of textual features, and they usually exert an equal impact on each type of feature, which is quite unreasonable while building contextual representation. Therefore, in this paper, we propose a novel Multi-Branch GCN (MBGCN)-based ASTE model to solve this problem. Specifically, our model first generates the enhanced semantic features via the structure-biased BERT, which takes the position of tokens into the transformation of self-attention. Then, a biaffine attention module is utilized to further obtain the specific semantic feature maps. In addition, to enhance the dependency among words in the sentence, four types of linguistic relations are defined, namely part-of-speech combination, syntactic dependency type, tree-based distance, and relative position distance of each word pair, which are further embedded as adjacent matrices. Then, the widely used Graph Convolutional Network (GCN) module is utilized to complete the work of integrating the semantic feature and linguistic feature, which is operated on four types of dependency relations repeatedly. Additionally, an effective refining strategy is employed to detect whether word pairs match or not, which is conducted after the operation of each branch GCN. At last, a shallow interaction layer is designed to achieve the final textual representation by fusing the four branch features with different weights. To validate the effectiveness of MBGCNs, extensive experiments have been conducted on four public and available datasets. Furthermore, the results demonstrate the effectiveness and robustness of MBGCNs, which obviously outperform state-of-the-art approaches.

**Keywords:** ASTE; biaffine attention; structure-biased BERT; GCN; linguistic feature
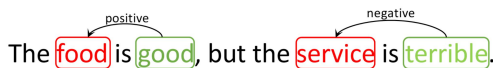
## 1. Introduction

Recently, a tremendous advance has been achieved in the development of social media platforms, which largely encourage people to express their emotional states online [1,2]. Furthermore, it has become popular to publish users' comments or opinions about services and products on specific electronic platforms in a timely manner. These perspectives expressed directly by consumers are extremely important for merchants to improve their service while in a dealing. Thus, how to extract the exact aspect terms, opinion terms, and their corresponding sentiment from a specific sentence is a significant Natural Language Processing (NLP) subtask [3–5]. The recent developing task, Aspect-Based Sentiment

Analysis (ABSA), aims to mine the explicit or implicit sentiment information about the opinion terms with regard to the specific aspect terms, which implements sentiment analysis about consumers' reviews effectively. Generally, the ABSA task contains seven fundamental subtasks (Figure 1), which are Aspect Term Extraction (ATE) [6], Aspect Term Extraction and Sentiment Classification (AESC) [7], Opinion Term Extraction (OTE) [6], Aspect-Based Sentiment Classification (ABSC) [8], Aspect-Oriented Opinion Term Extraction (AOE) [9], Pair Extraction (PE) [10], and Aspect Sentiment Triplets Extraction (ASTE) [11]. In particular, as the fine-grained subtask in ABSA, the ASTE task takes aspect terms, opinion terms, and sentiment polarities into consideration simultaneously, which is challenging but significant. For example, as shown in Figure 1, the review "The food is good, but the service is terrible". contains two triplets, (food, good, positive) and (service, terrible, negative). Unlike the other subtasks, such triplets extracted by the ASTE task can better reflect multiple emotional factors (aspect, opinion, sentiment) from the user reviews and are more proper for practical application scenarios.

The food is good, but the service is terrible.
positive — negative

| Tasks | Definitions | Targets |
|---|---|---|
| ATE | aims to extract aspect terms from the text. | food, service |
| AESC | aims to extract aspect terms from the text, and classifies its sentiment polarity. | (food, positive), (service, negative) |
| OTE | aims to extract opinion terms from the text. | good, terrible |
| ABSC | aims to predict the sentiment polarity with respect to the specific aspect terms. | food => positive, service => negative |
| AOE | aims to extract opinion terms according to aspect terms. | food => good, service => terrible |
| PE | aims to extract aspect and opinion terms simultaneously. | (food, good), (service, terrible) |
| ASTE | aims to complete the extraction of aspect and opinion terms and the prediction of sentiment polarity simultaneously. | (food, good, positive), (service, terrible, negative) |

**Figure 1.** Example of the task categories in (ABSA) Aspect-Based Sentiment Analysis.

In previous studies, the pipeline manner is widely applied in the approaches to ASTE. Peng et al. [12] first introduced the ASTE task and extracted the triplet {aspect, opinion, sentiment} via utilizing a pipeline method, which contains a two-stage framework. The first stage provided predictions about aspect, opinion, and sentiment, respectively. Furthermore, the second stage was designed to pair up the predictions achieved from the first stage and output triplets. However, the interactions among them were totally ignored, and the potential error was propagated between these two stages [13,14]. To take the dependencies among the multiple subtasks into consideration, the multi-turn machine reading comprehension (MRC) manner [15,16] was utilized to jointly train multiple subtasks together, and it has achieved significant results. In addition, the fashion of end-to-end [17,18] also attracts many researchers' attentions, which is constructed based on the new tagging scheme.

Although the paradigm of the framework is important to enhance the performance of the ASTE task, the effective utilization of various linguistic relations between words is also decisive to the task's success [19]. Specifically, the syntactic dependency tree is widely used to present the structure of a sentence, which tends to depict the syntactic relations among words. Zhao et al. [20] adopted the dependency tree as the support to capture relations between aspect and opinion terms. Furthermore, the work [21] directly employed an interactive attention mechanism to integrate syntactic and semantic relations between words. In addition, the contribution of part-of-speech categories to ASTE is also noticed, which straightly impacts the semantic representation of sentences. Except for the

dependency tree, relative position also largely influences the expression of the sentence. Xu et al. [22] applied a position-aware tagging scheme to mark the relative position between words in a sentence. Furthermore, the semantic features in this work are represented by Long Short-term Memory (LSTM) with the pre-trained Glove, which cannot handle contextual ambiguity comprehensively. Moreover, the tree-based distance and relative position distance of each word pair in the sentence also contribute a lot to the improvement in the ASTE task [23], and the utilization of Bidirectional Encoder Representation from Transformers (BERT) can largely enhance the feature representation from the semantic perspective. However, although significant progress has been achieved by previous studies, there are still remaining limitations: the effective optimization of semantic features is not enough, and the powerful utilization of multi-type textual features is unsolved yet.

To address these two problems, motivated by the impressive performance achieved by BERT, we propose a novel BERT- and Graph Convolutional Network-based (GCN-based) model Multi-branch Graph Convolutional Network (MBGCN) for the ASTE task. In detail, in our model, to evacuate the potential capability of BERT and obtain a more exquisite contextual representation, a structure-biased BERT [24] is firstly utilized as the semantic feature encoder. Subsequently, depending on the generated representations, aspect-oriented and opinion-oriented feature maps are extracted by two multi-layer perceptions (MLP). Then, before incorporating other relations of words, a biaffine attention module is applied to unify the aspect-oriented and opinion-oriented semantic features effectively. Unlike fusing textual features via a single GCN, an MBGCN employs four branch GCNs to integrate semantic representation with syntactic dependency type, part-of-speech combination, tree-based distance, and relative position distance among each word pair, respectively. Through the complementary of these four branches, a more precise textual representation is achieved. Finally, a shallow interaction strategy is designed to complete the work of information fusion before the triplet decoding layer. To validate the effectiveness of the MBGCN, a series of experiments are conducted on four widely used and available datasets. The experimental results prove that MBGCNs can efficiently deal with the complex relations among sentences and outperform the state-of-the-art (SOTA) ASTE approaches.

The main contributions of this work can be summarized as follows:

- We propose a framework MBGCN to extract the aspect, opinion, and sentiment triplet from review sentences in an end-to-end fashion, which can avoid error propagation among different subtasks;
- We utilize a structure-biased BERT to improve the ability to extract abundant contextual information, which provides rich textual features for subsequent task-oriented operations;
- Our proposed MBGCN adopts four branch GCNs to integrate the semantic feature with four types of linguistic relations, including syntactic dependency type, part-of-speech combination, tree-based distance, and relative position distance of each word pair. Furthermore, a shallow interaction layer is introduced to output the final textual representation;
- The extensive experiments conducted on multiple ASTE datasets prove that the proposed MBGCN outperforms the mentioned SOTA baselines.

The remainder of this article is organized as follows. In Section 2, we present a brief overview of the development of ABSA, previous research about ASTE, and the application of GCNs. The proposed framework MBGCN is introduced in detail in Section 3. In Section 4, we provide detailed experimental studies and performance analyses. Finally, Section 5 provides a conclusion of this study and an outlook for future work.

## 2. Related Works

In the past decade, fine-grained sentiment analysis and opinion extraction have been attractive research in the NLP community, and have firmly attracted many researchers' attentions. In this section, we will first briefly review the development of the ABSA task.

Secondly, a succinct summary of existing approaches for ASTE will be introduced. Lastly, the application of GCNs in ASTE will be shortly summarized.

### 2.1. Aspect-Based Sentiment Analysis

ABSA is a fine-grained task that aims to recognize the explicit or implicit sentiment information in a given sentence [25–27]. Normally, a sentence usually includes several aspect terms and opinion terms simultaneously, which means multiple sentiment expressions are contained in it. Specifically, with the development of e-commerce, this situation usually happens in the reviews of products and services, which are published on online platforms [28,29]. Through mining the opinions from these reviews, the merchants can learn the real and direct requests from consumers about their services. Thus, many efforts have been contributed to this task since it was proposed. Additionally, we can categorize the existing ABSA approaches into three types: the lexicon-based method [30], machine learning method [31], and deep learning method [32]. In traditional methods, the performance of the ABSA task largely depends on feature engineering, such as bag-of-words [33] and part-of-speech [23]. Although impressive performance has been achieved by traditional methods, the cost of handcrafted features is unbearable for human experts. Currently, the rapid development of deep learning promotes the improvement in contextual representation, which also encourages the progress of ABSA tasks straightforwardly [34,35]. In deep learning methods, they usually fine-tune the pre-trained language model (PLM) with the specific training data to generate task-oriented feature maps. As a representative of PLM, BERT makes a remarkable impression on vast NLP researchers with its outstanding ability to model contextual information. Thus, it is also utilized as a backbone in our proposed model for the ASTE task.

### 2.2. ASTE Methods

As a subtask of sentiment analysis, ASTE has been studied by many NLP researchers after being proposed [36,37], and aims to extract aspect terms, opinion terms, and the corresponding sentiment polarity in a sentence, simultaneously. From the above investigation, it has been known that the pipeline manner method proposed by [12] had an error propagation problem between different subtasks. However, the methods with an end-to-end manner can avoid this problem with their unique architecture. Chen et al. [11] decomposed the ASTE task into three subtasks: target tagging, opinion tagging, and sentiment tagging. Furthermore, a new target-aware tagging scheme was used to identify the correspondences between opinion targets and the whole sentence. In addition, span-level features also contribute a lot to the ASTE task. Chen et al. [38] proposed a joint training framework to process all potential entities as independent spans, and the related representations of the spans were utilized to classify their corresponding sentiment polarities. Moreover, to reduce the cost of sequence tagging, a tagging-free solution was proposed by Mukherjee et al. [39]. In the method, an encoder–decoder architecture with a pointer network-based decoding framework was introduced, which effectively captured the interactions between the aspects and opinions by considering the whole detected spans in predicting sentiment polarity. To prove the simple span-based method is also effective for ASTE, Xu et al. [40] proposed a three-layers framework, which consisted of a BERT-based encoding layer, a span representation layer, and an aspect–sentiment–opinion prediction layer. This work verified that the performance of the model for ASTE was impacted by explicit local context information largely. Through the above summary, it is obviously learned that the approaches with the end-to-end manner contribute a lot to the ASTE task, and it is essential to pay more attention to the research of effectively utilizing the relations among words in a sentence. Thus, in this work, we propose a novel model to integrate five kinds of words' relations together to enhance the performance of ASTE.

### 2.3. Application of GCN in ASTE

In (ABSA) Aspect-Based Sentiment Analysis tasks, the syntax dependency tree plays an important role in catching the key feature from the review sentence [41–43]. Furthermore, it is well known that the regular method GCN is popular in handling dependency graphs in previous works. Regarding ASTE, GCNs are also used widely to fuse different sources of information. As mentioned above, Shi et al. [21] employed a GCN to enhance the interaction between syntactic and semantic features. To fully exploit the potential information implied in syntactic and semantic features, the work [18] also integrated semantic and syntactic representations through a GCN module, which preserved the sequential information and enhanced the linguistic representation, simultaneously. Moreover, to overcome the problem of many aspect terms to one opinion term or one aspect term to many opinion terms, Li et al. [44] combined a GCN with a base encoder to build the span representations, which included both aspect terms and opinion terms. In [45], a GCN was also employed to model the graph based on the concatenated representations of aspects terms and opinion terms. Thus, it is quite clear that GCNs are extremely important in enhancing the feature representations in the ASTE task. Motivated by their impressive ability, we also process the work of feature fusion under the guidance of the GCN in this paper.

Conclusively, as aforementioned, ASTE is a difficult and challenging subtask in ABSA, which attracts a lot of researchers' attentions. In this paper, inspired by the existing works which apply BERT and GCNs to NLP tasks, we propose a novel model MBGCN to process semantic feature, syntactic dependency type, part-of-speech combination, tree-based distance, and relative position distance, simultaneously.
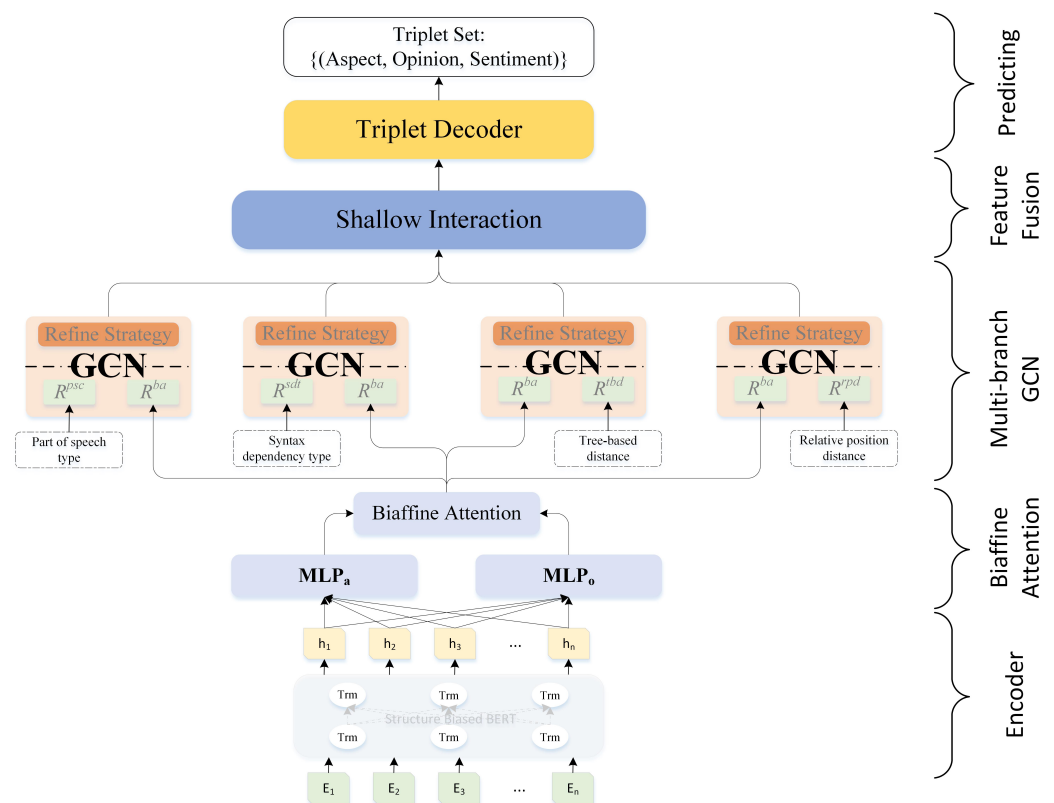
## 3. Framework of MBGCN

In this section, the detailed framework of the MBGCN is described. Firstly, the definition of the ASTE task is introduced briefly. Then, the mechanism of feature generation through the backbone structure-biased BERT is depicted, and this step is utilized to generate semantic features. After that, multi-branch GCNs are employed to integrate semantic features with the other four types of linguistic feature representations. Lastly, the shallow interaction, output layer, and training are introduced shortly. Additionally, the overall architecture of the MBGCN is described in detail in Figure 2.

### 3.1. Task Formulation

Given a sentence with a sequence of words $\mathbf{X} = \{w_1, w_2, \ldots, w_n\}$ as input, where $n$ is the number of words, the goal of the ASTE task is to extract and output a set of triplets $\{(a, o, s)_k\}_{k=1}^{m}$, where $a$, $o$, and $s$ are the aspect term, opinion term, and the corresponding sentiment polarity, respectively, and $m$ is the number of triplets. Concretely, the aspect $a$ can be decomposed into two or more elements, i.e., $(a_b, a_e)$, where $b$ and $e$ mean the start and end positions. The opinion $o$ can be decomposed as $(o_b, o_e)$ similarly. Furthermore, $s$ is selected from the set (position, netural, negative) to represent the sentiment polarity of the corresponding opinion term on the aspect term. For the sentence shown in Figure 1, the triplets are collected as (food, good, positive) and (service, terrible, negative).

Specifically, to make the target of our ASTE task more explicit, ten types of relations between words in a review are defined, which are collected in Table 1. Similarly, the mentioned relations also can be seen as the labels, and these labels are introduced to present the relations in the word pairs, which are also the eventual predictions of our MBGCN.

**Figure 2.** The overall framework of the proposed Multi-branched Graph Convolutional Network (MBGCN). $[E_1, E_2, E_3, \ldots, E_n]$ is the input vector **E** of self-attention (Equations (1) and (2)).

**Table 1.** The definitions of our defined relations.

| Items | Relation | Definition |
|:---:|:---:|:---:|
| 1 | B-A | beginning of aspect term. |
| 2 | I-A | inside of aspect term. |
| 3 | A | aspect term. |
| 4 | B-O | beginning of opinion term. |
| 5 | I-O | inside of opinion term. |
| 6 | O | opinion term. |
| 7 | POS | sentiment polarity is positive. |
| 8 | NEU | sentiment polarity is neutral. |
| 9 | NEG | sentiment polarity is negative. |
| 10 | N | belong to no aforementioned relations. |

### 3.2. Embedding via Structure-Biased BERT

As aforementioned, BERT has an impressive performance in modeling contextual representation in various NLP tasks [46–49]. Therefore, in our proposed Multi-branches Graph Convolutional Netwrok (MBGCN),we also utilize it to generate the semantic features by the version of the `bert-uncased-base`. To be precise, before feeding the review **X** into the MBGCN, the input is always formulated in three formats: segment embedding $X_s$, position embedding $X_p$, and tokens embedding $X_t$. Then, these three aspects of embedding are summarized as the input to the selective feature generator, which is shown in Equation (1),

$$\mathbf{E} = X_s + X_p + X_t, \tag{1}$$

where $\mathbf{E} = [E_1, E_2, E_3, \ldots, E_n]$ is the input of self-attention (Equation (2)). In detail, BERT is a PLM with the structure of a stacked transformer, which has 12 transformer layers in total. Furthermore, in each transformer layer, the feature representations are trans-

formed by multi-head self-attention with a residual structure (Figure 3a). Furthermore, this transformation can be formulated as follows:

$$h^0 = LN(\mathbf{E}), \tag{2}$$

$$\hat{h}^l = LN(h^{l-1} + MHSA(h^{l-1})), \tag{3}$$

$$h^l = LN(\hat{h}^l + FFN(\hat{h}^l)), \tag{4}$$

where $l \in [1, 12]$ is the $l$-th layer transformer, and $h^0$ means the input embedding of BERT, which is built from **E** with a liner function. The outputs of 12-layer transformers are denoted as $[h^1, h^2, \ldots, h^{12}]$. *FFN* includes two linear functions with a *ReLU* activation function between. *MHSA* is the core of the transformer, which has a stacked structure with 12 heads of self-attention. Thus, we can formulate the architecture of attention as follows:

$$\hat{h}^l_{Mj} = softmax(e^l_j)(h^{l-1}\mathbf{W}_V), \tag{5}$$

$$e^l_j = \frac{h^{l-1}\mathbf{W}_Q(h^{l-1}\mathbf{W}_K)^\top}{\sqrt{d}}), \tag{6}$$

$$\hat{h}^l_M = \sum_{j=1}^{N} \hat{h}^l_{Mj}, \tag{7}$$

where parameters $\mathbf{W}_Q$, $\mathbf{W}_K$, and $\mathbf{W}_V$ are the learnable weights for query **Q**, key **K**, and value **V**, and $d$ is the head dimensionality. $\hat{h}^l_{Mj}$ is the single attention, and $\hat{h}^l_M$ denotes the sum of $N$ heads attention (MHSA).
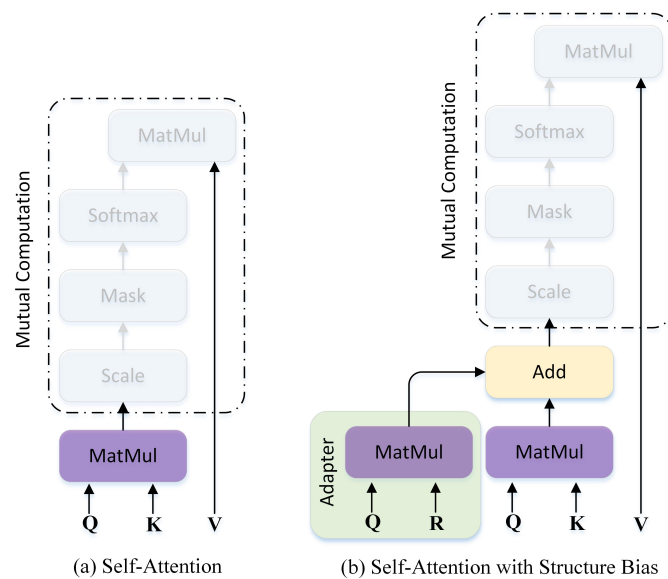


**Figure 3.** The mechanism of structure bias utilized in BERT. Furthermore, **Q**, **K**, and **V** denote the query vector, key vector, and value vector, respectively, which are standardized inputs for the transformer module. **R** indicates the relation distance embedding (Equation (8)).

Inspired by the structure-biased BERT utilized in [24,50], we also introduce it into our MBGCN for generating more informative feature maps. In the optimized approach, self-attention is re-constructed by inserting the relative distance or the dependency between words. Furthermore, the effectiveness of this modification has been obviously proven by

the NLP task [51]. Thus, we describe this change in our model as Equation (8), which can be implemented in Equation (6) directly. Additionally, the procedure is depicted below:

$$
\begin{aligned}
e_j^l &= \frac{h^{l-1}\mathbf{W}_Q(h^{l-1}\mathbf{W}_K + \mathbb{R}^{l-1})^\top}{\sqrt{d}} \\
&= \underbrace{\frac{h^{l-1}\mathbf{W}_Q(h^{l-1}\mathbf{W}_K)^\top}{\sqrt{d}})}_{\text{Raw}} + \underbrace{\frac{h^{l-1}\mathbf{W}_Q(\mathbb{R}^{l-1})^\top}{\sqrt{d}}}_{\text{Bias}},
\end{aligned}
\tag{8}
$$

where $\mathbf{R}^{l-1} \in \mathcal{R}^{k \times k}$ indicates the relative distance embedding between the word pairs of the $k$-th sentence in $(l-1)$-th transformer layer. Note that each dependency embedding is independent from one layer to another layer, but it can be transformed across different heads as an entirety. Additionally, the sketch of the difference between raw self-attention (a) and biased self-attention (b) is shown in Figure 3.

With the backbone encoder of structure-biased BERT, the semantic features $h^l$ is obtained, which provides more accurate contextual information to the module of biaffine attention.

### 3.3. Biaffine Attention

Biaffine attention has been proven to have the ability to capture the relationship among the different words or word pairs [23,52]. Thus, in this paper, we also apply it to predict the relation probability of word pairs in a sentence. To present the process of biaffine attention, the hidden states $h_\varsigma$ and $h_\tau$ of $w_\varsigma$ and $w_\tau$ in $\mathbf{X}$ are extracted from $h^l$. With the aforementioned $\mathbf{MLP_a}$ and $\mathbf{MLP_o}$, the aspect-specific feature $h_\varsigma^a$ (Equation (9)) and opinion-specific feature $h_\tau^o$ (Equation (10)) are obtained, which are adopted into the processing of biaffine attention directly.

$$
h_\varsigma^a = \mathbf{MLP_a}(h_\varsigma),
\tag{9}
$$

$$
h_\tau^o = \mathbf{MLP_o}(h_\tau).
\tag{10}
$$

and the transformation of biaffine attention can be formulated as

$$
g_{\varsigma,\tau} = h_\varsigma^{a\top}\mathbf{U}_o h_\tau^o + \mathbf{U}_a(h_\varsigma^a \oplus h_\tau^o) + \mathbf{b},
\tag{11}
$$

$$
R_{\varsigma,\tau,\xi} = \frac{\exp(g_{\varsigma,\tau,\xi})}{\sum_{\xi=1}^{\Xi}\exp(g_{\varsigma,\tau,\xi})},
\tag{12}
$$

where $\mathbf{U}_o$, $\mathbf{U}_a$, and $\mathbf{b}$ are the trainable weights and biases, and $\oplus$ denotes the operation of concatenation. The relations between $w_\varsigma$ and $w_\tau$ are modeled as $R_{\varsigma,\tau} \in \mathbb{R}^{1 \times \Xi}$. $\Xi$ is the number of relation types. Furthermore, we use $R^{ba}$ to represent the relations obtained in this manner in the following sections.

With the aspect-oriented and opinion-oriented processing of biaffine attention, the probability of relation $R^{ba}$ between the word pairs in a sentence can be modeled effectively. Furthermore, this relation will be integrated with the other four types of linguistic features via GCNs adequately.

### 3.4. Multi-Branch GCN

Except for encoding text as semantic feature maps, it also can be represented in the linguistic feature types. Furthermore, the most widely utilized type is the syntax dependency graph, where the feature is formed in a graph $G = (V, E)$. $V$ is the vertex (i.e., node or word), and $E$ is the edge (i.e., dependency or syntactic relation) between two nodes. Generally, we usually denote this kind of relation through a matrix, namely adjacent matrix $A$. $A_{\tau,\varsigma} = 1$ if the relation between $w_\varsigma$ and $w_\tau$ exists, and $A_{\tau,\varsigma} = 0$ otherwise. In addition, in this paper, we also introduce three other types of linguistic features for each word pair to enhance the contextual representation of the sentence, which are the part-of-speech combination, tree-based distance, and relative position distance (Figure 4). Before

feeding them into the GCN, they are all encoded in the fashion of adjacency matrices. Then, these four types of linguistic features are integrated with semantic features, respectively. For instance, we apply the GCN to integrate the $R^{ba}$ with $R^{sdt}$ encoded from syntactic dependency type tensor $E_{sdt}$, and the process is depicted as follows:

$$R_{\tau,\varsigma}^{sdt} = \sigma(E_{sdt}), \tag{13}$$

$$H_{\tau,\varsigma}^{f1} = f((R_{\tau,\varsigma}^{ba} \oplus R_{\tau,\varsigma}^{sdt})H_{\tau,\varsigma}^{Bert}), \tag{14}$$

where $H^{Bert}$ is obtained from the original contextual representation $h^l$ through a dense layer and a *ReLU* activation layer; $\sigma$ is the function *softmax*. Furthermore, $f(\cdot)$ is an average pooling function applied on the node hidden representations of all channels. To make the extracted relations more accurate, a refining strategy is employed to enhance the relations among words, which can be described as

$$R_{\tau,\varsigma}^{F} = h_{\tau,\varsigma}^{l} \oplus E_{\tau,\varsigma}^{sdt}, \tag{15}$$

$$\hat{F}_{ba+sdt}^{G(\tau,\varsigma)} = R_{\tau,\varsigma}^{F} \oplus R_{\varsigma,\varsigma}^{F} \oplus R_{\tau,\tau}^{F} \oplus H_{\varsigma}^{f1} \oplus H_{\tau}^{f1}, \tag{16}$$

$$F_{ba+sdt}^{G(\tau,\varsigma)} = \sigma(LN(\hat{F}_{ba+sdt}^{G(\tau,\varsigma)})), \tag{17}$$

we use $\oplus$ to concatenate contextual representation $h_{\tau,\varsigma}^l$ and syntactic dependency type tensor $E_{\tau,\varsigma}^{sdt}$. Furthermore, in Equation (16), $R_{\tau,\tau}^F$ and $R_{\varsigma,\varsigma}^F$ are the main diagonal and vice diagonal, which are used to refine the representation $\hat{F}_{ba+sdt}^{G(\tau,\varsigma)}$. Finally, with the operations of a linear layer and a softmax layer, the distribution of probabilities on ten defined relations between $\tau$ and $\varsigma$ is obtained, which is denoted as $F_{ba+sdt}^{G(\tau,\varsigma)}$.
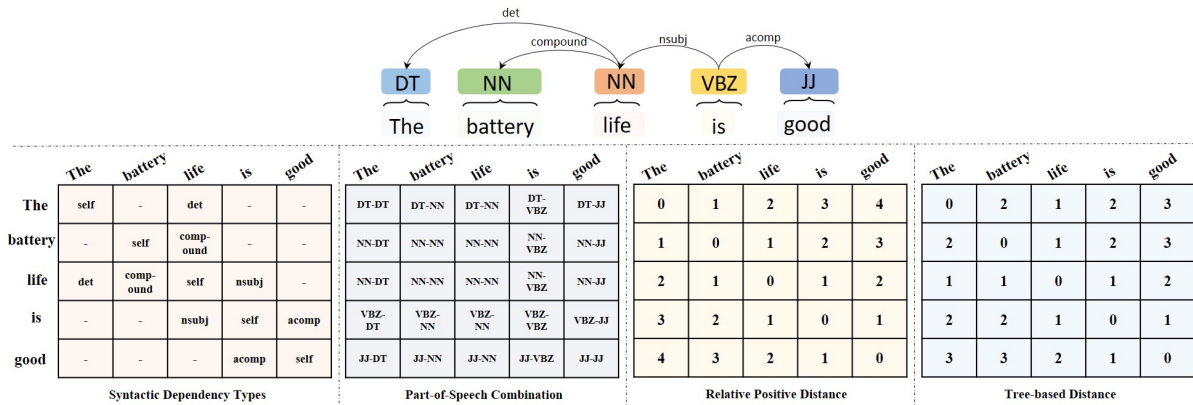


**Figure 4.** The example of four mentioned types of dependency relations among words in reviews.

Similarly, we integrate $R_{\tau,\varsigma}^{ba}$ with $E_{\tau,\varsigma}^{psc}$, $E_{\tau,\varsigma}^{tbd}$, and $E_{\tau,\varsigma}^{rpd}$ via different branches of the GCN to obtain the refined feature representations $F_{ba+psc}^{G(\tau,\varsigma)}$, $F_{ba+tbd}^{G(\tau,\varsigma)}$ and $F_{ba+rpd}^{G(\tau,\varsigma)}$, respectively. Through the operations described in this part, we enhance the contextual feature $R_{\tau,\varsigma}^{ba}$ with four types linguistic features, respectively.

### 3.5. Shallow Interaction and Output Layer

To further enhance the performance of our MBGCN, we apply a shallow interaction layer to fuse the four types of integrated feature representations, which can be depicted as follows:

$$T_{\tau,\varsigma}^{F} = [\alpha, \beta, \gamma, \mu] \cdot [F_{ba+sdt}^{G(\tau,\varsigma)}, F_{ba+psc}^{G(\tau,\varsigma)}, F_{ba+tbd}^{G(\tau,\varsigma)}, F_{ba+rpd}^{G(\tau,\varsigma)}]^{\top}, \tag{18}$$

where $\alpha$, $\beta$, $\gamma$ and $\mu$ are manually selective hyper-parameters to control the weights of different feature representations. Furthermore, $\top$ is the transposition operation for the

related matrix. With this layer, the MBGCN achieves the final textual representation fused from four branches of the GCN, which take five types of textual features into consideration simultaneously.

### 3.6. Training

Generally, the deep learning models are always optimized by minimizing a loss function, and cross entropy is usually applied to complete this work. Without simply applying cross entropy in the proposed MBGCN, due to various contextual information involved, it is necessary to take these into the final fine tuning. For instance, the separated loss $\mathcal{L}_{ba}$ to measure the influence of $R^{ba}$ is modeled as

$$\mathcal{L}_{ba} = -\sum_{\varsigma}^{n}\sum_{\tau}^{n}\sum_{\xi \in \Xi} \mathbb{I}(y_{\varsigma,\tau} = \xi)log(g_{\varsigma,\tau|\xi}), \tag{19}$$

where $\mathbb{I}(\cdot)$ is the indicator function, $y_{\varsigma,\tau}$ is the ground truth relation of word pair $(w_{\varsigma}, w_{\tau})$. Furthermore, $\Xi$ denotes the whole relations set. With a similar operation, the other four separated linguistic features' losses $\mathcal{L}_{psc}$, $\mathcal{L}_{sdt}$, $\mathcal{L}_{tbd}$, and $\mathcal{L}_{rpd}$ are all obtained likewise. Thus, with the prediction, the final loss function $\mathcal{L}$ in the paper is designed as

$$\mathcal{L} = \mathcal{L}_{T^F} + \rho\mathcal{L}_{ba} + \kappa(\mathcal{L}_{psc} + \mathcal{L}_{sdt} + \mathcal{L}_{tbd} + \mathcal{L}_{rpd}), \tag{20}$$

$$\mathcal{L}_{T^F} = \mathbb{I}(Y = \Xi)log(T^F|\Xi) \tag{21}$$

where $\rho$ and $\kappa$ are the manual hyper-parameters to control the influence of each part on the final loss function. Through this manner, our MBGCN can adjust its fine-tuning from six aspects simultaneously.

## 4. Experiments and Discussion

In this section, the results of the conducted experiments are depicted in Tables and Figures, and the corresponding analyses are also provided in detail. We first introduce the widely used ASTE datasets and the related settings of experiments. The detailed experimental results are clearly shown in the analysis secondly.

### 4.1. Datasets

In this paper, extensive experiments are conducted on four benchmarks, namely Laptop14, Restaurant14, Restaurant15, and Restaurant16, which are public and available for ASTE tasks. Furthermore, these four datasets are all collected from the SemEval ABSA challenges [53–55]. It's worth noting that these four datasets are revised by Wu et al. [56] and Xu et al. [22] for ASTE tasks, respectively, which are denoted as $\mathcal{V}_1$ and $\mathcal{V}_2$ in this paper. Moreover, the statistics for these two versions of datasets are shown in Table 2.

**Table 2.** Statistics of two groups of experiment datasets.

| Datasets | | Laptop14 | | Restaurant14 | | Restaurant15 | | Restaurant16 | |
|---|---|---|---|---|---|---|---|---|---|
| | | #S | #T | #S | #T | #S | #T | #S | #T |
| $\mathcal{V}_1$ | train | 899 | 1452 | 1259 | 2356 | 603 | 1038 | 863 | 1421 |
| | dev | 225 | 383 | 315 | 580 | 151 | 239 | 216 | 348 |
| | test | 332 | 547 | 493 | 1008 | 325 | 493 | 328 | 525 |
| $\mathcal{V}_2$ | train | 906 | 1460 | 1266 | 2338 | 605 | 1013 | 857 | 1394 |
| | dev | 219 | 346 | 310 | 577 | 148 | 249 | 210 | 339 |
| | test | 328 | 543 | 492 | 994 | 322 | 485 | 326 | 514 |

Note: #S denotes the number of sentences; #T means the number of triplets contained in the datasets.

### 4.2. Experimental Setup

To conduct the extensive experiments successfully, we use the BERT [57] with structure bias as our review encoder, and it consists of 12 transformer layers, where 12 heads self-attention are included in each layer. Furthermore, the size of the hidden state in self-attention is 768. In addition, the total number of the model's parameters is approximately 110 M. The optimizer AdamW is employed to optimize the training process, where the learning rate is set as $2 \times 10^{-5}$. The dropout is set to 0.5. Moreover, we train our model with 100 epochs with a batch size of 8. The hyper-parameters $\alpha$, $\beta$, $\gamma$, and $\mu$ of fusion work in Equation (18) are set as 0.625, 0.125, 0.125 and 0.125, respectively, and the parameters $\rho$ and $\kappa$ to control the weights of each loss in Equation (16) are set as 0.1 and 0.01, respectively. In addition, the experiments are conducted on a system on NVIDIA GeForce RTX 3080Ti with 12GB of graphics memory. To validate our MBGCN effectively, the widely and popularly used evaluations of Precision (P), Recall (R), and macro-F1 (F1) are employed to present the performance of the proposed approach on four benchmarks.

### 4.3. Baselines

Specifically, to demonstrate the validity of the proposed MBGCN, we make comparisons with several existing SOTA methods designed for ASTE tasks, which are shown as follows:

- GTS-BERT [56] proposes an end-to-end tagging scheme, Grid Tagging Scheme (GTS) with cooperation with BERT, to address the extraction task;
- GTS-CNN [56] is the Grid Tagging Scheme (GTS) that cooperates with CNN;
- GTS-BiLSTM [56] is the Grid Tagging Scheme (GTS) that cooperates with BiLSTM;
- $S^3E^2$ [18] exploits the syntactic and semantic relationships between word pairs in a sentence by a graph-sequence dual representation and modeling paradigm for the ASTE task;
- Peng-two-stage+IOG [56] is the combination of Peng-two-stage [12] and IOG [58];
- Peng-two-stage [12] is a two-stage pipeline model. It extracts both aspect–sentiment pairs and opinion terms in the first stage, and pairs the extraction results into triplets in the second stage;
- OTE-MTL [59] treats the ABSA task as an opinion triplet extraction work, and jointly extracts aspect terms, opinion terms, and parses their sentiment via a multi-task learning framework;
- JET-BERT [22] builds a joint model to extract the triplets using a position-aware tagging approach, which is capable of jointly extracting aspect terms, opinion terms, and their sentiment together;
- BMRC [16] transforms the ASTE task into a Multi-Turn Machine Reading Comprehension (MTMRC) task, and three types of queries are devised to handle the related inputs;
- EMC-GCN [23] transforms the sentence into a multi-channel graph by treating words and edges as nodes and edges, respectively, while ten types of relations for ASTE are defined;
- MuG-Bert [24] proposes an approach, Multi-task learning with Grid decoding (MuG), to integrate the multi-task learning framework with grid triplets decoding from GTS;
- UniASTE$_{BERT}$ [11] proposes an end-to-end method that decomposes ASTE into three subtasks, namely target tagging, opinion tagging, and sentiment tagging. Furthermore, a target-aware tagging scheme is introduced to identify the correspondences between opinion targets and opinion expressions;
- Dual-MRC [15] solves the ASTE task via constructing two machine reading comprehension problems, and trains two BERT-MRC models jointly with parameters sharing.

### 4.4. Main Results

In this subsection, we report the main results of ASTE tasks in Table 3 for version $\mathcal{V}_1$ and Table 4 for version $\mathcal{V}_2$, respectively. According to the results reported in these two Tables, two observations can be concluded and stated as follows.

First, the performances of the PLM-based approaches on ASTE are much better than the normal word2vector-based models. Furthermore, this is quite clear in the comparisons

among GTS-BERT, GTS-CNN, and GTS-BiLSTM. Observing from Table 3, it is obvious that our proposed MBGCN acquires the optimal performance when compared with the previously mentioned SOTA baselines. To be precise, for experimental results in four benchmarks, our MBGCN achieves 72.33%, 57.46%, 59.57%, and 70.43% on the main indicator F1, respectively; while compared with the best baseline EMC-GCN$^\dagger$, the proposed MBGCN obtains 1.13% (72.33–71.20), 0.92% (57.46–56.54), 1.53% (59.57–58.04), and 1.40% (70.42–69.03) improvements on F1 in the four datasets, respectively, and it achieves the new SOTA. This observation from the comparison indicates the effectiveness of our proposed model with the multi-branch framework.

Second, even within the comparison with other BERT-based approaches, the MBGCN also enhances the ASTE performance through its excellent contextual understanding. Notably, in Table 4, the experimental results conducted on $\mathcal{V}_2$ are collected clearly. For the vital evaluation F1, the performances of the MBGCN are improved to 71.37%, 58.89%, 63.07%, and 67.34% in four corpora, respectively, which outperform the mentioned baselines obviously and are only a little worse than EMC-GCN$^\dagger$ on F1 in Restaurant14. Specifically, there is a dramatic increase in F1 in Restaurant15, which is nearly 3.46% (63.07–59.61). Furthermore, in Laptop14 and Restaurant16, the MBGCN also achieves 0.58% (58.89–58.31) and 0.31% (67.34–66.74) increases on the F1 indicator. This can be viewed as direct evidence to support the usefulness of the combination of structure-biased BERT and multi-branch GCNs.

Conclusively, as the results show, structure-biased BERT-based multi-branch GCNs can further boost the performance of ASTE tasks, which is beneficial in excavating the semantic and syntactic information in reviews comprehensively.

**Table 3.** The performance of Multi-branches Graph Convolutional Network (MBGCN) on $\mathcal{V}_1$.

| Models | Restaurant14 | | | Laptop14 | | | Restaurant15 | | | Restaurant16 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **P** | **R** | **F1** | **P** | **R** | **F1** | **P** | **R** | **F1** | **P** | **R** | **F1** |
| Peng-two-stage+IOG© [56] | 58.89 | 60.41 | 59.64 | 48.62 | 45.52 | 47.02 | 51.70 | 46.04 | 48.71 | 59.25 | 58.09 | 58.67 |
| GTS-CNN© [56] | 70.79 | 61.71 | 65.94 | 55.93 | 47.52 | 51.38 | 60.09 | 53.57 | 56.64 | 62.63 | 66.98 | 64.73 |
| GTS-BiLSTM© [56] | 67.28 | 61.91 | 64.49 | 59.42 | 45.13 | 51.30 | 63.26 | 50.71 | 56.29 | 66.07 | 65.05 | 65.56 |
| GTS-BERT© [56] | 70.92 | 69.49 | 70.20 | 57.52 | 51.92 | 54.58 | 59.29 | 58.07 | 58.67 | 68.58 | 66.60 | 67.58 |
| S$^3$E$^2$© [18] | 69.08 | 64.55 | 66.74 | 59.43 | 46.23 | 52.01 | 61.06 | 56.44 | 58.66 | 71.08 | 63.13 | 66.87 |
| Dual-MRC© [15] | - | - | 70.32 | - | - | 55.58 | - | - | 57.21 | - | - | 67.40 |
| EMC-GCN $^\dagger$ [23] | 70.92 | 71.49 | 71.20 | 58.96 | 54.31 | 56.54 | 54.99 | 61.46 | 58.04 | 65.74 | 72.66 | 69.03 |
| MBGCN | 72.89 | 71.79 | **72.33** | 57.30 | 57.62 | **57.46** | 60.76 | 58.42 | **59.57** | 71.68 | 69.22 | **70.43** |

**Note:** The "$\dagger$" denotes that we reproduce the models using released code with original parameters on the dataset. The "©" denotes the results are referred from the original paper. The "-" denotes not mentioned in original paper. And the bold format denotes the optimal performance.

**Table 4.** The performance of MBGCN on $\mathcal{V}_2$. The "$\S$" means the results are retrieved from [8]. The "$\mathbb{1}$" denotes the results are retrieved from [23].

| Models | Restaurant14 | | | Laptop14 | | | Restaurant15 | | | Restaurant16 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **P** | **R** | **F1** | **P** | **R** | **F1** | **P** | **R** | **F1** | **P** | **R** | **F1** |
| Peng-two-stage $^\S$ [12] | 43.24 | 63.66 | 51.46 | 37.38 | 50.38 | 42.87 | 48.07 | 57.51 | 52.32 | 46.96 | 64.24 | 54.21 |
| OTE-MTL $^{\mathbb{1}}$ [59] | 62.00 | 55.97 | 58.71 | 49.53 | 39.22 | 43.42 | 56.37 | 40.94 | 47.13 | 62.88 | 52.10 | 56.96 |
| JET-BERT $^\S$ [22] | 70.56 | 55.94 | 62.40 | 55.39 | 47.33 | 51.04 | 64.45 | 51.96 | 57.53 | 70.42 | 58.37 | 63.83 |
| BMRC $^{\mathbb{1}}$ [16] | 75.61 | 61.77 | 67.99 | 70.55 | 48.98 | 57.82 | 68.51 | 53.40 | 60.02 | 71.20 | 61.08 | 65.75 |
| EMC-GCN $^\dagger$ [23] | 70.35 | 73.14 | **71.72** | 61.48 | 55.45 | 58.31 | 56.33 | 63.30 | 59.61 | 62.46 | 72.32 | 67.03 |
| MuG-BERT© [24] | 68.40 | 67.64 | 68.00 | 58.30 | 52.21 | 55.06 | 60.65 | 54.12 | 57.10 | 66.26 | 67.39 | 66.74 |
| UniASTE$^©_{BERT}$ [11] | 72.14 | 66.30 | 69.09 | 62.24 | 51.77 | 56.51 | 64.83 | 54.31 | 59.06 | 69.06 | 65.53 | 67.22 |
| MBGCN | 67.92 | 75.18 | 71.37 | 59.96 | 57.86 | **58.89** | 62.25 | 63.92 | **63.07** | 63.76 | 71.35 | **67.34** |

Note: The "$\dagger$" denotes that we reproduce the models using released code with original parameters on the dataset. The "©" denotes the results are referred from the original paper. And the bold format denotes the optimal performance.
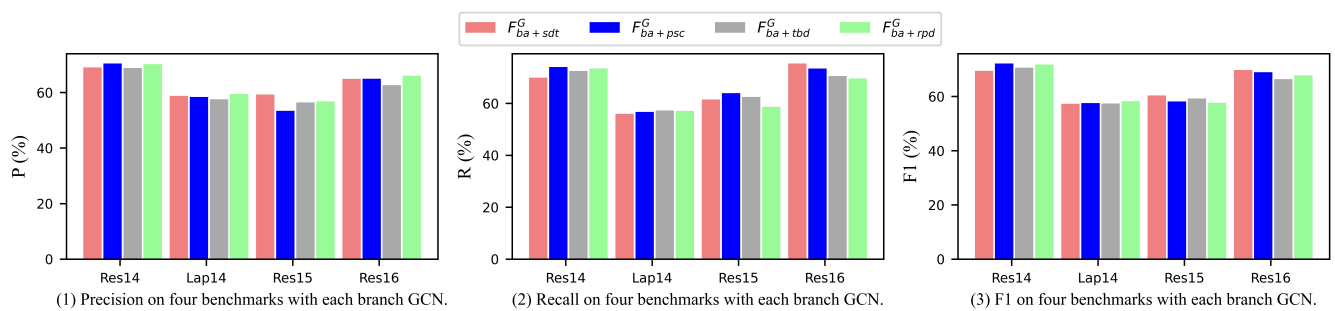
*4.5. Ablation Study*

To further validate the effectiveness of each component in the MBGCN, we conduct ablation experiments and answer the following questions:

- Is the contribution of each linguistic feature equal?
- Does the structure-biased BERT promote the performance of the MBGCN on the ASTE task?

### 4.5.1. Effect of Each Linguistic Feature

We first validate whether each type of linguistic feature is equal to improve the performance of the MBGCN in modeling the textual representation. Accordingly, a single branch GCN is constructed to integrate semantic features achieved by structure-biased BERT with a single linguistic feature. Moreover, the experimental results are shown in Figures 5 and 6.



**Figure 5.** The Results of Ablation Study on $\mathcal{V}_1$. Res14 means Restaurant14, Lap14 means Laptop14, Res15 denotes Restaurant15, and Res16 denotes Restaurant16.

Above all, we compare the effectiveness of the four types of linguistic features on version $\mathcal{V}_1$ of the four benchmarks, respectively. From Figure 5, we can observe that the approach utilizing $F_{ba+psc}^G$ achieves the optimal performance in the experimental results, in which the representation is implemented by semantic feature and part-of-speech combination $R^{psc}$ only. This suggests that linguistic feature $R^{psc}$ can largely enhance the textual representation of semantic features generated by structure-biased BERT. Conversely, the effectiveness of $F_{ba+tbd}^G$ is slightly worse than the observation of the performance of the ablation experiments, but it still contributes to improving the model's capability to extract triplets. In addition, $F_{ba+sdt}^G$ and $F_{ba+rpd}^G$ both have a significant impact on improving the performance of the proposed model. Additionally, extensive experiments are conducted on version $\mathcal{V}_2$ of four datasets, and the results are presented in Figure 6. Observing from the figure, the same conclusion can be obtained from the experimental results based on $F_{ba+psc}^G$ and $F_{ba+tbd}^G$. Therefore, we believe that the triplets extraction task benefits from the cooperation of all four GCN branches directly.
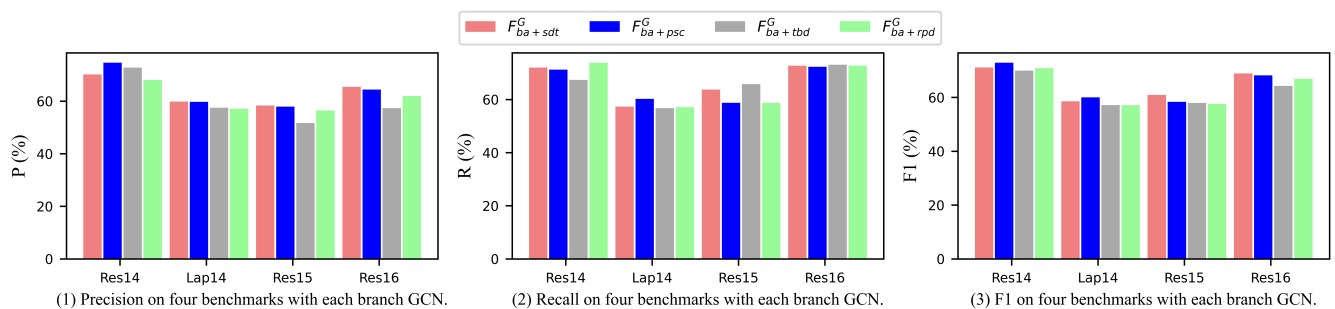


**Figure 6.** The Results of Ablation Study on $\mathcal{V}_2$. Res14 means Restaurant14, Lap14 means Laptop14, Res15 denotes Restaurant15, and Res16 denotes Restaurant16.

### 4.5.2. Effect of Adapter BERT

To validate the influence of structure-biased BERT on textual semantic representation, extensive experiments are conducted on the aforementioned datasets. Table 5 shows the results on version $\mathcal{V}_1$ and $\mathcal{V}_2$ of the four datasets. We can see that structure-biased BERT strengthens the performance of the proposed model in extracting triplets on three datasets (i.e., Restaurant14, Restaurant15, and Restaurant16). In particular, for $\mathcal{V}_1$, the model without structure bias only achieves 71.66%, 58.55%, and 68.52% of F1 on Restaurant14, Restaurant15 and Restaurant16, respectively, which are obviously worse than structure-biased BERT based MBGCN. Furthermore, when it comes to version $\mathcal{V}_2$, the method based on structure-biased BERT also achieves higher F1 scores on three datasets, which are Restaurant14, Laptop14 and Restaurant15, respectively. Furthermore, the corresponding improvements are 1.53% (71.37–69.84), 0.29% (58.89–58.70), and 3.23% (63.07–59.84). Conclusively, the above description indicates that the employment of structure-biased BERT can extract more abundant textual semantic features in the current ASTE task.

**Table 5.** The contribution of adapter to MBGCN for ASTE task on $\mathcal{V}_1$ and $\mathcal{V}_2$.

| Versions | Models | Restaurant14 | | | Laptop14 | | | Restaurant15 | | | Restaurant16 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | P | R | F1 | P | R | F1 | P | R | F1 | P | R | F1 |
| $\mathcal{V}_1$ | MBGCN | 72.89 | 71.79 | **72.33** | 57.30 | 57.62 | 57.46 | 60.76 | 58.42 | **59.57** | 71.68 | 69.22 | **70.43** |
| | *w/o* Structure bias | 70.38 | 72.99 | 71.66 | 60.86 | 54.50 | **57.50** | 56.76 | 60.45 | 58.55 | 64.53 | 73.04 | 68.52 |
| $\mathcal{V}_2$ | MBGCN | 67.92 | 75.18 | **71.37** | 59.96 | 57.86 | **58.89** | 62.25 | 63.92 | **63.07** | 63.76 | 71.35 | 67.34 |
| | *w/o* Structure bias | 69.49 | 70.19 | 69.84 | 59.77 | 57.67 | 58.70 | 58.88 | 60.83 | 59.84 | 65.36 | 71.74 | **68.40** |

Note: The "w/o" denotes the abbreviation for without.

### 4.6. Case Study

To further analyze the role of each linguistic feature in our task, two samples are selected and visualized by attention weights on each word, and they are expressed through Figure 7. As shown in the figure, each row means the visualization of the representation obtained by the single branch GCN, and each column denotes the visualization of each word presented by the four branch GCN. From Figure 7, we can conclude two observations related to the core idea of the proposed model. First, it is obvious that the attention of each branch of the GCN is attracted by the different words in the sentence. For example, in Figure 7b, the key word in branch $F_{ba+sdt}^G$ is "is", while "is" is the last word in sorted attention sequence from $F_{ba+psc}^G$, and $F_{ba+psc}^G$ gives a heavy attention weight to "this". The same conclusion can be summarized from $F_{ba+tbd}^G$ and $F_{ba+rpd}^G$. Second, we find that if one branch misses the specific word, such as "is" in $F_{ba+psc}^G$ and $F_{ba+rpd}^G$ in Figure 7b, another branch of the GCN would provide a higher attention weight on this word, such as $F_{ba+sdt}^G$ and $F_{ba+tbd}^G$. Furthermore, this phenomenon directly corresponds to the core of feature integration. In addition, Figure 7a also provides the same information to us about the attention distributions via a four branch GCN. For this case, our proposed MBGCN completes its work of fusing various branch features to enhance the textual representation and finally improve the performance of the ASTE task.
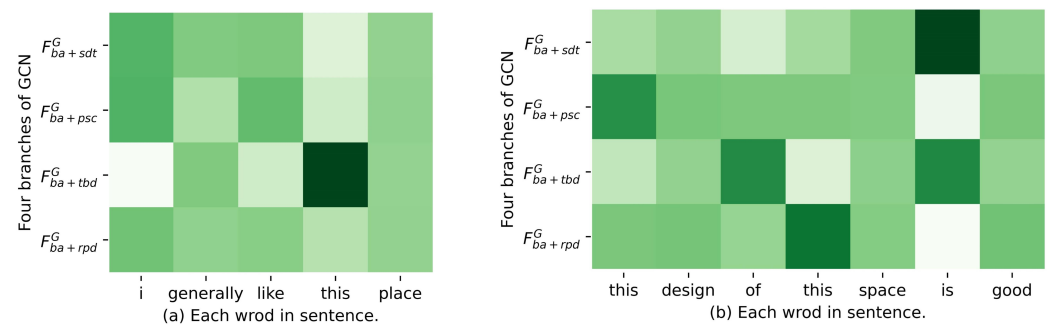
**Figure 7.** Attention distribution on each word in samples for case study.

*4.7. Attempts via Prompt Learning*

Prompt learning is the process of creating a prompt format to guide the training of the model on the downstream tasks [60]. Furthermore, from the investigation of previous research, we learn that creating intuitive templates based on human introspection is the most widely used method that has been adopted in many studies [61–63]. In addition to the mentioned strategies, we also tried to exploit the usefulness of prompt learning in our designed experiments. Following the core idea of prompt learning, the comprehensive prompt template in this task is designed as "the targets are aspect, opinion, sentiment". Thus, the input to the model is remodeled as {REVIEW, the targets are aspect, opinion, sentiment.}, which directly tells the PLM the exact target of the current task.

Furthermore, the relative experiments are conducted both on version $\mathcal{V}_1$ and $\mathcal{V}_2$, and the results are collected in Table 6. From the observations, first, we find that the improvement in the model's performance on $\mathcal{V}_1$ is confined to Laptop14 and Restaurant15, and different declines happen to the experimental results in Restaurant14 and Restaurant16. In other words, the effectiveness of prompt learning is limited for current version $\mathcal{V}_1$ under the framework of our proposed MBGCN. Second, while observing the experimental results on $\mathcal{V}_2$ shown in Table 6, we can learn that the approach with prompts outperforms the baseline MBGCN on three benchmarks clearly. However, it fails in the experiments conducted in Laptop14, which suggests that the optimized model by current prompts is not sensitive to the reviews in Laptop14. Finally, we can conclude that prompt learning with the aforementioned template can improve the model's capability in modeling textual representation in some aspects, but it is not the most proper manner for the current designed framework, which means a lot of effort is essential to improve the performance of prompt learning in the ASTE task.

**Table 6.** The contribution of prompt learning to MBGCN for ASTE task on $\mathcal{V}_1$ and $\mathcal{V}_2$.

| Versions | Models | Restaurant14 | | | Laptop14 | | | Restaurant15 | | | Restaurant16 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | P | R | F1 | P | R | F1 | P | R | F1 | P | R | F1 |
| $\mathcal{V}_1$ | MBGCN | 72.89 | 71.79 | **72.33** | 57.30 | 57.62 | 57.46 | 60.76 | 58.42 | 59.57 | 71.68 | 69.22 | **70.43** |
| | *w/* Prompts | 73.27 | 70.18 | 71.69 | 56.35 | 59.45 | **57.86** | 58.32 | 64.71 | **61.35** | 64.11 | 68.64 | 66.30 |
| $\mathcal{V}_2$ | MBGCN | 67.92 | 75.18 | 71.37 | 59.96 | 57.86 | **58.89** | 62.25 | 63.92 | 63.07 | 63.76 | 71.35 | 67.34 |
| | *w/* Prompts | 73.63 | 71.31 | **72.46** | 58.93 | 56.75 | 57.82 | 64.68 | 62.68 | **63.67** | 65.96 | 72.90 | **69.26** |

Note: The "*w/*" denotes the abbreviation of with.

## 5. Conclusions

In this work, we propose an end-to-end model MBGCN for the ASTE task, which processes Aspect Term Extraction, Opinion Term Extraction, and sentiment polarity prediction in a sentence, simultaneously. For modeling the textual semantic feature more accurately, an optimized attention module is inserted into BERT, namely structure-biased BERT, which is employed to enhance the representation of the specific sentence. In addition,

to emphasize the key features in the generated representation, biaffine attention is utilized to absorb the crucial components from both aspect-oriented and opinion-oriented feature maps. Furthermore, a novel fusion architecture with a multi-branch GCN is proposed to integrate the semantic feature with the linguistic feature. In this part, through each branch GCN, attentive semantic representation is integrated with syntactic dependency types, part-of-speech combination, relative positive distance, and tree-based distance, respectively. Eventually, four branch features are synthesized as an entirety via a designed shallow interaction layer. To validate the effectiveness of our proposed model, we conduct extensive experiments on the benchmark datasets, and the results show that the MBGCN achieves SOTA performances.

Although outstanding performances were achieved by our proposed MBGCN, several limitations still exist, which are the working aims of our future study. First, the working mechanism of prompt learning should be optimized to be more proper for our current task. Second, a more robust integration strategy is essential in our future study for feature fusion.

## Reference

1. Deng, J.; Ren, F. Multi-label Emotion Detection via Emotion-Specified Feature Extraction and Emotion Correlation Learning. *IEEE Trans. Affect. Comput.* **2020**, *9*, 162018–162034. [CrossRef]
2. Liang, B.; Su, H.; Gui, L.; Cambria, E.; Xu, R. Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. *Knowl.-Based Syst.* **2022**, *235*, 107643. [CrossRef]
3. Zhao, A.; Yu, Y. Knowledge-enabled BERT for aspect-based sentiment analysis. *Knowl.-Based Syst.* **2021**, *227*, 107220. [CrossRef]
4. Bie, Y.; Yang, Y. A multitask multiview neural network for end-to-end aspect-based sentiment analysis. *Big Data Min. Anal.* **2021**, *4*, 195–207. [CrossRef]
5. Zhang, Y.; Du, J.; Ma, X.; Wen, H.; Fortino, G. Aspect-based sentiment analysis for user reviews. *Cogn. Comput.* **2021**, *13*, 1114–1127. [CrossRef]
6. Zhao, H.; Huang, L.; Zhang, R.; Lu, Q.; Xue, H. Spanmlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 3239–3248.
7. Akhtar, M.S.; Garg, T.; Ekbal, A. Multi-task learning for aspect term extraction and aspect sentiment classification. *Neurocomputing* **2020**, *398*, 247–256. [CrossRef]
8. Xu, Q.; Zhu, L.; Dai, T.; Yan, C. Aspect-based sentiment classification with multi-attention network. *Neurocomputing* **2020**, *388*, 135–143. [CrossRef]
9. Ying, C.; Wu, Z.; Dai, X.; Huang, S.; Chen, J. Opinion transmission network for jointly improving aspect-oriented opinion words extraction and sentiment classification. In Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, Zhengzhou, China, 14–18 October 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 629–640.
10. Li, Z.; Li, Q.; Zou, X.; Ren, J. Causality extraction based on self-attentive BiLSTM-CRF with transferred embeddings. *Neurocomputing* **2021**, *423*, 207–219. [CrossRef]

11. Chen, F.; Yang, Z.; Huang, Y. A multi-task learning framework for end-to-end aspect sentiment triplet extraction. *Neurocomputing* **2022**, *479*, 12–21. [CrossRef]

12. Peng, H.; Xu, L.; Bing, L.; Huang, F.; Lu, W.; Si, L. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 8600–8607.

13. Xu, L.; Chia, Y.K.; Bing, L. Learning Span-Level Interactions for Aspect Sentiment Triplet Extraction. 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), Virtual, 1–6 August 2021; pp. 4755–4766.

14. Yu Bai Jian, S.; Nayak, T.; Majumder, N.; Poria, S. Aspect sentiment triplet extraction using reinforcement learning. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, 17–21 October 2021; pp. 3603–3607.

15. Mao, Y.; Shen, Y.; Yu, C.; Cai, L. A joint training dual-mrc framework for aspect based sentiment analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 13543–13551.

16. Chen, S.; Wang, Y.; Liu, J.; Wang, Y. Bidirectional machine reading comprehension for aspect sentiment triplet extraction. In Proceedings of the AAAI Conference On Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 12666–12674.

17. Yan, H.; Dai, J.; Ji, T.; Qiu, X.; Zhang, Z. A Unified Generative Framework 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), Virtual, 1–6 August 2021; pp. 2416–2429.

18. Chen, Z.; Huang, H.; Liu, B.; Shi, X.; Jin, H. Semantic and Syntactic Enhanced Aspect Sentiment Triplet Extraction. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Online, 1–6 August 2021; pp. 1474–1483.

19. Wu, S.; Li, B.; Xie, D.; Teng, C.; Ji, D. Neural transition model for aspect-based sentiment triplet extraction with triplet memory. *Neurocomputing* **2021**, *463*, 45–58. [CrossRef]

20. Zhao, Y.; Meng, K.; Liu, G.; Du, J.; Zhu, H. A Multi-Task Dual-Tree Network for Aspect Sentiment Triplet Extraction. In Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Republic of Korea, 12–17 October 2022; pp. 7065–7074.

21. Shi, L.; Han, D.; Han, J.; Qiao, B.; Wu, G. Dependency graph enhanced interactive attention network for aspect sentiment triplet extraction. *Neurocomputing* **2022**, *507*, 315–324. [CrossRef]

22. Xu, L.; Li, H.; Lu, W.; Bing, L. Position-Aware Tagging for Aspect Sentiment Triplet Extraction. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Virtual, 16–20 November 2020; pp. 2339–2349.

23. Chen, H.; Zhai, Z.; Feng, F.; Li, R.; Wang, X. Enhanced Multi-Channel Graph Convolutional Network for Aspect Sentiment Triplet Extraction. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Dublin, Ireland, 22–27 May 2022; pp. 2974–2985.

24. Zhang, C.; Ren, L.; Ma, F.; Wang, J.; Wu, W.; Song, D. Structural Bias for Aspect Sentiment Triplet Extraction. In Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Republic of Korea, 12–17 October 2022; pp. 6736–6745.

25. Tang, F.; Fu, L.; Yao, B.; Xu, W. Aspect based fine-grained sentiment analysis for online reviews. *Inf. Sci.* **2019**, *488*, 190–204. [CrossRef]

26. Xiao, L.; Xue, Y.; Wang, H.; Hu, X.; Gu, D.; Zhu, Y. Exploring fine-grained syntactic information for aspect-based sentiment classification with dual graph neural networks. *Neurocomputing* **2022**, *471*, 48–59. [CrossRef]

27. Consoli, S.; Barbaglia, L.; Manzan, S. Fine-grained, aspect-based sentiment analysis on economic and financial lexicon. *Knowl.-Based Syst.* **2022**, *247*, 108781. [CrossRef]

28. Phan, M.H.; Ogunbona, P.O. Modelling context and syntactical features for aspect-based sentiment analysis. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 3211–3220.

29. Wang, K.; Shen, W.; Yang, Y.; Quan, X.; Wang, R. Relational Graph Attention Network for Aspect-based Sentiment Analysis. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 3229–3238.

30. Zhang, B.; Xu, D.; Zhang, H.; Li, M. STCS lexicon: Spectral-clustering-based topic-specific Chinese sentiment lexicon construction for social networks. *IEEE Trans. Comput. Soc. Syst.* **2019**, *6*, 1180–1189. [CrossRef]

31. Kiritchenko, S.; Zhu, X.; Cherry, C.; Mohammad, S. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 23–24 August 2014; pp. 437–442.

32. Dai, J.; Yan, H.; Sun, T.; Liu, P.; Qiu, X. Does syntax matter? A strong baseline for Aspect-based Sentiment Analysis with RoBERTa. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Mexico City, Mexico, 6–11 June 2021; pp. 1816–1829.

33. Zhang, Y.; Jin, R.; Zhou, Z.H. Understanding bag-of-words model: A statistical framework. *Int. J. Mach. Learn. Cybern.* **2010**, *1*, 43–52. [CrossRef]

34. Tang, H.; Ji, D.; Li, C.; Zhou, Q. Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 6578–6588.

35. Xiao, Z.; Wu, J.; Chen, Q.; Deng, C. 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, Online, 7–11 November 2021; pp. 9193–9200.

36. Zhang, Z.; Zuo, Y.; Wu, J. Aspect Sentiment Triplet Extraction: A Seq2Seq Approach with Span Copy Enhanced Dual Decoder. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2022**, *30*, 2729–2742. [CrossRef]

37. Dai, D.; Chen, T.; Xia, S.; Wang, G.; Chen, Z. Double embedding and bidirectional sentiment dependence detector for aspect sentiment triplet extraction. *Knowl.-Based Syst.* **2022**, *253*, 109506. [CrossRef]

38. Chen, Y.; Zhang, Z.; Zhou, G.; Sun, X.; Chen, K. Span-based dual-decoder framework for aspect sentiment triplet extraction. *Neurocomputing* **2022**, *492*, 211–221. [CrossRef]

39. Mukherjee, R.; Nayak, T.; Butala, Y.; Bhattacharya, S.; Goyal, P. PASTE: A Tagging-Free Decoding Framework Using Pointer Networks for Aspect Sentiment Triplet Extraction. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Online, 7–11 November 2021; pp. 9279–9291.

40. Xu, K.; Li, F.; Xie, D.; Ji, D. Revisiting Aspect-Sentiment-Opinion Triplet Extraction: Detailed Analyses Towards a Simple and Effective Span-Based Model. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2022**, *30*, 2918–2927. [CrossRef]

41. Zhu, X.; Zhu, L.; Guo, J.; Liang, S.; Dietze, S. GL-GCN: Global and local dependency guided graph convolutional networks for aspect-based sentiment classification. *Expert Syst. Appl.* **2021**, *186*, 115712. [CrossRef]

42. Cai, H.; Tu, Y.; Zhou, X.; Yu, J.; Xia, R. Aspect-category based sentiment analysis with hierarchical graph convolutional network. In Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain, 8–13 December 2020; pp. 833–843.

43. Feng, S.; Wang, B.; Yang, Z.; Ouyang, J. Aspect-based sentiment analysis with attention-assisted graph and variational sentence representation. *Knowl.-Based Syst.* **2022**, *258*, 109975. [CrossRef]

44. Li, Y.; Lin, Y.; Lin, Y.; Chang, L.; Zhang, H. A span-sharing joint extraction framework for harvesting aspect sentiment triplets. *Knowl.-Based Syst.* **2022**, *242*, 108366. [CrossRef]

45. Fei, H.; Ren, Y.; Zhang, Y.; Ji, D. Nonautoregressive Encoder-Decoder Neural Framework for End-to-End Aspect-Based Sentiment Triplet Extraction. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–13. early access. [CrossRef]

46. Xu, H.; Shu, L.; Philip, S.Y.; Liu, B. Understanding Pre-trained BERT for Aspect-based Sentiment Analysis. In Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain, 8–13 December 2020; pp. 244–250.

47. Wu, Z.; Ong, D.C. Context-guided bert for targeted aspect-based sentiment analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021; Volume 35, pp. 14094–14102.

48. Zhu, L.; Xu, Y.; Zhu, Z.; Bao, Y.; Kong, X. Fine-Grained Sentiment-Controlled Text Generation Approach Based on Pre-Trained Language Model. *Appl. Sci.* **2022**, *13*, 264. [CrossRef]

49. Mutinda, J.; Mwangi, W.; Okeyo, G. Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network. *Appl. Sci.* **2023**, *13*, 1445. [CrossRef]

50. Shaw, P.; Uszkoreit, J.; Vaswani, A. Self-Attention with Relative Position Representations. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, LA, USA, 1–6 June 2018; Volume 2 (Short Papers), pp. 464–468.

51. Wang, B.; Shin, R.; Liu, X.; Polozov, O.; Richardson, M. RAT-SQL: Relation-Aware Schema Encoding and Linking for Text-to-SQL Parsers. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 7567–7578.

52. Dozat, T.; Manning, C.D. Deep Biaffine Attention for Neural Dependency Parsing. *arXiv* **2016**, arXiv:1611.01734.

53. Pontiki, M.; Galanis, D.; Pavlopoulos, J.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Association for Computational Linguistics, Dublin, Ireland, 23–24 August 2014; pp. 27–35. [CrossRef]

54. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; Androutsopoulos, I. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Co, USA, 4–5 June 2015; pp. 486–495. [CrossRef]

55. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S.; AL-Smadi, M.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; De Clercq, O.; et al. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics, San Diego, CA, USA, 16–17 June 2016; pp. 19–30. . [CrossRef]

56. Wu, Z.; Ying, C.; Zhao, F.; Fan, Z.; Dai, X.; Xia, R. Grid Tagging Scheme for Aspect-oriented Fine-grained Opinion Extraction. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2020, Punta Cana, Dominican Republic, 8–12 November 2020; pp. 2576–2585.

57. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. 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, Minneapolis, USA, 2–7 June 2019; Volume 1 (Long and Short Papers), pp. 4171–4186.

58. Fan, Z.; Wu, Z.; Dai, X.Y.; Huang, S.; Chen, J. Target-oriented Opinion Words Extraction with Target-fused Neural Sequence Labeling. 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); Association for Computational Linguistics: Minneapolis, MN, USA, 2–7 June 2019; pp. 2509–2518. [CrossRef]

59. Zhang, C.; Li, Q.; Song, D.; Wang, B. A Multi-task Learning Framework for Opinion Triplet Extraction. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2020, Punta Cana, Dominican Republic, 8–12 November 2020; pp. 819–828.
60. Liu, P.; Yuan, W.; Fu, J.; Jiang, Z.; Hayashi, H.; Neubig, G. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *arXiv* **2021**, arXiv:2107.13586.
61. Petroni, F.; Rocktäschel, T.; Riedel, S.; Lewis, P.; Bakhtin, A.; Wu, Y.; Miller, A. Language Models as Knowledge Bases? 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), Hong Kong, China, 3–7 November 2019; pp. 2463–2473.
62. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
63. Schick, T.; Schütze, H. Exploiting Cloze-Questions for Few-Shot Text Classification and Natural Language Inference. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Kiev, Ukraine, 19–23 April 2021; pp. 255–269.