

Article

Triplet Contrastive Learning for Aspect Level Sentiment Classification

Haoliang Xiong ^{1,†} , Zehao Yan ^{1,†}, Hongya Zhao ², Zhenhua Huang ³ and Yun Xue ^{1,*}¹ School of Electronics and Information Engineering, South China Normal University, Foshan 528225, China² Industrial Center, Shenzhen Polytechnic, Shenzhen 518055, China³ School of Computer Science, South China Normal University, Guangzhou 510631, China

* Correspondence: xueyun@m.scnu.edu.cn

† These authors contributed equally to this work.

Abstract: The domain of Aspect Level Sentiment Classification, in which the sentiment toward a given aspect is analyzed, attracts much attention in NLP. Recently, the state-of-the-art Aspect Level Sentiment Classification methods are devised by using the Graph Convolutional Networks to deal with both the semantics and the syntax of the sentence. Generally, the parsing of syntactic structure inevitably incorporates irrelevant information toward the aspect. Besides, the syntactic and semantic alignment and uniformity that contribute to the sentiment delivery is currently neglected during processing. In this work, a **Triplet Contrastive Learning Network** is developed to coordinate the syntactic information and the semantic information. To start with, the aspect-oriented sub-tree is constructed to replace the syntactic adjacency matrix. Further, a sentence-level contrastive learning scheme is proposed to highlight the features of sentiment words. Based on The Triple Contrastive Learning, the syntactic information and the semantic information are thoroughly interacted and coordinated whilst the global semantics and syntax can be exploited. Extensive experiments are performed on three benchmark datasets and achieve accuracies (BERT-based) of 87.40, 82.80, 77.55 on Rest14, Lap14, and Twitter datasets, which demonstrate that our approach achieves state-of-the-art results in Aspect Level Sentiment Classification task.

Keywords: Aspect Level Sentiment Classification; Contrastive Learning; Graph Convolutional Networks

MSC: 18C50



Citation: Xiong, H.; Yan, Z.; Zhao, H.; Huang, Z.; Xue, Y. Triplet Contrastive Learning for Aspect Level Sentiment Classification. *Mathematics* **2022**, *10*, 4099. <https://doi.org/10.3390/math10214099>

Academic Editors: Jianping Gou, Weihua Ou, Shaoning Zeng and Lan Du

Received: 7 October 2022

Accepted: 1 November 2022

Published: 3 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Aspect Level Sentiment Classification (ALSC) is a fundamental subtask of fine-grained sentiment analysis, which currently receives a great deal of attention [1]. The main focus of ALSC is to identify the sentiment polarity (e.g., positive, neutral or negative) of aspects explicitly given in sentences. For example, in the sentence “*The price is reasonable although the service is poor*” (Figure 1), the sentiment toward aspects *price* and *service* is positive and negative, respectively.

Advances of deep neural networks bring paradigm shift to various tasks of NLP and the ALSC is no different [2–4]. The attention-based network is a most common approach that exploits the semantic information to capture the sentiment words of the given aspect. In Figure 1, more attentive weights can be assigned to the sentiment words *reasonable* and *poor* via attention mechanism. However, the use of semantic feature alone can result in the misunderstanding of contextual words, especially for sentences of complex syntax structure. More recently, the application of Graph Convolutional Networks (GCN) in ALSC is both creative and practical [5]. For one thing, the encoding of syntactic information using GCN mitigates the deficiencies of long-distance dependencies among words [6,7]. For another, not just the syntax, but also the semantic information can be processed by

GCN, which gives rise to opportunities to the integration of semantic features. As such, the state-of-the-art approaches work on developing multi-channel GCNs to deal with multiple information [8,9].

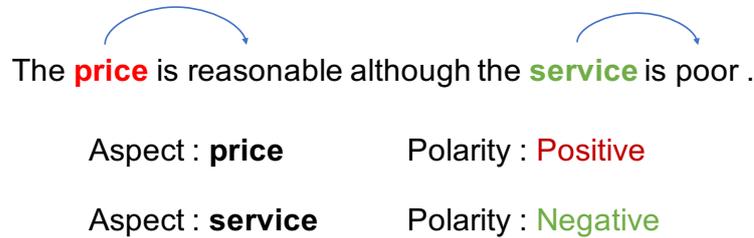


Figure 1. An example of ALSC.

Despite the progress of GCN-based method in ALSC, two main limitations are observed. **On the one hand**, most syntactic parsing is performed on the whole sentence without considering the importance of key phrases (e.g., aspect words, opinion words and etc.) to sentiment determination. In such a manner, redundant information or even noise can be incorporated during feature extraction. **On the other hand**, current methods set the semantic information and syntactic information in two individual spaces for feature extraction and fuse their features in a elementary way. But the alignment and uniformity of these two categories of features are ignored [10].

Inspired by the methods reported by [8,11], a **Triplet Contrastive Learning Network (TCL)** for ALSC is proposed to address the aforementioned issues. For the exploiting of syntactic information, we start with reconstructing the syntax dependency tree by setting the aspect as the root according to [12] (Figure 2). The dependencies between aspect word and other words are explicitly established, which contributes to the capturing of opinions words to the aspect and restricting the introduction of redundant information. As presented in [13], the key phrase plays a pivot role in delivering the essence of texts. To further filter the noise and highlight the key information, a contrastive learning scheme is proposed to magnify the significance of sentiment-related words. In ALSC tasks, the key phrases are either nouns, verbs, adjectives, or adverbs of degree [14]. With the application of masking mechanism, both positive and negative examples are generated and fed into the contrastive learning module to enhance the impacts of key phrases and distill the syntactic features.

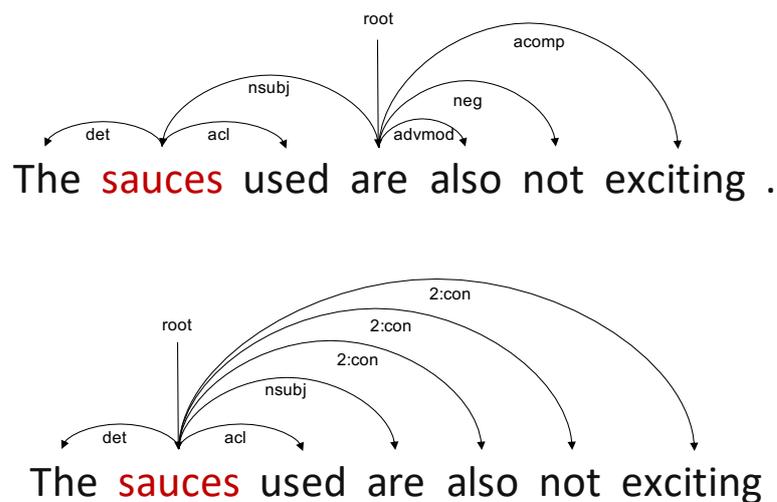


Figure 2. Reconstruction of aspect-oriented syntax dependency tree.

With respect to the integration of sentence syntax and semantics, recent publications reveal that they are distinct and related [8,15]. Likewise, the features from both space, conveying sentiment toward the aspect, also have a similar relationship between each other.

For this reason, the alignment of both features can facilitate the information integration. Concretely, the features, within either syntactic or semantic space, expressing the same sentiment polarity can be aligned while those expressing different sentiment polarities can be separated. With this, the interaction between syntactic information and semantic information is carried out, based on which a dual-contrastive learning scheme is devised. For each data within the mini-batch, the features of the same sentiment polarity are getting closer based on dual-contrastive learning, and vice versa. In this way, features of both categories are thoroughly interacted and aligned. We can thus leverage feature integration to improve the ALSC performance.

The contributions of this paper are as follows:

- The syntactic adjacency matrix of the dual-channel graph convolutional neural network is replaced with an aspect-oriented tree structure, which helps the model to better capture the information of opinion words related to aspect words.
- A syntactic contrastive learning scheme is designed to encourage the model to focus on keywords that are helpful for sentiment polarity classification, and to better learn features related to aspect words.
- Constructing the dual contrastive learning module can make the semantic features and syntactic features of sentences more fully interact and align.
- Experiments show that our method outperforms baseline models on three benchmark datasets.

This work is organized as follows. Section 2 gives an overview of relevant work of ALSC and contrastive learning. Section 3 describes the TCL model in details. In Section 4, the experiment is depicted, as well as the presentation of result analysis. Concluding remarks are given in Section 5.

2. Related Work

2.1. Aspect Level Sentiment Classification

Sentiment classification tasks mainly focus on capturing the sentiment information from the given text [16,17]. ALSC aims to classify the sentiment polarity of a specific aspect from given texts. Within ALSC, a more detailed analysis about the sentiment associated with the aspect is performed by using the textual information. Early research focuses on employing CNN- and RNN-based method, together with the integration of attention mechanisms or knowledge distillation [18,19], to obtain aspect-related information. As such, the utilization of attention mechanism to precisely capture the aspect-aware contextual information becomes a main topic [2,3]. In recent years, GCN-based models rise to prominence in a variety of NLP tasks, which is capable of alleviating the defects of attention networks. On the task of ALSC, Ref. [6] first apply GCN to tackle the syntax dependency and resolve the long-term multi-word dependencies. Later work aims to establish the syntax structure and extract aspect-related features [7]. Ref. [12] re-shape the syntax dependency tree into an aspect-oriented sub-tree, in order to determine the connection between aspects and its opinion words. Ref. [20], fuse the syntax dependency types into GCN, based on which to highlight the syntax corresponds to sentiment classification. So far, there is an ongoing trend to combine the sentence syntax and semantics [8,9,21]. Most approaches tend to separately construct adjacency matrix for syntactic and semantic information, generate corresponding feature representations, and concatenate the representations for sentiment classification.

2.2. Contrastive Learning

A fundamental focus of contrastive learning is the learning of alignment and uniformity of given data [10]. Comprehensively, alignment is taken to indicate the similarity among positive examples while uniformity refers to informative-distribution of features, so that negative examples are isolated from positive ones. In practical use, both alignment and the uniformity are used as indexes to optimize the feature learning. That is, the cap-

turing intra-class similarities and inter-class differences can benefit the performance in downstream tasks.

Recently, a number of studies apply contrastive learning to NLP tasks and achieve satisfying results [22–24]. Ref. [22] devise a simple contrastive sentence embedding framework, which can produce superior sentence embeddings on semantic textual similarity tasks. For the aspect words absent from the training set, Ref. [25] take contrastive learning to capture aspect-invariant and aspect-dependent features to distinguish the roles of valuable sentiment features. Ref. [11] propose a novel contrastive-learning-based approach that simultaneously learns the features of input samples and the parameters of classifiers in the same space on the task of text classification.

3. Proposed Method

Figure 3 shows the framework of the TCL Network. Let $X = \{x_1, x_2, \dots, x_a, \dots, x_{a+l_a}, \dots, x_N\}$ be a N -word sentence with aspect $A = \{x_a, \dots, x_{a+l_a}\}$ in it where a represents the starting index of A and l_a is the length of A . We feed the sentence into GloVe [26] or BERT [27] encoder for sentence embedding establishment. For GloVe-based model, each word is mapped into a low-dimensional vector by looking up in a pretrained word embedding matrix $E \in \mathbb{R}^{|V| \times d_E}$ where $|V|$ is the lexicon size and d_E is the dimension of word vector. The sentence embedding is given as $x = \{e_1, e_2, \dots, e_N\}$. The hidden states of the sentence are extracted via Bi-LSTM. The contextual feature vector is $H = \{h_1, h_2, \dots, h_N\}$ with $H \in \mathbb{R}^{N \times 2d}$ and d representing the hidden layer dimension. In addition, the sequence $[CLS]X[SEP]A[SEP]$ can also sent to BERT encoder to obtain the contextual feature vector H . Subsequently, H is taken as the input of both semantic-learning GCN module and syntactic-aware GCN module. A multi-layer Biaffine unit is proposed to integrate the semantic features and syntactic features. To further align the features from both space, the dual contrastive learning scheme is carried out. More details of each component are presented as follows.

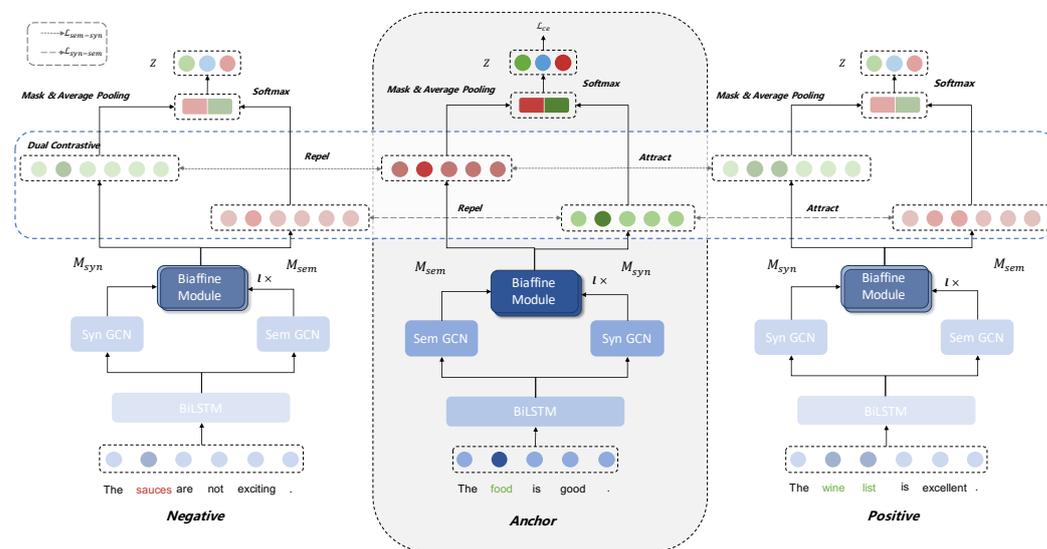


Figure 3. The overall architecture of our Triplet Contrastive Learning Network.

3.1. Syntactic-Aware GCN Module

The architecture of syntactic-aware module is exhibited in Figure 4. As pointed out in the Introduction, the syntactic-aware GCN in our model tends to precisely capture the aspect-related context words and remove the redundant information. According to [12], a relational graph attention network is devised. Specifically, we construct an aspect-oriented dependency tree to replace the adjacency matrix of classical syntax dependency tree. Then, the attention mechanism is applied to the reshape sub-tree to capture the aspect-specific contextual features. Moreover, to resolve the long-dependencies among words, we set four

categories of words as the key phrases that contributes to sentiment delivery, i.e., nouns, verbs, adjectives, and adverbs of degree. As such, the contrastive learning is performed to enhance the features of key phrases and effectively capture the word feature of long dependency.

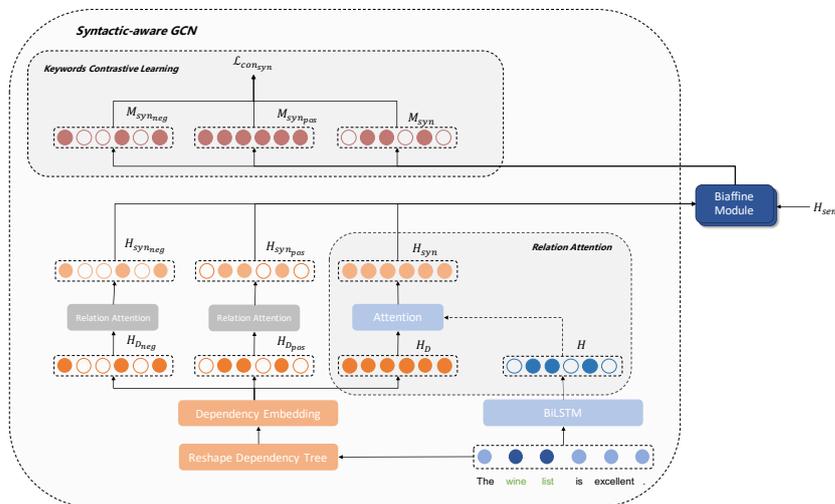


Figure 4. Architecture of syntactic-aware GCN module

3.1.1. Relational Graph Attention Module

At this stage, the aspect A is taken as the central word to construct the aspect-oriented dependency tree; see Algorithm 1 For words syntactically related to the central word of one hop, the corresponding dependency types are established. Through iteration, for words syntactically related to the central word of n hops ($n \geq 2$), the dependency types are characterized by $(con : n)$. If the aspect contains multiple words, these words are considered as a whole. In such a manner, we shall thus obtain the re-constructed dependency tree as $D = \{dep_1, dep_1, \dots, dep_N\}$ and map it into embedding space to generate the dependency representation $H_D = \{h_{D_1}, h_{D_2}, \dots, h_{D_N}\}$. Notably, the randomly initialized dependency embedding $E_D \in \mathbb{R}^{|V_d| \times d_D}$ is employed with $|V_d|$ standing for the number of dependency types. For $H_D \in \mathbb{R}^{N \times d_D}$, we have d_D representing the dimension of dependency type embeddings.

The relational attention between aspect and the dependency type representation is computed. Specifically, the syntactic dependency of context toward the aspect is incorporated within H_D . Thus, the attentive weight between H_D and H is calculated using a simplified inner product operation, which is:

$$att = f \left(\frac{(W_D H_D + b_D) \times (W_h H + b_h)^T}{\sqrt{d_m}} \right) \tag{1}$$

where $W_D \in \mathbb{R}^{d_D \times d_m}$ and $W_h \in \mathbb{R}^{2d \times d_m}$ are linear layer weights; b_D and b_h are bias terms; $f(\cdot)$ stands for the softmax activation function; and d_m is the hidden layer dimension of the attention module.

Then, the syntactic representation is given as:

$$H_{syn} = att * H \tag{2}$$

Algorithm 1 Aspect-Oriented Dependency Tree

Input: sentence $X = \{x_1, x_2, \dots, x_N\}$, aspect $A = \{x_a, \dots, x_{a+l_a}\}$, dependency tree T , and dependency relations R .

Output: aspect-oriented dependency \tilde{T} .

- 1: Construct the aspect root \tilde{R} for \tilde{T}
 - 2: **for** a to $a + l_a$ **do**
 - 3: **for** $j = 1$ to n **do**
 - 4: **if** $x_j \notin A$ and $x_j \xrightarrow{R_{ja}} x_a$ **then**
 - 5: $x_j \xrightarrow{R_{ja}} \tilde{R}$
 - 6: **else if** $x_j \notin A$ and $x_j \xleftarrow{R_{ja}} x_a$ **then**
 - 7: $x_j \xleftarrow{R_{ja}} \tilde{R}$
 - 8: **else**
 - 9: $n = \text{distance}(a, j)$
 - 10: $x_j \xrightarrow{n:con} \tilde{R}$
 - 11: **end if**
 - 12: **end for**
 - 13: **end for**
 - 14: **return** \tilde{T}
-

3.1.2. Syntactic Contrastive Learning Scheme

The effectiveness of key phrases (i.e., nouns, verbs, adjectives or adverbs of degree) is highlighted by using based on a sentence-level key phrases contrastive learning module. To be specific, a mask operation, based on the POS information of phrases in the sentence, is performed. Only if the position mask 1 assigned to key phrase and mask 0 to others, can this representation defined as a positive example, i.e., $M_{pos} \in \mathbb{R}^N$. Conversely, a negative example indicates a key phrase with a position mask 0 while other words with a mask 1, i.e., $M_{neg} \in \mathbb{R}^N$.

The dependency type can be integrated into both positive and negative examples. We shall thus compute the positive example dependency type representation and the positive example dependency type representation as:

$$H_{D_{pos}} = H_D * M_{pos} \tag{3}$$

$$H_{D_{neg}} = H_D * M_{neg} \tag{4}$$

Similar to Equation (1), the attention weights of $H_{D_{pos}}$ and $H_{D_{neg}}$ toward the context representation are available, as presented in Equations (5) and (6). Thus, the syntactic representation of both positive examples and negative examples can be obtained (Equations (7) and (8)):

$$att_{pos} = f \left(\frac{(W_{D_{pos}} H_{D_{pos}} + b_{D_{pos}}) \times (W_{h_{pos}} H + b_{h_{pos}})^T}{\sqrt{d_m}} \right) \tag{5}$$

$$att_{neg} = f \left(\frac{(W_{D_{neg}} H_{D_{neg}} + b_{D_{neg}}) \times (W_{h_{neg}} H + b_{h_{neg}})^T}{\sqrt{d_m}} \right) \tag{6}$$

$$H_{syn_{pos}} = att_{pos} * H \tag{7}$$

$$H_{syn_{neg}} = att_{neg} * H \tag{8}$$

For every sentence, we have its syntactic representation H_{syn} , the syntactic representation with key phrases $H_{syn_{pos}}$ and syntactic representation without key phrases $H_{syn_{neg}}$. Each of these syntactic representations is fed into a shared-weight biaffine unit to fuse with

the semantic representation in following section. The final syntactic representations, with the integration of semantic information, are presented as M_{syn} (derived from Equation (13)), $M_{syn_{pos}}$ and $M_{syn_{neg}}$, respectively.

Aiming to focus more on the key phrases, the contrastive learning scheme is carried out with the loss function set as:

$$\mathcal{L}_{con_{syn}} = -\frac{1}{B} \sum_{j=1}^B \frac{1}{N} \sum_{i=1}^N \log \frac{e^{\text{sim}(M_{syn_{pos}}^i, M_{syn}^i) / \tau_1}}{\sum_{t=1}^N (e^{\text{sim}(M_{syn_{pos}}^i, M_{syn}^i) / \tau_1} + e^{\text{sim}(M_{syn_{neg}}^t, M_{syn}^i) / \tau_1})} \tag{9}$$

where τ_1 is the temperature coefficient, B is the batch size and N is sentence length mentioned above.

Distinguishing from the current contrastive learning approaches, in addition to the positive example $M_{syn_{pos}}^i$, the other examples, containing $N - 1$ key-phrases-related syntactic representations $M_{syn_{pos}}^t (t \neq i)$ and N syntactic representations without key phrases $M_{syn_{neg}}$, are all considered as negative examples. In other words, the negative examples of each word in the sentence is $2N - 1$.

3.2. Semantic-Learning GCN Module

The sentence semantics is also encoded via GCN to enhance the modelling of sentiment information. Seeing that the self-attention mechanism is capable of extracting the semantic relevance of other words and the given word, we use self-attention network to construct a semantic adjacency matrix $A^{sem} \in \mathbb{R}^{N \times N}$:

$$A^{sem} = f \left(\frac{QW^q \times (KW^k)^T}{\sqrt{d}} \right) \tag{10}$$

where both Q and K equal the context representation H , W^q and W^k are trainable weighting parameters and d is the hidden layer size of attention network.

The semantic representation is derived from graph convolution, which is:

$$H_{sem} = \sigma(A^{sem}WH + b) \tag{11}$$

where $\sigma(\cdot)$ stands for the linear activation function, such as ReLU function.

3.3. Biaffine Unit

The interaction of semantic information and syntactic information is conducted via multi-layer mutual Biaffine transformation. In Equation (12), H_{syn} and H_{sem} are first multiplied to obtain a syntactic-related matrix containing the semantic information. Then, the syntactic-related matrix is mapped via Softmax and multiplied by the original semantic information to obtain the final syntactic feature representation with semantic information integrated. Via multi-layers of Biaffine unit, the semantic features can be fused to the syntactic representation for sentiment polarity classification. So is Equation (13).

$$H_{syn}^{(l)} = f \left(H_{syn}^{(l-1)} W_1^{(l-1)} \left(H_{sem}^{(l-1)} \right)^T \right) H_{sem}^{(l-1)} \tag{12}$$

$$H_{sem}^{(l)} = f \left(H_{sem}^{(l-1)} W_2^{(l-1)} \left(H_{syn}^{(l-1)} \right)^T \right) H_{syn}^{(l-1)} \tag{13}$$

where $l (l = 1, 2, \dots)$ stands for the layer number of the biaffine unit; both $W_1 \in \mathbb{R}^{2d \times 2d}$ and $W_2 \in \mathbb{R}^{2d \times 2d}$ are learnable parameters. Specifically, we take $H_{sem}^{(0)}$ and $H_{syn}^{(0)}$ to represent $H_{sem} \in \mathbb{R}^{N \times 2d}$ and $H_{syn} \in \mathbb{R}^{N \times 2d}$, which are the inputs of biaffine unit.

With the mutual Biaffine transformation, we thus obtain the final semantic representation with fused syntactic features $H_{sem}^{(l)}$ which also presented as M_{sem} and the final

syntactic representation with fused semantic features $H_{syn}^{(l)}$ which also presented as M_{syn} . The average pooling is performed on the outcomes in relation to the aspect.

$$M_{sem}^A = \text{avgpool}(M_{sem_a}, \dots, M_{sem_{a+la}}) \tag{14}$$

$$M_{syn}^A = \text{avgpool}(M_{syn_a}, \dots, M_{syn_{a+la}}) \tag{15}$$

Then, both the semantic representation and the syntactic representation of the aspect are concatenated and sent to the linear classifier to determine the sentiment polarity of the given aspect:

$$Z = f\left(W \left[M_{sem}^A; M_{syn}^A \right] + b\right) \tag{16}$$

where $[\cdot]$ stands for the vector concatenation, W and b are learnable parameters in the linear layer.

3.4. Dual Contrastive Learning Scheme

In the proposed model, the main purpose of the dual contrastive learning is to comprehensively align the features of both syntactic space and semantic space. The global syntactic features and semantic features can thus be captured. Notably, the output of the biaffine unit (i.e., M_{syn} and M_{sem}) are taken as the input of the dual contrastive learning module. For each input X_i , the sequence with the same sentiment polarity within the same batch is considered as the positive example \mathcal{P} , otherwise as negative example \mathcal{N} . The loss function of the dual contrastive learning is presented as:

$$\mathcal{L}_{syn-sem} = -\frac{1}{B} \sum_{i=1}^B \frac{1}{|\mathcal{P}|} \sum_{j \in \mathcal{P}} \log \frac{e^{\text{sim}(M_{syn_i}, M_{sem_j}) / \tau_2}}{\sum_{t=1}^B e^{\text{sim}(M_{syn_i}, M_{sem_t}) / \tau_2}} \tag{17}$$

$$\mathcal{L}_{sem-syn} = -\frac{1}{B} \sum_{i=1}^B \frac{1}{|\mathcal{P}|} \sum_{j \in \mathcal{P}} \log \frac{e^{\text{sim}(M_{sem_i}, M_{syn_j}) / \tau_3}}{\sum_{t=1}^B e^{\text{sim}(M_{sem_i}, M_{syn_t}) / \tau_3}} \tag{18}$$

where τ_2 and τ_3 are the temperature coefficients of contrastive loss.

3.5. Loss Function

The loss function for model training is expressed:

$$\begin{aligned} \mathcal{L} = & \mathcal{L}_{CE} + \alpha \mathcal{L}_o + \beta (\mathcal{L}_{syn-sem} + \mathcal{L}_{sem-syn}) \\ & + \gamma \mathcal{L}_{con_{syn}} + \lambda \|\Theta\| \end{aligned} \tag{19}$$

with

$$\mathcal{L}_o = \|A^{sem} A^{sem T} - I\|_F \tag{20}$$

where α , β and γ are hyperparameters; \mathcal{L}_{CE} represents the cross-entropy loss for sentiment polarity classification; Θ denotes the training parameter set; λ represents the coefficient of L2 regularization. Inspired by [8], for each word in the sentence, its attention distribution on every other word is distinguishing. In other words, the overlap of attentive weights has to be minimized especially for the application of semantic graph adjacency matrix. Therefore, an additional orthogonal regularized loss function L_o is thereby introduced. The parameter I in Equation (20) is an identity matrix and the subscript F stands for the Frobenius norm.

Since the contrastive learning loss results are derived from various weighting parameters, the back propagation can be applied to optimize these parameters during the loss function optimizing.

4. Experiments

4.1. Datasets and Settings

Datasets: We evaluate the working performance of TCL network on three benchmark datasets, which are Rest14 and Lap14 from SemEval 2014 Task4 [28] and Twitter [29]. Each sample in these datasets is either a product review or tweet sentence, which contains explicit aspect words and the corresponding sentiment polarities. Each aspect from the product reviews or tweets in our experiments is labeled as positive, neutral or negative. Details of each dataset are exhibited in Table 1.

Experimental Settings: For GloVe-based model, we initialize the word embeddings with 300-dimensional vectors pretrained by GloVe [26]. The dimension of the dependent syntactic embeddings is set to 30. The hidden layer dimension of BiLSTM is 50. All the weights in the model are initialized by Xavier uniform distribution. The layer number of biaffine unit is set as 2. For the contrastive learning scheme, the temperature coefficient determines how much attention the contrastive learning loss assign to the outlier negative samples. The larger the temperature coefficient is, the greater the tolerance to negative samples, and vice versa. In the syntactic contrastive learning module, it is desirable that more attention is given to key phrases with a certain tolerance to other words. Therefore, τ_1 of the syntactic contrastive learning is 1 while τ_2 and τ_3 of the dual contrastive learning is set to 0.1. In addition, the Adam optimizer is adopted with a learning rate of 2×10^{-3} . The batch size ranges from 16 to 64. The L2 regularization coefficient λ is set to 1×10^{-4} . Notably, the values of α , β and γ vary in line with the datasets, which are 0.1, 0.5 and 0.5 for Rest14, 0.5, 0.7 and 0.8 for Lap14 and 0.2, 0.2 and 0.7 for Twitter.

Table 1. Statistics of datasets.

Dataset		#Pos.	#Neu.	#Neg.	Total
Rest14	Train	2164	637	807	3608
	Test	728	196	196	1120
Lap14	Train	994	464	870	2328
	Test	341	169	128	638
Twitter	Train	1561	3127	1560	6248
	Test	173	346	173	692

4.2. Baselines

In order to validate the effectiveness of the proposed model in ALSC, we take 10 state-of-the-art methods for comparison:

1. **ASGCN [6]** The syntactical features are obtained using GCN via syntax dependency tree while the aspect-specific attention is applied to extract the features related to aspects.
2. **CDT [30]** The Bi-LSTM is taken to learn the sentence representations and the GCN encodes the syntactic information and capture the aspect-related syntactic features.
3. **RGAT [12]** The aspect-oriented dependency tree is constructed, based on which the relation graph attention network is developed to learn the dependencies between aspect and other words.
4. **BiGCN [31]** A global lexical graph and a concept hierarchy graph are constructed, which aims to integrate word pair co-occurrence and syntactic dependencies.
5. **DualGCN [8]** A dual-channel GCN method is proposed to extract both syntactic and semantic information, and then fuse the two categories of information.
6. **BERT-SPC [27]** The sentence-aspect pair is sent to BERT model with its token [CLS] used for sentiment classification.
7. **T-GCN [20]** A multilayer type-aware GCN is established to learn the relationship among words.
8. **BERT4GCN [32]** The intermediate layers of BERT is employed to augment GCN for ALSC.

9. **DR-BERT [33]** The Dynamic Re-weighting Adapter is proposed to encourage model to better understand aspect-aware sentiment through

4.3. Experimental Results and Analysis

We take two metrics, accuracy and Macro-F1, to evaluate the working performance of the proposed model. The experimental results of 13 different methods are presented in Table 2. Comparing with the state-of-the-arts, the TCL network is the best performing method in most datasets. There is a considerable performance gap between the proposed model and the baselines. According to Table 2, one can easily see that models using BERT-based embeddings have a better performance than those of GloVe-based embeddings. Indeed, the employment of GCN substantially contributes to the encoding of sentence syntax and semantics. With respect to our model, the effectively use of syntactic information highlights the contextual words related to the aspect. As a result, more attentive weights are given to words that contribute to the sentiment delivery. In comparison with the single-channel GCN (i.e., [6,12,30]), the dual-channel GCN methods (i.e., [8]), which deal with both the syntactic information and the semantic information, shows their superiorities in ALSC tasks. In this way, our model not just integrates different types of features, but also exploits the global information to further optimize the sentiment classification results.

Table 2. Experimental results. Bold numbers represent the best results among methods of the same type.

Models	Rest14		Lap14		Twitter	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
ASGCN [6]	80.77	72.02	75.55	71.05	72.15	70.40
CDT [30]	82.30	74.02	77.19	72.99	74.66	73.66
RGAT [12]	83.30	76.08	77.42	73.76	75.57	73.82
BiGCN [31]	81.97	73.48	74.59	71.84	74.16	73.35
DualGCN [8]	84.27	78.08	78.48	74.74	75.92	74.29
Our TCL	84.27	77.04	79.27	76.05	76.81	75.53
BERT-SPCBERT-SPC [27]	86.15	80.29	81.01	76.69	75.18	74.01
RGAT+BERT [12]	86.60	81.35	78.21	74.07	76.15	74.88
T-GCN [20]	86.16	77.11	77.49	73.01	74.73	73.76
DualGCN+BERT [8]	87.13	81.16	81.80	78.10	77.40	76.02
BERT4GCN [32]	84.75	77.11	77.49	73.01	74.73	73.36
DR-BERT [33]	87.72	82.31	81.45	78.16	77.24	76.10
Our TCL+BERT	87.40	82.12	81.80	78.96	77.55	76.57

However, the TCL network fails to overperform DR-BERT on the Rest14. A possible explanation is that the samples of distinguishing sentiment occupy significantly different proportion in the Rest14 dataset, which affects performance of contrastive learning scheme as the generation of positive and negative samples is obtained by random sampling.

4.4. Ablation Study

An ablation study is carried out on three datasets to investigate the importance of the contrastive learning losses; see Table 3. The dual contrastive learning scheme concerns the syntactic-based semantic learning loss function $\mathcal{L}_{sem-syn}$ and the semantic-based syntactic learning loss function $\mathcal{L}_{syn-sem}$. The results show that the ablating of both loss functions leads to the most significant drop. The main reason is that the employment of global features within the minibatch does benefit the sentiment delivery. We see that the contribution of $\mathcal{L}_{sem-syn}$ is slightly higher than that of $\mathcal{L}_{syn-sem}$, which indicates the effectiveness of semantic alignment. By contrast, the contribution of \mathcal{L}_{consyn} in the syntactic learning module is relatively small, but its removal still results in an average decrease of 1.2% in accuracy.

Table 3. Ablation study results. Bold numbers represent the best results.

Models	Rest14		Lap14		Twitter	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
TCL <i>w/o</i> $\mathcal{L}_{syn-sem}$	82.31	74.14	77.69	74.11	75.18	73.59
TCL <i>w/o</i> $\mathcal{L}_{sem-syn}$	82.30	74.73	78.01	74.72	75.33	74.01
TCL <i>w/o</i> $\mathcal{L}_{syn-sem}$ & $\mathcal{L}_{sem-syn}$	81.94	74.17	77.53	74.57	74.00	72.76
TCL <i>w/o</i> $\mathcal{L}_{con_{syn}}$	83.02	74.96	78.32	74.75	75.48	74.27
TCL	84.27	77.04	79.27	76.05	76.81	75.53

4.5. Case Study

Four examples of ALSG tasks are conducted and presented in Figure 5. The aspect words in green, blue, and red represent the positive, neutral, and negative sentiment polarities, respectively. The first case is a sentence of simple syntax and semantics. All the three models are capable of identifying the sentiment as negative. Sentence 2 contains multiple aspects. The ASGCN fails to determine the sentiment of aspect ‘disc drive’, because ‘disc drive’ is syntactically close to the word negative word ‘not’. Similarly, in sentence 3, the aspect ‘apple OS’ has a long distance dependency with its opinion word, which results in the misunderstanding of the sentiment using ASGCN. By contrast, the DualGCN, which integrates both syntactic and semantic information, can classify the sentiment toward aspect ‘apple OS’ correctly. In the last sentence, despite the complexity in both the syntax and semantics, the TCL network is capable of identifying the sentiment polarities of all aspects. The application of triplet contrastive learning effectively obtains alignment between semantic and syntactic features, indicating its efficacy in ALSG of complex sentences.

Sentence	ASGCN	DualGCN	Our TCL
but the mountain lion is just too slow .	(✓)	(✓)	(✓)
the latest version does not have a disc drive .	(✓, X)	(✓, ✓)	(✓, ✓)
Works well, and I am extremely happy to be back to an apple OS .	(✓, X)	(✓, ✓)	(✓, ✓)
the power plug has to be connected to the power adaptor to charge the battery but won't stay connected .	(✓, X, X)	(✓, X, X)	(✓, ✓, ✓)

Figure 5. Case study. ALSG results of TCL, ASGCN and DualGCN on testing examples, along with their predictions and correspondingly, golden labels. The marker ✓ and X indicate the correct classification and incorrect classification, respectively

4.6. Visualization

4.6.1. Comparison of Syntactic and Semantic Vectors

The distribution of semantic and syntactic representations aims to verify the effectiveness of the dual contrastive learning scheme. Figure 6 shows the visualization of semantic and syntactic outputs of the dual contrastive learning module using t-SNE algorithm [34]. To facilitate the comparison, we only take the data with positive and negative sentiment polarities for visualization. Apparently, both the basic TCL network and TCL without dual contrastive learning can distinguish one type of representations. Notably, the proposed model without dual contrastive learning fails to resolve the two types of vectors with the same sentiment polarity, such as the distribution of red dots, which indicates the importance of alignment between the semantic and syntactic spaces. Moreover, there are large amount of overlapping for vectors with different sentiment polarities. The uniformity of syntax and semantics is absent. In comparison, the TCL network considers both the alignment and the uniformity of features. With the application of dual contrastive learning scheme, not only the distribution of the same-polarity-representations is more concentrated, but also the overlapping within different-polarities-representations are reduced to a large extent.

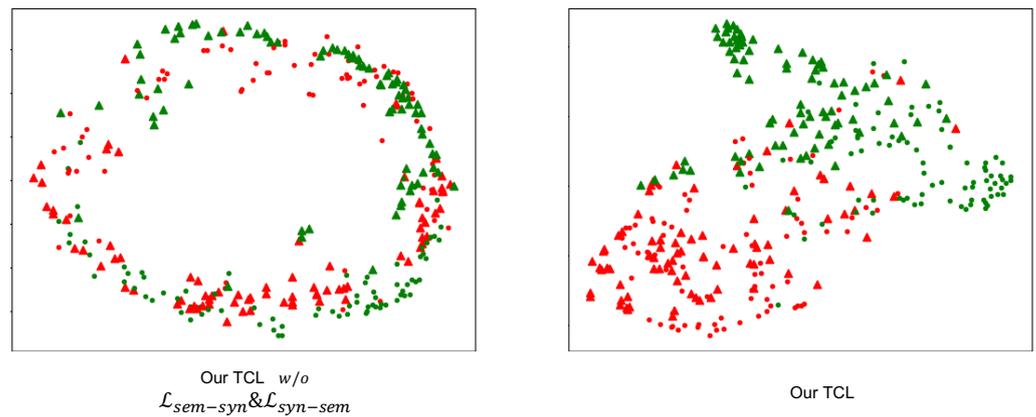


Figure 6. Visualization of semantic and syntactic vectors. Triangle dots represent syntactic vectors; round dots represent semantic vectors; dots in red represent positive samples; dots in green represent negative samples.

4.6.2. Sentiment Classification Visualization

Similarly, the visualization of triplet contrastive learning is also performed; see Figure 7. For the ASGCN that merely exploits the syntactic features, the neural samples can be distinguished from those of other two sentiment polarities. Whereas, the classification between positive and negative samples is challenging, with large amount of misunderstanding of the sentiment. Since DualGCN tackles both syntactic and semantic information, the samples of three sentiment polarities can be better discriminated. The distribution of neural samples is still not that distinctive, especially comparing with the negative samples. By contrast, our model shows its dominance in sentiment classification. It is clearly that a more concentrated distribution of samples with the same sentiment is accessible. Due to the introduction of triple contrastive learning, a better performance of feature learning and sentiment classification can be expected.

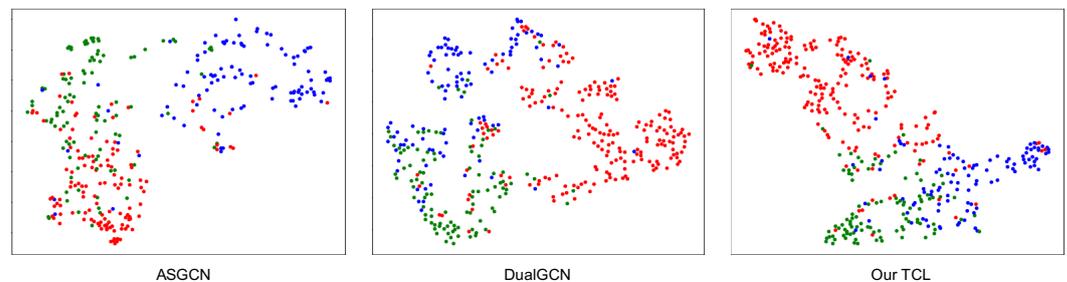


Figure 7. Visualization of sentiment classification results. The dots in green, red and blue respectively represent the positive, neutral and negative samples.

5. Conclusions

In this work, a TCL network is developed to deal with the ALSC tasks, which not just exploits the global information, but also obtains the alignment of semantics and syntax. To start with, an aspect-oriented dependency tree is constructed by reshaping the syntactic adjacency matrix. Then, the sentence-level contrastive learning is applied to highlight the effectiveness of key phrases toward sentiment delivery. Two GCNs are employed to respectively encode the syntactic and semantic information. A dual contrastive learning scheme is proposed to align the features from both syntactic and semantic spaces. Experiments are carried out on three benchmark datasets. Our method produces results considerably better than the state-of-the-art methods on the task of ALSC.

Author Contributions: Conceptualization, H.X. and Y.X.; methodology, H.X.; formal analysis, H.X. and Z.Y.; writing—original draft preparation, H.X. and Z.Y.; writing—review and editing, Y.X. and H.Z.; supervision, Y.X., Z.H. and H.Z.; funding acquisition H.Z. and Z.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Characteristic Innovation Projects of Guangdong Colleges and Universities (Nos. 2018KTSCX049), the Science and Technology Plan Project of Guangzhou under Grant Nos. 202102080258 and 201903010013.

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. Tang, D.; Qin, B.; Feng, X.; Liu, T. Effective LSTMs for target-dependent sentiment classification. *arXiv* **2015**, arXiv:1512.01100.
2. Ma, D.; Li, S.; Zhang, X.; Wang, H. Interactive attention networks for aspect-level sentiment classification. *arXiv* **2017**, arXiv:1709.00893.
3. 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, Copenhagen, Denmark, 7–11 September 2017; pp. 452–461.
4. Xu, G.; Zhang, Z.; Zhang, T.; Yu, S.; Meng, Y.; Chen, S. Aspect-level sentiment classification based on attention-BiLSTM model and transfer learning. *Knowl.-Based Syst.* **2022**, *245*, 108586. [[CrossRef](#)]
5. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
6. Zhang, C.; Li, Q.; Song, D. Aspect-based sentiment classification with aspect-specific graph convolutional networks. *arXiv* **2019**, arXiv:1909.03477.
7. Xu, K.; Zhao, H.; Liu, T. Aspect-specific heterogeneous graph convolutional network for aspect-based sentiment classification. *IEEE Access* **2020**, *8*, 139346–139355. [[CrossRef](#)]
8. Li, R.; Chen, H.; Feng, F.; Ma, Z.; Wang, X.; Hovy, E. Dual graph convolutional networks for aspect-based sentiment analysis. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Bangkok, Thailand, 1–6 August 2021; pp. 6319–6329.
9. Pang, S.; Xue, Y.; Yan, Z.; Huang, W.; Feng, J. Dynamic and multi-channel graph convolutional networks for aspect-based sentiment analysis. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Online, 1–6 August 2021; pp. 2627–2636.
10. Wang, T.; Isola, P. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In Proceedings of the International Conference on Machine Learning. PMLR, Virtual, 13–18 July 2020; pp. 9929–9939.
11. Chen, Q.; Zhang, R.; Zheng, Y.; Mao, Y. Dual Contrastive Learning: Text Classification via Label-Aware Data Augmentation. *arXiv* **2022**, arXiv:2201.08702.
12. Wang, K.; Shen, W.; Yang, Y.; Quan, X.; Wang, R. Relational graph attention network for aspect-based sentiment analysis. *arXiv* **2020**, arXiv:2004.12362.
13. Hu, J.; Li, Z.; Chen, Z.; Li, Z.; Wan, X.; Chang, T.H. Graph Enhanced Contrastive Learning for Radiology Findings Summarization. *arXiv* **2022**, arXiv:2204.00203.
14. Karamibekr, M.; Ghorbani, A.A. Sentiment analysis of social issues. In Proceedings of the 2012 international conference on social informatics, Alexandria, VA, USA, 14–16 December 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 215–221.
15. Pytkäinen, L. The neural basis of combinatory syntax and semantics. *Science* **2019**, *366*, 62–66. [[CrossRef](#)]
16. Shahi, T.; Sitaula, C.; Paudel, N. A Hybrid Feature Extraction Method for Nepali COVID-19-Related Tweets Classification. *Comput. Intell. Neurosci.* **2022**, *2022*, 5681574. [[PubMed](#)]
17. Sitaula, C.; Basnet, A.; Mainali, A.; Shahi, T.B. Deep learning-based methods for sentiment analysis on Nepali covid-19-related tweets. *Comput. Intell. Neurosci.* **2021**, *2021*, 2158184. [[CrossRef](#)] [[PubMed](#)]
18. Gou, J.; Yu, B.; Maybank, S.J.; Tao, D. Knowledge distillation: A survey. *Int. J. Comput. Vis.* **2021**, *129*, 1789–1819. [[CrossRef](#)]
19. Yang, M.; Jiang, Q.; Shen, Y.; Wu, Q.; Zhao, Z.; Zhou, W. Hierarchical human-like strategy for aspect-level sentiment classification with sentiment linguistic knowledge and reinforcement learning. *Neural Netw.* **2019**, *117*, 240–248. [[CrossRef](#)]
20. Tian, Y.; Chen, G.; Song, Y. Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 June 2021; pp. 2910–2922.
21. Yan, Z.; Pang, S.; Xue, Y. Semantic Enhanced Dual-Channel Graph Communication Network for Aspect-Based Sentiment Analysis. In Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, Guilin, China, 24–25 September 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 531–543.

22. Gao, T.; Yao, X.; Chen, D. Simcse: Simple contrastive learning of sentence embeddings. *arXiv* **2021**, arXiv:2104.08821.
23. Xu, P.; Chen, X.; Ma, X.; Huang, Z.; Xiang, B. Contrastive Document Representation Learning with Graph Attention Networks. *arXiv* **2021**, arXiv:2110.10778.
24. Li, Z.; Xu, B.; Zhu, C.; Zhao, T. CLMLF: A Contrastive Learning and Multi-Layer Fusion Method for Multimodal Sentiment Detection. *arXiv* **2022**, arXiv:2204.05515.
25. Liang, B.; Luo, W.; Li, X.; Gui, L.; Yang, M.; Yu, X.; Xu, R. Enhancing aspect-based sentiment analysis with supervised contrastive learning. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Online, 1–5 November 2021; pp. 3242–3247.
26. Pennington, J.; Socher, R.; Manning, C. GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; Association for Computational Linguistics: Doha, Qatar, 2014; pp. 1532–1543. [[CrossRef](#)]
27. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
28. 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), Dublin, Ireland, 23–24 August 2014; Association for Computational Linguistics: Dublin, Ireland, 2014; pp. 27–35. [[CrossRef](#)]
29. 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 (Volume 2: Short Papers), Baltimore, MD, USA, 22–27 June 2014; Association for Computational Linguistics: Baltimore, MD, USA, 2014; pp. 49–54. [[CrossRef](#)]
30. Sun, K.; Zhang, R.; Mensah, S.; Mao, Y.; Liu, X. Aspect-level sentiment analysis via convolution over dependency tree. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 5679–5688.
31. Zhang, M.; Qian, T. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Punta Cana, Dominican Republic, 8–12 November 2020; pp. 3540–3549.
32. Xiao, Z.; Wu, J.; Chen, Q.; Deng, C. BERT4GCN: Using BERT Intermediate Layers to Augment GCN for Aspect-based Sentiment Classification. *arXiv* **2021**, arXiv:2110.00171.
33. Zhang, K.; Zhang, K.; Zhang, M.; Zhao, H.; Liu, Q.; Wu, W.; Chen, E. Incorporating Dynamic Semantics into Pre-Trained Language Model for Aspect-based Sentiment Analysis. *arXiv* **2022**, arXiv:2203.16369.
34. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.