*mathematics*

*Article*

# Contextual Graph Attention Network for Aspect-Level Sentiment Classification

**Yuqing Miao [1,2,3,†], Ronghai Luo [1,†], Lin Zhu [1], Tonglai Liu [4,*], Wanzhen Zhang [4,5,*], Guoyong Cai [1,2,3] and Ming Zhou [6]**

1 School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China; miaoyuqing@guet.edu.cn (Y.M.); 20032303090@mails.guet.edu.cn (R.L.); 20032303180@mails.guet.edu.cn (L.Z.); ccgycai@guet.edu.cn (G.C.)
2 Guangxi Key Laboratory of Image & Graphics Intelligent Processing, Guilin 541004, China
3 Guangxi Key Laboratory of Cryptography and Information Security, Guilin 541004, China
4 College of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China
5 Engineering Comprehensive Training Center, Guilin University of Aerospace Technology, Guilin 541004, China
6 Guilin Hivision Technology Company, Guilin 541004, China; zhouming@hivision.com.cn
* Correspondence: tonglailiu@zhku.edu.cn (T.L.); wanzhenzhang@zhku.edu.cn (W.Z.)
† These authors contributed equally to this work.

**Abstract:** Aspect-level sentiment classification aims to predict the sentiment polarities towards the target aspects given in sentences. To address the issues of insufficient semantic information extraction and high computational complexity of attention mechanisms in existing aspect-level sentiment classification models based on deep learning, a contextual graph attention network (CGAT) is proposed. The proposed model adopts two graph attention networks to aggregate syntactic structure information into target aspects and employs a contextual attention network to extract semantic information in sentence-aspect sequences, aiming to generate aspect-sensitive text features. In addition, a syntactic attention mechanism based on syntactic relative distance is proposed, and the Gaussian function is cleverly introduced as a syntactic weight function, which can reduce computational complexities and effectively highlight the words related to aspects in syntax. Experiments on three public sentiment datasets show that the proposed model can make better use of semantic information and syntactic structure information to improve the accuracy of sentiment classification.

**Keywords:** aspect-based sentiment analysis; syntactic relative distance; attention mechanism; graph attention network; BERT; deep learning

**MSC:** 68T07

check for updates

**Citation:** Miao, Y.; Luo, R.; Zhu, L.; Liu, T.; Zhang, W.; Cai, G.; Zhou, M. Contextual Graph Attention Network for Aspect-Level Sentiment Classification. *Mathematics* **2022**, *10*, 2473. https://doi.org/10.3390/math10142473

Academic Editors: Zhijia Zhao and Zhijie Liu

Received: 26 June 2022
Accepted: 14 July 2022
Published: 15 July 2022



## 1. Introduction

Aspect-level sentiment classification (ALSC), also known as target sentiment classification, is a subtask of aspect-based sentiment analysis (ABSA), which aims to identify sentiment polarity expressed by aspects [1] in sentences. As shown in Figure 1, a user can mention aspects "*food*" and "*service*" and express two opposite sentiments over them. The opinion word for "*food*" is "*great*", which is *positive*, and the opinion word for "*service*" is "*dreadful*", which is *negative*. It is inappropriate to assign one sentence-level sentiment label (*positive* or *negative*) for a sentence containing aspects with different sentiment polarities. ALSC can identify the sentiment polarities of different aspects for fine-grained sentiment analysis.

The core of ALSC is to find out opinion words of target aspects accurately. Tang [2] and Wang [3] adopted long short-term memory networks (LSTMs) and an attention mechanism to mine semantic information to achieve semantic alignment between opinion words and target aspects. However, these methods did not fully utilize syntactic structure information. It is difficult to deal with sentences with complex syntactic structures, such as double negative sentences. Zhang [4] and Wang [5] applied graph neural networks (GNNs) to explicitly exploit syntactic structure information to capture long-range dependencies of words. However, these models did not treat sentences as word sequences and ignored semantic information of sentences, making it difficult to cope with sentences that need to understand context words, such as sentences with implicit opinion words.

great **food**$_{pos}$ but the **service**$_{neg}$ was dreadful.

**Figure 1.** A sentence containing aspects with different sentiment polarities.

In sentiment classification tasks, context words that get closer to the target aspects are generally more significant for identifying sentiment polarities. Based on the shortest path length on dependency trees between words and aspects (syntactic relative distance, SRD), He [6] proposed an attention mechanism to find context words that are close to target aspects in syntax. However, this attention mechanism has a high computational cost, and the assigned syntactic weight decreases sharply with SRD increases, resulting in sentiment information loss.

To tackle these problems, a novel aspect-level sentiment classification model (contextual graph attention network, CGAT) is proposed, which combines a syntactic attention mechanism and a contextual graph attention network.

The contributions of our work are summarized as follows.

- A contextual graph attention network is proposed. After taking target aspects as roots to reconstruct dependency trees, the contextual graph attention network applies two graph attention networks to mine syntactic structure information related to aspects and adopts a contextual attention network to extract semantic information, aiming to enhance sentiment expression of target aspects.
- A syntactic attention mechanism based on SRD is proposed, which has low computational complexity and cleverly introduces the Gaussian function as a syntactic weight function. The syntactic attention mechanism avoids the loss of sentiment information from weight decaying sharply and effectively focuses on opinion words of aspects.

The remainder of this paper is organized as follows. In Section 2, recent works in ALSC are described. In Section 3, the implementation details of the proposed model CGAT are explicitly introduced. In Section 4, extensive experiments are reported to assess the performance of CGAT. In Section 5, our work is summarized.

## 2. Related Work

Many aspect-level sentiment classification works employ bidirectional long short-term memory (Bi-LSTM) and an attention mechanism to extract semantic information of sentences. Huang [7] used a Bi-LSTM to model sentences and aspects and adopted an attention mechanism to capture interactive features between them. Chen [8] utilized multiple attention layers to capture aspect-related long-range dependency information. Fan [9] presented a multi-grained attention mechanism to capture interaction features between sentences and aspects. Park [10] adopted a Bi-LSTM to encode left contexts and right contexts of aspects separately and applied a gated recurrent unit (GRU) to extract aspect sentiment features.

Since convolutional neural networks (CNNs) have the advantage of modeling local features, there are some attempts to exploit CNNs for ALSC. For modeling sentences and aspects by utilizing a Bi-LSTM, Cheng [11] adopted different sizes of convolution kernels to extract local features of words. After dividing sentences into multiple regions, Liu [12] applied a CNN to extract local features of different regions and employed a hierarchical LSTM to mine temporal relations of sentences and associations between sentences.

Cognitive linguistics [13] holds that syntactic structure information helps to interpret semantics expressed by sentences. To utilize syntactic structure information to assist text analysis, early approaches tried hand-crafted coders [14], though they required careful design. Afterward, for mining more comprehensive syntactic structure information, Dozat [15] used deep learning to parse dependencies between words without manual design. Moreover, to detect sentiment and emotions around aspects "gender" and "gap" in 10,000 tweets, Stella [16] used a multi-layer perceptron to extract syntactic structure information and build a network of non-stopwords. Similarly, to investigate sentiments and emotions around aspects "love" and "live" in suicide notes, Sofia [17] used syntactic structure information to construct subject–verb–object triads.

Recently, because of the flexibility of graph neural networks (GNNs) in dealing with complex topological structures, GNNs have been widely applied to mine syntactic structure information of sentences. Zhang [4] adopted a graph convolutional network (GCN) to capture syntactic structure information on dependency trees. Lu [18] designed a gated mechanism based on Bi-LSTM to guide the encoding of sentiment information related to aspects, followed by a GCN to capture long-range dependencies of words. After concatenating dependency relations and their two sides words, Du [19] employed a GCN and a multi-head attention mechanism to capture aspect-related sentiment information. Wang [5] reconstructed dependency trees by taking aspects as roots and applied BERT and graph attention networks (GATs) to explicitly encode dependency relations. Wang [20] applied BERT and attention mechanisms to fuse syntactic structure information extracted by GCNs with other information, such as semantics, location, parts of speech, and aspects.

Some studies show that words in different positions contribute differently to recognizing aspect sentiments. The closer to the target aspect, the more important the word may be. Tang [21] assigned weights to words based on their index in the sentence. Zhang [4], Chen [8], Fan [9], Park [10], Wang [20], and Chen [22] assigned weights to words based on the number of words between words and aspects (position distance, PD). He [6] and Su [23] assigned weights to words based on the shortest path lengths on dependency trees between words and aspects (syntax relative distance, SRD).

## 3. Method

The overall framework of proposed model CGAT is illustrated in Figure 2. It essentially consists of three components: an input layer, a syntactic attention mechanism, and a contextual graph attention network.

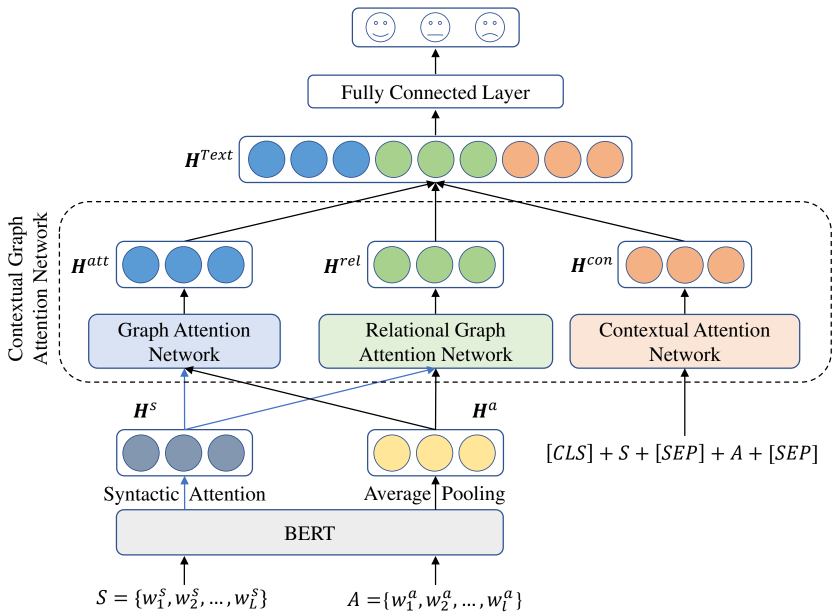


**Figure 2.** The overall framework of the proposed model CGAT.

The input layer converts each word into word embedding via BERT to obtain sentence and aspect representations. The syntactic attention mechanism calculates syntactic weights based on SRD and assigns them to each word. The contextual graph attention network applies two graph attention networks to extract syntactic structure information and utilizes a contextual attention network to extract semantic information from sentence-aspect sequences.

Our task is to predict the sentiment polarity of the target aspect, which can be *positive*, *neutral*, or *negative*.

### 3.1. Input Layer

Given an aspect–sentence pair $(S, A)$, a target aspect (i.e., words or phrases) $A = \{w_1^a, w_2^a, \cdots, w_l^a\}$ containing $l$ words is a sub-sequence of the sentence $S = \{w_1^s, w_2^s, \cdots, w_L^s\}$ with $L$ words. The details of the input layer are as follows.

Firstly, a sentence $S$ and its target aspect $A$ are input into BERT [24], one of the state-of-the-art word embedding models, and $\boldsymbol{X}^s = \{\boldsymbol{x}_1^s, \boldsymbol{x}_2^s, \cdots, \boldsymbol{x}_L^s\}$ and $\boldsymbol{X}^a = \{\boldsymbol{x}_1^a, \boldsymbol{x}_2^a, \cdots, \boldsymbol{x}_l^a\}$ are obtained, where $\boldsymbol{x}_i \in \mathbb{R}^e$ is an $e$-dimensional word vector.

$$\boldsymbol{X}^s = \mathrm{BERT}(S) \tag{1}$$

$$\boldsymbol{X}^a = \mathrm{BERT}(A) \tag{2}$$

Secondly, $\boldsymbol{X}^s \in \mathbb{R}^{L\times e}$ and $\boldsymbol{X}^a \in \mathbb{R}^{l\times e}$ are fed into a low-dimensional linear space, and the hidden layer representations of a sentence, $\boldsymbol{V}^s = \{\boldsymbol{v}_1^s, \boldsymbol{v}_2^s, \cdots, \boldsymbol{v}_L^s\}$ and an aspect, $\boldsymbol{V}^a = \{\boldsymbol{v}_1^a, \boldsymbol{v}_2^a, \cdots, \boldsymbol{v}_l^a\}$, are obtained.

$$\boldsymbol{V}^s = \boldsymbol{X}^s\boldsymbol{W}^s + \boldsymbol{b}^s \tag{3}$$

$$\boldsymbol{V}^a = \boldsymbol{X}^a\boldsymbol{W}^a + \boldsymbol{b}^a \tag{4}$$

where $\boldsymbol{W}^s$, $\boldsymbol{W}^a \in \mathbb{R}^{e\times d}$, $\boldsymbol{b}^s$, $\boldsymbol{b}^a \in \mathbb{R}^d$, and $d$ are the dimensions of the hidden layer.

Finally, average pooling on $\boldsymbol{V}^a \in \mathbb{R}^{l\times d}$ is used to acquire aspect representation $\boldsymbol{H}^a \in \mathbb{R}^d$.

$$\boldsymbol{H}^a = \mathrm{mean}(\boldsymbol{V}^a) \tag{5}$$

### 3.2. Syntactic Attention Mechanism

Researchers [4,10,20,22,25] showed that a context word closer to the target aspect is generally more significant than a farther one for determining sentiment polarity. In the proposed model, the syntactic attention mechanism calculates syntactic weights based on SRD and assigns them to $\boldsymbol{V}^s$ to acquire the sentence representation with syntactic weight $\boldsymbol{H}^s$.

As shown in Figure 3, the opinion word of the target aspect "*size*" is "*busy*", and the number of words between them is 7, which is remote. It indicates that weighting context words based on position distance (PD) may fail to focus on those words close to target aspects in syntax.

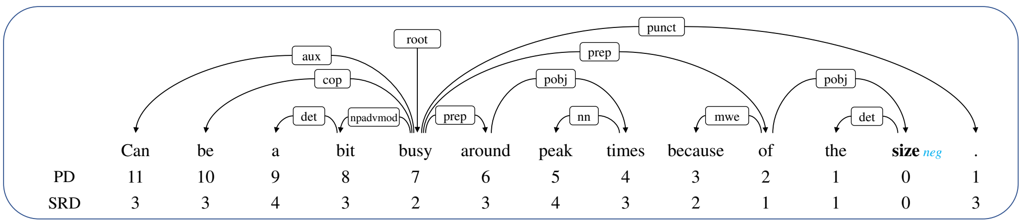


**Figure 3.** Position distance (PD) and syntactic relative distance (SRD).

After generating a syntactic dependency tree by parsing the sentence, the shortest path length between "*size*" and "*busy*" on the dependency tree is two. This indicates that the syntactic relative distances (SRD) between words and aspects can effectively emphasize those context words remote in PD but close to aspects in syntax.

Some studies [6,23,25] found that a context word far away from aspects should not be abandoned, as it may contribute to recognizing aspect sentiment. Further analyses found that the SRDs and the importance of words approximately satisfy a negative correlation when the amount of data reaches a certain scale. The remoter the SRD between words and aspects, the smaller the contribution of words.

Based on these analyses, the Gaussian function is cleverly introduced into the syntactic attention mechanism as a syntactic weight function, aiming to calculate syntactic weights based on SRDs.

$$\rho_i^s = g(d_i^s) = A * \mathrm{e}^{-\frac{(d_i^s-\mu)^2}{2\sigma^2}} \tag{6}$$

$$d_i^s = \begin{cases} 0 & n = 0, \\ n & 1 \leq n < \omega, \\ \omega & n \geq \omega. \end{cases} \tag{7}$$

where peak $A = 1$, mean $\mu = 0$, standard deviation $\sigma = 5$, attention window size $\omega = 4$, $n$ denotes SRD, and $d_i^s$ is the improved SRD. When $1 \leq n < \omega$, $d_i^s = n$, syntactic weight $\rho_i^s$ decreases slowly with increases in SRD. When $n \geq \omega$, set $d_i^s = \omega$, and the contribution of the word to aspect sentiment recognition is considered.

Without involving other factors, the calculation of syntactic weight only relates to SRD, which reduces computational complexity. In addition, syntactic weight decreases slowly with increases in SRD, avoiding the loss of sentiment information and enabling the syntactic attention mechanism to pay more attention to aspect-related words.

After calculating the syntactic weight of each word in the sentence through the syntactic weight function, a syntactic weight vector $\boldsymbol{\rho}^s = \{\rho_1^s, \rho_2^s, \cdots, \rho_L^s\}$ can be acquired by connecting. Finally, the corresponding elements of $\boldsymbol{\rho}^s \in \mathbb{R}^L$ and $\boldsymbol{V}^s \in \mathbb{R}^{L\times d}$ are multiplied to obtain the sentence representation with syntactic weight $\boldsymbol{H}^s \in \mathbb{R}^d$.

$$\boldsymbol{H}^s = \boldsymbol{\rho}^s \odot \boldsymbol{V}^s \tag{8}$$

where $\odot$ is element-wise multiplication.

### *3.3. Contextual Graph Attention Network*

The contextual graph attention network consists of a graph attention network, a relational graph attention network, and a contextual attention network, in order to extract syntactic structure information and semantic information.

#### 3.3.1. Aspect-Oriented Dependency Tree

A dependency tree is a graph structure that describes dependency relations between words, which contains abundant syntactic structure information. However, it is commonly not aspect-oriented, and there is a lot of redundant information that is irrelevant to the target aspect. Therefore, to capture aspect-related syntactic structure information, the dependency tree is reconstructed as an aspect-oriented dependency tree to highlight the syntactic relationships between words and aspects.

Figure 4 is the reconstructed dependency tree, and the details of reshaping are as follows.

Firstly, taking a target aspect as a root node. If the aspect contains multiple words, it would be treated as an entity and maintain dependencies with other words.

Secondly, retaining dependency relations that connect directly to the target aspect and removing other dependency relations.

Finally, establishing new connections ($n$ connected) with the target aspect for those words without dependency relations, where $n$ represents SRD.

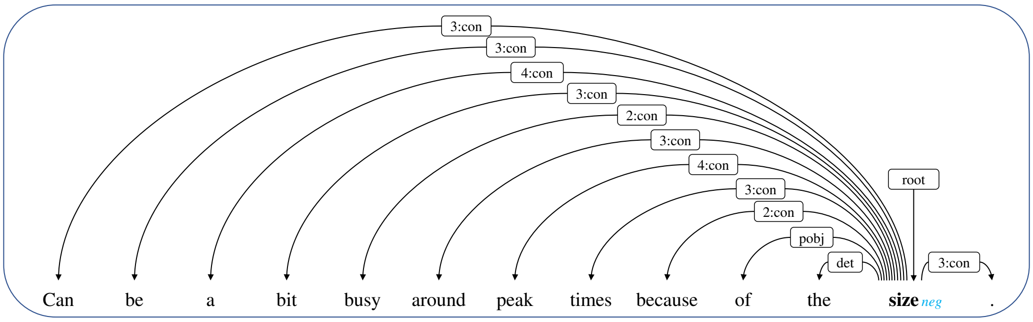


**Figure 4.** An aspect-oriented dependency tree.

#### 3.3.2. Graph Attention Network

After reconstructing dependency trees, target aspects connect with context words. A graph attention network can aggregate sentiment information to target aspects.

A dependency tree can be represented as a graph structure with $L$ nodes, where each node denotes a word in the sentence. Additionally, the neighborhood nodes of node $i$ can be denoted by $\mathcal{N}_i$. Applying a GAT and a multi-head attention mechanism can aggregate neighborhood node representations and iteratively update each node representation.

$$\boldsymbol{h}_{att_i}^{n+1} = \|_{k=1}^{K} \sum_{j \in \mathcal{N}_i} \alpha_{ij}^{n} \boldsymbol{h}_{att_j}^{n} \boldsymbol{W}_k^n \tag{9}$$

$$\alpha_{ij}^{n} = \mathrm{softmax}(\boldsymbol{h}_{att_i}^{n} \cdot \boldsymbol{h}_{att_j}^{n}) \tag{10}$$

where $\boldsymbol{h}_{att_i}^{n+1} \in \mathbb{R}^d$ is the attention head representation of node $i$ at layer $n$+1, $\|_{k=1}^{K}$ denotes the vector concatenation operation, $\alpha_{ij}^{n} \in \mathbb{R}^1$ is a normalized attention coefficient between node $i$ and node $j$ at layer $n$, and $\boldsymbol{W}_k^n \in \mathbb{R}^{d \times d/K}$ is a weight matrix of the $k$-th attention head at layer $n$. Additionally, we adopt a dot-product operation to compute $\alpha_{ij}^{n}$.

Initialize $\boldsymbol{h}_{att_i}^{0}$ as $\boldsymbol{h}_i^s \in \boldsymbol{H}^s$. The final node representation $\boldsymbol{H}_{att}^{n} = \{\boldsymbol{h}_{att_1}^{n}, \boldsymbol{h}_{att_2}^{n}, \cdots, \boldsymbol{h}_{att_L}^{n}\}$ can be obtained at the last GAT layer, where $H_{att}^{n} \in \mathbb{R}^{L \times d}$. The interactive features between $\boldsymbol{H}_{att}^{n}$ and aspect representation $\boldsymbol{H}^a \in \mathbb{R}^d$ are captured; we name the multi-head text representation $\boldsymbol{H}^{att} \in \mathbb{R}^d$.

$$\boldsymbol{H}^{att} = \boldsymbol{H}^a \boldsymbol{W}^k \boldsymbol{H}_{att}^{n} + \boldsymbol{b}^k \tag{11}$$

where $\boldsymbol{W}^k \in \mathbb{R}^{d \times L}$ is a weight matrix and $\boldsymbol{b}^k \in \mathbb{R}^d$ is a bias.

#### 3.3.3. Relational Graph Attention Network

After reconstructing dependency trees, dependency relations connect with target aspects and context words, indicating their syntactic dependency or relative distance. The relational graph attention network can aggregate the syntactic structure information to the target aspect.

A dependency relation between node $i$ and node $j$ is mapped into the relation embedding $\boldsymbol{r}_{ij} \in \mathbb{R}^d$. Similarly, we employ another GAT and multi-head attention mechanism to encode each dependency relation.

$$\boldsymbol{h}_{rel_i}^{n+1} = \|_{m=1}^{M} \sum_{j \in N_i} \beta_{ij} \boldsymbol{h}_{rel_j}^{n} \boldsymbol{W}_m^n \tag{12}$$

$$\beta_{ij} = \mathrm{softmax}(\boldsymbol{r}_{ij} \boldsymbol{W}^r + b^r) \tag{13}$$

where $\boldsymbol{h}_{rel_i}^{n+1} \in \mathbb{R}^d$ is the relational head representation of node $i$ at layer $n$+1, $\beta_{ij} \in \mathbb{R}^1$ is a normalized attention coefficient between node $i$ and node $j$, $\boldsymbol{W}_m^n \in \mathbb{R}^{d \times d/M}$, $\boldsymbol{W}^r \in \mathbb{R}^{d \times 1}$, and $b^r \in \mathbb{R}^1$.

Similarly, we initialize $\boldsymbol{h}_{rel}^{0}$ as $\boldsymbol{h}_i^s \in \boldsymbol{H}^s$. The final dependency relation representation $\boldsymbol{H}_{rel}^n = \{\boldsymbol{h}_{rel_1}^n, \boldsymbol{h}_{rel_2}^n, \cdots, \boldsymbol{h}_{rel_L}^n\}$ can be obtained at the last GAT layer, where $\boldsymbol{H}_{rel}^n \in \mathbb{R}^{L\times d}$. Finally, we align the dimensions of $\boldsymbol{H}_{rel}^n$ with $\boldsymbol{H}^a$ to generate multi-head relation representation $\boldsymbol{H}^{rel} \in \mathbb{R}^d$.

$$\boldsymbol{H}^{rel} = \boldsymbol{H}^a \boldsymbol{W}^m \boldsymbol{H}_{rel}^n + \boldsymbol{b}^m \tag{14}$$

where $\boldsymbol{W}^m \in \mathbb{R}^{d\times L}$ and $\boldsymbol{b}^m \in \mathbb{R}^d$.

#### 3.3.4. Contextual Attention Network

After constructing a sentence-aspect sequence, the contextual attention network adopts BERT to adaptively model words based on context words, aiming to fully extract semantic information of the sentence and obtain semantic features $\boldsymbol{H}^{con}$.

Firstly, construct a sentence-aspect sequence $SA$, "[$CLS$] + $S$ +[$SEP$]+ $A$ +[$SEP$]", where the symbol "[$CLS$]" aggregates semantic information and "[$SEP$]" separates sentences and aspects.

Secondly, input $SA$ to BERT and take out the hidden layer feature of the first symbol "[$CLS$]."

$$\boldsymbol{X}^c = \mathrm{BERT}(SA) \tag{15}$$

$$\boldsymbol{H}^{cls} = \boldsymbol{X}^c[1] \tag{16}$$

Finally, linearly transform $\boldsymbol{H}^{cls} \in \mathbb{R}^e$ to obtain the aspect-sensitive semantic feature $\boldsymbol{H}^{con} \in \mathbb{R}^d$.

$$\boldsymbol{H}^{con} = \boldsymbol{H}^{cls} \boldsymbol{W}^c + \boldsymbol{b}^c \tag{17}$$

where $\boldsymbol{W}^c \in \mathbb{R}^{e\times d}$ and $\boldsymbol{b}^c \in \mathbb{R}^d$.

### 3.4. Sentiment Classification

The final text representation $\boldsymbol{H}^{Text} \in \mathbb{R}^{3d}$ can be obtained by concatenating the output of the contextual graph attention network: $\boldsymbol{H}^{att}$, $\boldsymbol{H}^{rel}$ and $\boldsymbol{H}^{con}$.

$$\boldsymbol{H}^{Text} = \boldsymbol{H}^{att} \| \boldsymbol{H}^{rel} \| \boldsymbol{H}^{con} \tag{18}$$

where $\|$ denotes the vector concatenation.

Feed $\boldsymbol{H}^{Text}$ to a fully connected layer and softmax activation function to obtain sentiment classification results $\boldsymbol{p}(a)$.

$$\boldsymbol{H} = \sigma(\sigma(\boldsymbol{H}^{Text} \boldsymbol{W}_1^h + \boldsymbol{b}_1^h) \boldsymbol{W}_2^h + \boldsymbol{b}_2^h) \tag{19}$$

$$\boldsymbol{p}(a) = \mathrm{softmax}(\boldsymbol{H} \boldsymbol{W}^p + \boldsymbol{b}^p) \tag{20}$$

where $\boldsymbol{H} \in \mathbb{R}^d$, $\boldsymbol{W}_1^h \in \mathbb{R}^{3d\times d}$, $\boldsymbol{W}_2^h \in \mathbb{R}^{d\times d}$, $\boldsymbol{W}^p \in \mathbb{R}^{d\times 3}$, $\boldsymbol{b}_1^h \in \mathbb{R}^d$, $\boldsymbol{b}_2^h \in \mathbb{R}^d$, and $\boldsymbol{b}^p \in \mathbb{R}^3$. Additionally, $\sigma$ is a HardSwish activation function [26].

The proposed model employs standard cross-entropy and L1 normalization as the objective function to optimize network parameters. The L1 normalization reduces the number of non-zero parameters to prevent overfitting.

$$L(\theta) = -\sum_{(S,A)\in D} \log \boldsymbol{p}(a) + \gamma * \mathrm{L1}(\theta) \tag{21}$$

where $D$ denotes the collection of sentence-aspect pairs, $A$ represents the aspects appearing in sentence $S$, $\theta$ contains all the trainable parameters, and $\gamma$ is a normalized coefficient.

## 4. Experiments

### 4.1. Datasets and Settings

To assess the classification performance of the proposed model CGAT, experiments were conducted on three public sentiment analysis datasets. Two of them contain reviews from the SemEval 2014 Task [27], *Restaurant* and *Laptop*, and the third dataset contains

*Twitter* data collected by Dong [28] in 2014. The basic statistics of the three datasets are summarized in Table 1.

In the experiment, word vector dimension $e$ was set to 768, and hidden layer dimensions $d$ were set to 400. The Biaffine parser [15] was adopted to parse dependency relations to generate syntactic dependency trees. The numbers of attention head $K$ and relational head $M$ were both 6 in the contextual graph attention network. The L1 normalization coefficient $\gamma$ was $1 \times 10^{-8}$. The batch size was 16, and the learning rate was $5 \times 10^{-5}$. All training of the model was conducted on GPU (NVIDIA Tesla P100).

**Table 1.** The basic statistics of three datasets.

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

### *4.2. Baselines*

The original experiment results of LSTM+SynATT+TarRep, RAM, MGAN, HRT_Bi, ASGCN, BE-GCN, and Mem+BERT on three datasets are cited. To ensure the fairness of experiments, model R-GAT+BERT was reimplemented in the same experiment environment as the proposed model CGAT.

**LSTM+SynATT+TarRep** [6] pre-sets $K$ representative aspects to extract important information related to target aspects and assigns weights to context words based on SRD and association degree with target aspects.

**RAM** [8] adopts multi-layer attention to capture aspect-related sentiment information and utilizes position information to assign weights to words.

**MGAN** [9] designs a multi-grained attention mechanism to capture interaction features between context words and aspects and assigns weights based on position distance.

**HRT_Bi** [10] utilizes a Bi-LSTM to encode the left contexts and the right contexts of aspects separately and uses a GRU to mine aspect-related sentiment information.

**ASGCN** [4] employs a GCN to extract syntactic structure information.

**BE-GCN** [19] combines words and their dependencies and adopts a GCN and a multi-head attention mechanism to capture aspect-related sentiment information.

**Mem+BERT** [20] adopts a GCN to extract syntactic structure information and utilizes an attention mechanism to fuse syntactic structure information, semantic information, location information, parts of speech, and aspects.

**R-GAT+BERT** [5] reconstructs ordinary dependency trees into aspect-oriented and exploits two GATs to model dependency relations.

### *4.3. Results and Analysis*

#### 4.3.1. Main Results

The performances of different models on the three datasets are shown in Table 2, and there several observations can be noted.

Firstly, the proposed model CGAT outperformed other models in accuracy and MF1.

Secondly, on the three datasets, compared with LSTM+SynATT+TarRep, RAM, MGAN, HRT_Bi, ASGCN, and BE-GCN, CGAT achieved average increases of 5.30%, 5.82%, and 5.59% in accuracy, and 8.32%, 5.43%, and 6.16% in average MF1. The main reason is that CGAT is based on BERT, which can adaptively model words according to context words and effectively extract semantic information.

Thirdly, the performance of CGAT significantly improved compared with Mem+BERT and R-GAT+BERT. This phenomenon shows that the proposed syntactic attention mechanism can effectively highlight aspect-related context words, which is beneficial for mining syntactic structure information.

**Table 2.** Comparison results of different models on the three datasets.

| Category | Model | PD | SRD | Restaurant | | Laptop | | Twitter | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | ACC | MF1 | ACC | MF1 | ACC | MF1 |
| Att. | LSTM+synATT+TarRep | | √ | 80.63 | 71.32 | 71.94 | 69.23 | - | - |
| | RAM | √ | | 80.23 | 70.80 | 74.49 | 71.35 | 69.36 | 67.30 |
| | MGAN | √ | | 81.25 | 71.94 | 75.39 | 72.47 | 72.54 | 70.81 |
| | HRT_Bi | √ | | 81.96 | 74.09 | 74.45 | 70.83 | 73.27 | 71.98 |
| Syn. | ASGCN | √ | | 80.77 | 72.02 | 75.55 | 71.05 | 72.15 | 70.40 |
| | BE-GCN | | | 80.86 | 72.18 | 75.70 | 71.82 | 72.01 | 70.55 |
| | Mem+BERT | √ | | 85.18 | 77.32 | 78.68 | 73.93 | 72.98 | 72.08 |
| | R-GAT+BERT | | | 85.09 | 79.22 | 79.00 | 75.78 | 75.15 | 73.97 |
| Att.+Syn. | CGAT(ours) | | √ | **86.25** | **80.38** | **80.41** | **76.48** | **77.46** | **76.37** |

#### 4.3.2. Ablation Study

To evaluate the effectiveness of different components in CGAT, four groups of ablation experiments were designed. The specific description of each group is as follows.

**CGAT/S** removes syntactic attention mechanism S.

**CGAT/G** removes graph attention network G.

**CGAT/R** removes relational graph attention network R.

**CGAT/C** removes contextual attention network C.

As shown in Table 3, removing any components would lead to decreases in accuracy and MF1.

**Table 3.** Ablation results of different components.

| Model | Restaurant | | Laptop | | Twitter | |
|---|---|---|---|---|---|---|
| | ACC | MF1 | ACC | MF1 | ACC | MF1 |
| CGAT/S | 84.55 | 76.77 | 77.74 | 73.94 | 72.69 | 70.58 |
| CGAT/G | 84.29 | 76.42 | 78.37 | 74.08 | 73.12 | 71.97 |
| CGAT/R | 86.07 | 79.70 | 79.62 | 76.08 | 74.42 | 72.65 |
| CGAT/C | 80.98 | 73.07 | 74.92 | 70.22 | 72.25 | 70.58 |
| CGAT | **86.25** | **80.38** | **80.41** | **76.48** | **77.46** | **76.37** |

Removing syntactic attention mechanism S, the accuracy of CGAT was decreased by 1.70%, 2.67%, and 4.77% on the three datasets, respectively. This indicates that the syntactic attention mechanism S can effectively highlight words close to aspects in syntax, which is beneficial to improving the performance of sentiment classification.

Removing G, R, or C would result in varying degrees of accuracy drop. This phenomenon shows that the contextual attention network can effectively extract syntactic structure information and semantic information, which is helpful for sentiment classification. In particular, removing C would lead to a sharp drop in accuracy, indicating that semantic information extracted by C is crucial for recognizing aspect sentiment.

#### 4.3.3. Effect of Different Parameters

To observe the effects of different peaks $A$ and attention window sizes $\omega$ in the syntactic attention mechanism, two series of experiments were designed.

(1) Effect of different peaks

To observe the effects of different peaks, we designed 10 groups of comparison experiments in which peak $A$ increased from 0.6 to 1.5 in Equation (6). The experimental results are shown in Figure 5a.

As depicted in Figure 5a, the overall trend of the accuracy is an increase at first and then a decrease; the highest point is when the peak $A$ is 1. Therefore, peak $A$ was selected as 1.

(2) Effect of different attention window sizes

To observe the effects of different attention window sizes, we designed three groups of comparison experiments in which the window size $\omega$ increased from 3 to 5 in Equation (7). The experimental results are shown in Figure 5b.

As depicted in Figure 5b, CGAT reached the highest accuracy on three datasets when the attention window size $\omega$ was 4. Therefore, attention window size $\omega$ was set to 4.

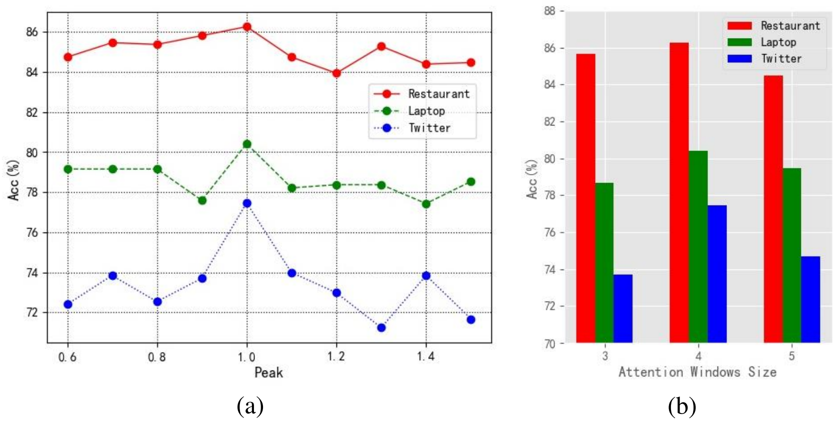


**Figure 5.** Effects of different parameters in CGAT: (**a**) peak $A$; (**b**) attention window size $\omega$.

4.3.4. Syntactic Weight Visualization

To observe the performance of the proposed syntactic attention mechanism on multi-aspect sentences, the syntactic weight were visualized by selecting representative sentences from the dataset *Restaurant*.

As shown in Figure 6a, for a sentence containing aspects with the same sentiment polarity, "*The food is great and the milkshakes are even better!*", the syntactic attention mechanism successfully discriminates opinion words with different aspects.

As shown in Figure 6b, for a sentence containing aspects with different sentiment polarities, "*The appetizers are ok , but the service is slow*", the syntactic attention mechanism still captured the opinion words of different aspects and avoided the interference of other words.

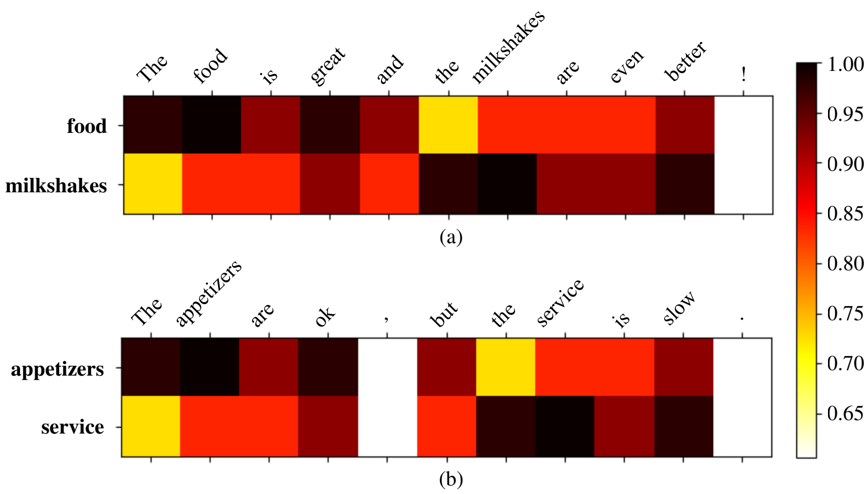


**Figure 6.** Visualization of syntactic weight: (**a**) a sentence containing aspects with the same sentiment polarity; (**b**) a sentence containing aspects with different sentiment polarities .

## 5. Conclusions

This paper has proposed a novel aspect-level sentiment classification model CGAT based on a contextual graph attention network. Rooted in target aspects to reconstruct dependency trees, the proposed model adopts two graph attention networks to mine syntactic structure information, and utilizes a context attention network to generate aspect-sensitive semantic features. In addition, a simple and efficient syntactic attention mechanism based on SRD was proposed, where the Gaussian function is cleverly introduced as the syntactic weight function.

Experimental results on three public sentiment datasets showed that CGAT outperforms other models and can effectively identify sentiment polarity expressed by aspects. The contextual graph attention network can effectively extract aspect-related semantic information and syntactic structure information. With a low computational cost, the syntactic attention mechanism effectively avoids the loss of sentiment information and emphasizes aspect-related words in syntax.

Future research will consider adopting image information to augment sentiment expressing of textual information.

**Author Contributions:** Conceptualization G.C.; investigation, L.Z.; methodology, T.L.; validation, W.Z.; validation, M.Z.; writing—original draft, Y.M. and R.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper was supported by the National Natural Science Foundation of China under grant 61866007, the Key Laboratory Project of Image and Graphics Intelligent Processing in Guangxi under grant GIIP201706, the Guangxi Natural Science Foundation Major Program under grant 2018GXNSFDA138006, the Guangxi Natural Science Fund Project under grant 2020GXNSFAA159094, the Guangdong Graduate Education Innovation Project under grant 2022JGXM115, the Basic Ability Enhancement Project for Young and Middle-aged Teachers of Universities in Guangxi under grant 2021KY0799, the Guangdong Science and Technology Planning Project under grant 2016A020210122, and the Innovation Project of GUET Graduate Education under grant 2022YCXS066.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Liu, B. *Sentiment Analysis and Opinion Mining*; Synthesis Lectures on Human Language Technologies; Morgan & Claypool Publishers: San Rafael, CA, USA, 2012.
2. Tang, D.; Qin, B.; Feng, X.; Liu, T. Effective LSTMs for Target-Dependent Sentiment Classification. In Proceedings of the COLING 2016, 26th International Conference on Computational Linguistics, Osaka, Japan, 11–16 December 2016; Calzolari, N., Matsumoto, Y., Prasad, R., Eds.; ACL: Stroudsburg, PA, USA, 2016; pp. 3298–3307.
3. Wang, Y.; Huang, M.; Zhu, X.; Zhao, L. Attention-based LSTM for Aspect-level Sentiment Classification. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, TX, USA, 1–4 November 2016; Su, J., Carreras, X., Duh, K., Eds.; The Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 606–615.
4. Zhang, C.; Li, Q.; Song, D. Aspect-based Sentiment Classification with Aspect-specific Graph Convolutional Networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, 3–7 November 2019; Inui, K., Jiang, J., Ng, V., Wan, X., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 4567–4577.
5. 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, ACL 2020, Online, 5–10 July 2020; Jurafsky, D., Chai, J., Schluter, N., Tetreault, J.R., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 3229–3238.
6. He, R.; Lee, W.S.; Ng, H.T.; Dahlmeier, D. Effective Attention Modeling for Aspect-Level Sentiment Classification. In Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, NM, USA, 20–26 August 2018; Bender, E.M., Derczynski, L., Isabelle, P., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 1121–1131.
7. Huang, B.; Ou, Y.; Carley, K.M. Aspect Level Sentiment Classification with Attention-over-Attention Neural Networks. In Proceedings of the Social, Cultural, and Behavioral Modeling—11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, 10–13 July 2018; Thomson, R., Dancy, C.L., Hyder, A., Bisgin, H., Eds.; Springer: New York, NY, USA, 2018; Volume 10899, pp. 197–206.

8. Chen, P.; Sun, Z.; Bing, L.; Yang, W. Recurrent Attention Network on Memory for Aspect Sentiment Analysis. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, 9–11 September 2017; Palmer, M., Hwa, R., Riedel, S., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 452–461.
9. Fan, F.; Feng, Y.; Zhao, D. Multi-grained Attention Network for Aspect-Level Sentiment Classification. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; Riloff, E., Chiang, D., Hockenmaier, J., Tsujii, J., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 3433–3442.
10. Park, H.; Song, M.; Shin, K. Deep learning models and datasets for aspect term sentiment classification: Implementing holistic recurrent attention on target-dependent memories. *Knowl. Based Syst.* **2020**, *187*, 104825–104839. [CrossRef]
11. Cheng, Y.; Yao, L.; Zhang, G.; Tang, T.; Xiang, G.; Chen, H. Text Sentiment Orientation Analysis of Multi-Channels CNN and BiGRU Based on Attention Mechanism. *J. Comput. Res. Dev.* **2020**, *57*, 2583–2595. (In Chinese)
12. Liu, Q.; Liang, B.; Xu, J.; Zhou, Q. A Deep Hierarchical Neural Network Model for Aspect-based Sentiment Analysis. *Chin. J. Comput.* **2018**, *41*, 2637–2652. (In Chinese)
13. Geeraerts, D. *Cognitive Linguistics: Basic Readings*; Walter de Gruyter: Berlin, Germany, 2006; Volume 34.
14. Carley, K. Coding choices for textual analysis: A comparison of content analysis and map analysis. *Sociol. Methodol.* **1993**, 23, 75–126. [CrossRef]
15. Dozat, T.; Manning, C.D. Deep Biaffine Attention for Neural Dependency Parsing. In Proceedings of the 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, 24–26 April 2017; pp. 1–8.
16. Stella, M. Text-mining forma mentis networks reconstruct public perception of the STEM gender gap in social media. *PeerJ Comput. Sci.* **2020**, *6*, e295. [CrossRef] [PubMed]
17. Teixeira, A.S.; Talaga, S.; Swanson, T.J.; Stella, M. Revealing semantic and emotional structure of suicide notes with cognitive network science. *Sci. Rep.* **2021**, *11*, 19423. [CrossRef] [PubMed]
18. Lu, Q.; Zhu, Z.; Zhang, G.; Kang, S.; Liu, P. Aspect-gated graph convolutional networks for aspect-based sentiment analysis. *Appl. Intell.* **2021**, *51*, 4408–4419. [CrossRef]
19. Du, J.; Zhang, Y.; Yue, B.; Chen, M. Bidirectional Edge-Enhanced Graph Convolutional Networks for Aspect-based Sentiment Classification. In Proceedings of the IEEE 45th Annual Computers, Software, and Applications Conference, COMPSAC 2021, Madrid, Spain, 12–16 July 2021; pp. 1599–1604.
20. Wang, G.; Li, H.; Qiu, Y.; Yu, B.; Liu, T. Aspect-based Sentiment Classification via Memory Graph Convolutional Network. *J. Chin. Inf. Process.* **2021**, *35*, 98–106. (In Chinese)
21. Tang, D.; Qin, B.; Liu, T. Aspect Level Sentiment Classification with Deep Memory Network. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, TX, USA, 1–4 November 2016; Su, J., Carreras, X., Duh, K., Eds.; The Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 214–224.
22. Chen, Y.; Zhuang, T.; Guo, K. Memory network with hierarchical multi-head attention for aspect-based sentiment analysis. *Appl. Intell.* **2021**, *51*, 4287–4304. [CrossRef]
23. Su, J.; Ouyang, Z.; Yu, S. Aspect-Level Sentiment Classification for Sentences Based on Dependency Tree and Distance Attention. *J. Comput. Res. Dev.* **2019**, *56*, 1731–1745. (In Chinese)
24. Devlin, J.; Chang, M.; 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, NAACL-HLT 2019, Minneapolis, MN, USA, 2–7 June 2019; Burstein, J., Doran, C., Solorio, T., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; Volume 1, pp. 4171–4186.
25. Xing, B.; Tsang, I.W. DigNet: Digging Clues from Local-Global Interactive Graph for Aspect-level Sentiment Classification. *arXiv* **2022**, arXiv:2201.00989.
26. Howard, A.; Pang, R.; Adam, H.; Le, Q.V.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.; Tan, M.; Chu, G.; et al. Searching for MobileNetV3. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea, 27 October–2 November 2019; pp. 1314–1324.
27. 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@COLING 2014, Dublin, Ireland, 23–24 August 2014; Nakov, P., Zesch, T., Eds.; The Association for Computer Linguistics: Stroudsburg, PA, USA, 2014; pp. 27–35.
28. Dong, L.; Wei, F.; Tan, C.; Tang, D.; Zhou, M.; Xu, K. Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, Baltimore, MD, USA, 22–27 June 2014; Short Papers; The Association for Computer Linguistics: Stroudsburg, PA, USA, 2014; Volume 2, pp. 49–54.