MDPI

Article

Knowledge-Enhanced Dual-Channel GCN for Aspect-Based Sentiment Analysis

Zhengxuan Zhang [1], Zhihao Ma [2], Shaohua Cai [3,*], Jiehai Chen [1] and Yun Xue [1]

1   School of Electronics and Information Engineering, South China Normal University, Foshan 528225, China
2   Wechat Open Platform Department, Tencent, Guangzhou 510220, China
3   Center for Faculty Development, South China Normal University, Guangzhou 510631, China
*   Correspondence: caishaohua@m.scnu.edu.cn

Abstract: As a subtask of sentiment analysis, aspect-based sentiment analysis (ABSA) refers to identifying the sentiment polarity of the given aspect. The state-of-the-art ABSA models are developed by using the graph neural networks to deal with the semantics and the syntax of the sentence. These methods are challenged by two issues. For one thing, the semantic-based graph convolution networks fail to capture the relation between aspect and its opinion word. For another, minor attention is assigned to the aspect word within graph convolution, resulting in the introduction of contextual noise. In this work, we propose a knowledge-enhanced dual-channel graph convolutional network. On the task of ABSA, a semantic-based graph convolutional netwok (GCN) and a syntactic-based GCN are established. With respect to semantic learning, the sentence semantics are enhanced by using commonsense knowledge. The multi-head attention mechanism is taken to construct the semantic graph and filter the noise, which facilitates the information aggregation of the aspect and the opinion words. For syntactic information processing, the syntax dependency tree is pruned to remove the irrelevant words, based on which more attention weights are given to the aspect words. Experiments are carried out on four benchmark datasets to evaluate the working performance of the proposed model. Our model significantly outperforms the baseline models and verifies its effectiveness in ABSA tasks.

Keywords: aspect-based sentiment analysis; graph convolutional networks; commonsense knowledge graph

MSC: 18C50

1. Introduction

Aspect-based sentiment analysis (ABSA) is a sentiment classification task that aims to identify the sentiment of given aspects [1]. Within ABSA, the sentiment of each aspect is classified according to a predefined set of sentiment polarities, i.e., positive, neutral or negative [2]. In recent years, ABSA yields very fine-grained sentiment information, which is useful for applications in a variety of domains [3].

In the context of advancing deep neural networks, state-of-the-art ABSA methods report high accuracy and strong robustness on benchmark datasets. During the progressing stage in ABSA tasks, efforts are generally made in two directions: one is to enhance significant information from the given text and the other is to filter the irrelevant information and its impact. A major step toward the comprehension of semantic information is the integration of attention mechanism with deep neural networks [4–6]. More attentive weights are assigned to aspect-related words, based on which to classify the sentiment polarity. Nevertheless, it can be challenging to capture syntax dependencies between the aspect and its contexts for attention-based models. More recently, research on graph neural networks (GNNs) has given rise to dealing with the syntactic information from dependency trees, a manner in which to prevent the syntactically irrelevant contextual noise [7–9]. The widespread GNNs, such as graph convolutional networks (GCNs) and graph attention

networks (GATs), are capable of encoding both the semantics and the syntax. This has been an ongoing trend to incorporate syntactic information and semantic information into GNN-based models [10–12].

In spite of the collaborative exploiting of syntax and semantics, two main limitations can be observed :

(1) For one thing, GNNs are generally used for tackling global syntactic information, while the mask operation is lastly performed to conceal the context words. Thereby, the sentiment of the aspect is determined. In practical application, the contextual noise can be introduced, which results in minor importance given to the aspect words.

(2) For another, the semantic-based GNNs are typically built up based on attention weights. With respect to the delicate relationship between aspects and opinion words, more attention is assigned to other words instead of the sentiment words. This can further confuse the sentiment aggregation. As presented in Figure 1, in the sentence *'Meal is very expensive for what you get'*, the aspect *'meal'* and its opinion word *'expensive'* are semantically insensitive.

Figure 1. Attention weights towards aspects. Words in black bold are aspects; words with a blue background are predicted attention weights; words with a green background represent desirable attention distribution. A word in the darker color indicates a greater weight and vice versa.

On the task of ABSA, this work focuses on establishing a Knowledge-Enhanced Dual-Channel Graph Convolutional Network (KDGCN). Two GCN-based modules, referred to as syntax-based GCN and semantic-based GCN, are developed to separately deal with the syntax structure and the semantic information. On the one hand, the syntactical dependency tree of the sentence is pruned to remove the connections of minor relevance to the aspect. Hence, the aspect-oriented syntactic information is sent to the syntax-based GCN. Besides, the position information and the attention mechanisms are taken to highlight the importance of the aspect. On the other hand, the external knowledge is introduced to enhance the semantic-based GCN. The word sentiment vectors, together with the supplementary of the aspect, are obtained (derived) by using SenticNet (i.e., a commonsense knowledge base); see Figure 2. A multi-headed attention mechanism is carried out to re-assign the attentive weights among words. The sentiment of the opinion words can thus be aggregated to the aspect via the knowledge-enhanced semantic-based GCN.
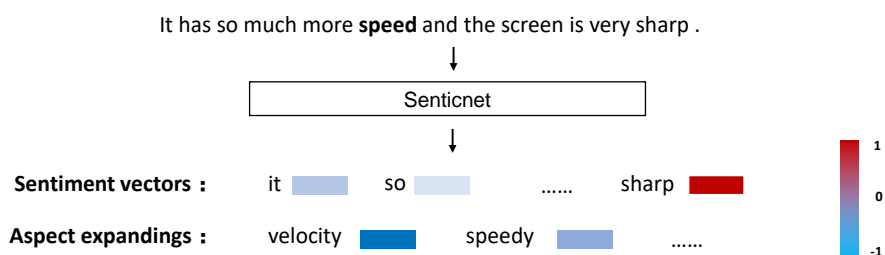
Figure 2. Sentiment vectors and aspect supplementary based on SenticNet. The different colors and shades represent the emotional polarity score of the word in SenticNet, where -1 is negative and 1 is positive.

Notably, a certain number of studies leverage the commonsense knowledge to enhance the sentiment expression and classify the sentiment polarity of the aspect [13,14].

Theoretically, the commonsense knowledge is involved with the background materials of the entities under discussion. The commonsense knowledge is preserved in the commonsense bases, such as ConceptNet [15], SenticNet [16] and WordNet [17], and recalled for processing. In most cases, the integration of semantic-related commonsense knowledge can generate noise from external information. Our model aims to exploit the sentence-related external knowledge, not just the sentiment information of each word, but also the relative knowledge of the aspect. In such a manner, the input of semantic-based GCN is distilled. Accordingly, the more-related information is preserved with the noise removed. The contributions of this paper are threefold and summarized as follows:

- Considering the deficiencies of the current ABSA methods, a dual-channel GCN based model is proposed, which processes both the syntax structure and the semantic information.
- The external knowledge is incorporated to enhance the semantics of the sentence, while the multi-head attention mechanism is taken to further filter the noise.
- Experiments on a variety of datasets indicate the effectiveness of the proposed method. Our model produces results considerably better than the baselines.

The paper is mainly divided into six sections. In the Introduction, we summarize the content of the article in general and propose our solutions for the challenge of the current ABSA task; in the Section 2, we will summarize the research related to our work; in the Section 3, we will introduce our proposed model and each module in detail; in the Section 4, we will conduct experiments on four public datasets and design ablation experiments; in the Section 5, we will further analyze the general situation of the model and the experimental results; in the Section 6, we summarize the full text.

2. Related Work

2.1. Aspect-Based Sentiment Analysis

As pointed out in the introduction, ABSA is a fine-grained sentiment classification task. Rather than assigning an overall sentiment polarity to a sentence or a document, ABSA aims at precisely determining the sentiment of a certain aspect. Early methods usually rely on manual features when predicting, which cannnot model the dependency relationship between the aspect and its context [18–20].

In recent years, advances in deep-learning algorithms significantly improved the working performance of ABSA, while a more detailed analysis of the textual information has risen [21,22]. The integrating of an attention mechanism into deep neural networks highlights the contribution of opinion words towards the aspects. [4–6,23–25] The relationship between aspect and its opinion words are reliably modeled in attention-based networks. Wang et al. [4] proposed an attention-long short-term memory (LSTM) method to obtain more-related information about a given aspect. Chen et al. [5] devised a hierarchical multi-attention model to address the long-range dependency between aspect and the opinion words. Whereas the attention mechanisms fails to cope with sentence syntax, by contrast, the employment of GCN takes advantages of the syntactic dependencies of the aspect and the opinion words. To be specific, an adjacency matrix is formed based on the syntactic dependent tree, which is further modeled to aggregate the sentiment information to the aspect by GCN [7,8]. Wang et al. [9] eliminated the noise from irrelevant contexts by constructing an aspect-oriented syntactic dependency tree, and then encoded the syntax relation by GNN. More recently, modules of multi-channel-GCNs have been carried out to resolve the syntax and semantics of the given sentence, which effectively optimizes the results of ABSA.

2.2. Graph Convolutional Networks

As a classical variant of GNN, GCN was originally proposed by Kipf et al. [26] in 2017. So far, GCN has shown its superiority in diversified NLP tasks, such as text classification [27,28], relation extraction [29,30], knowledge distillation [31] and machine translation [32].

Most studies [7,8] take GCNs to capture the syntactic information of a sentence where the nodes represent the words and the edges indicate the dependencies, which can induce representation vectors of nodes based on their neighborhoods' features. Likewise, the semantic relation within the sentence can also be obtained using GCN. In [10,11], the semantic graph was constructed with edges standing for the attention weights. Therefore, both semantic features and syntactic features can be extracted via GCN-based modules.

Considering a graph as structured data, the multilayers of GCN are responsible for information delivery. As such, every single node within the graph can learn the global information. Let $G = (V, E)$, where $V = \{v_i, v_2, \ldots, v_n\}$ is a set of $N = |V|$ nodes and E is the set of edges, and it represents an $n-$node graph with an adjacency matrix of $A \in \mathbb{R}^{k \times k}$. In a graph, let $v_i \in V$ to denote a node and $e_{ij} = (v_i, v_j) \in E$ to denote an edge between v_i and v_j.

GCN can only capture information about neighbors with a layer. However, information about more neighborhoods can be integrated when multilayers of GCN are stacked. We define h_i^l as the output of node i on the $l - th$ layer and h_i^0 as the initial state of node i. The graph convolution of node i can be written as:

$$h_i^l = \sigma(\sum_{j=1}^{k} A_{ij} W^l h_j^{l-1} + b^l) \tag{1}$$

where W^l is the weight of linear transformation, b^l is the bias and σ is a nonlinear function such as *Relu*.

2.3. Commonsense Knowledge

The commonsense knowledge for NLP is typically obtained through large-scale corpus training and saved in commonsense bases. The commonsense is taken as prior knowledge for the pre-training of knowledge-enhanced approaches. SenticNet [16] is one such commonsense knowledge base, which contains $100k$ concepts related to sentiment expression. (e.g., mood, polarity, semantics and so on). Additionally, these affective properties provide concept-level representation and semantic connections to the words.

To facilitate access to corresponding knowledge, SenticNet provides an application programming interface. A series of sentiment scores of the word and its related concepts can be obtained from the interface (as shown in Figure 2), which can expand the semantics of the sentence.

The application of SenticNet into ABSA shows its distinctiveness in sentiment representation learning [13,33]. Ma et al. [13] utilized the commonsense from SenticNet to generate essays more closely surrounding the semantics of the input topics. Zhou et al. [14] enlarged the sentence semantics using SenticNet 5, and then jointly modeled the syntactic dependency trees and commonsense graph. Regardless of additional key information, the filter of the noise during the external knowledge introducing remains unsettled.

3. Methodology

The architecture of KDGCN is presented in Figure 3. Our model consists of five key components, i.e., a sentence encoder, a knowledge enhancement module, a semantic learning module, a syntax aware module and a sentiment classifier. Firstly, each word of the sentence is encoded as a vector by the sentence encoder. At the same time, the sentence is input into the knowledge enhancement module, and the sentiment vector of each word and the expanding words of aspect are obtained from SenticNet; secondly, the hidden state vector of the sentence is sent into a semantic learning module and a syntax aware module, respectively, to obtain the syntactic and semantic representation. Finally, we can obtain the sentiment polarity of the aspect from the sentiment classifier.
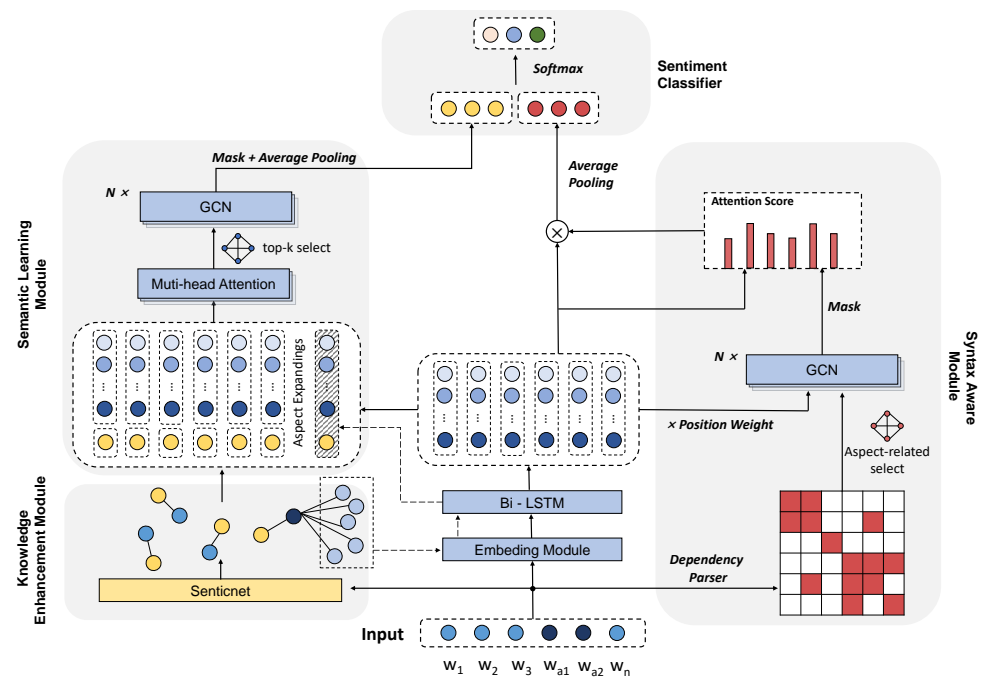
Figure 3. Overall architecture of the proposed Knowledge-Enhanced Dual-Channel Graph Convolutional Network.

3.1. Sentence Encoder

Glove embedding. For a sentence $c = \{w_1, w_2, \ldots, w_n\}$ with the aspects $a = \{w_{a1}, w_{a2}, \ldots, w_{an}\}$, we take the pre-trained embedding matrix $E \in \mathbb{R}^{|V| \times d_e}$ to map each word into a low-dimensional vector, where $|V|$ represents the lexicon size and d_e is the dimension of the word vector [34].

BERT embedding. BERT [35] is a commonly used sentence encoder in recent years. Each sentence is pre-processed by adding *[CLS]* at the beginning and *[SEP]* at the end, respectively, to obtain $c' = \{w_0, w_1, \ldots, w_{n+1}\}$, where w_0 and w_{n+1} denote the two special tokens inserted. Then, c' is fed into BERT to obtain the textual feature representation $X = \{x_0, x_1, \ldots, x_{n+1}\}$, where $x_i \in \mathbb{R}^{d_{bert}}$.

A Bidirectional LSTM (Bi-LSTM) is employed for sentence encoding. The given sentence embedding is sent to Bi-LSTM to generate the hidden state vector $H^{LSTM} = \{h_1, h_2, \ldots, h_n\}$. Specifically, the vector $H^{LSTM} \in \mathbb{R}^{2d_h}$ is the hidden state at a time step and is the hidden state vector dimension of LSTM.

3.2. Knowledge Enhancement Module

Word sentiment enhancement: For the given sentence c, the sentiment vector of each word can be obtained based on the commonsense from SenticNet. A 23-dimensional sentiment vector $H^{LSTM} \in \mathbb{R}^{23}$ hat represents the sentence that is derived. Besides, for the words that do not appear in SenticNet, the zero-vector is used instead. Then, H^{LSTM} and H^{sen} are fused to obtain the sentence representation, which is:

$$H^c = [H^{LSTM}; H^{sen}] \tag{2}$$

with $H^c \in \mathbb{R}^{2d_h + 23}$.

Aspect knowledge enhancement: In terms of the aspects a, the relative words of each word within a is collected from SenticNet, i.e., $\{w_{ex1}, w_{ex2}, \ldots, w_{exn}\}$. For the purpose of word supplementary, the first five words in relation to the aspect are used. All the relative words are also mapped to word embeddings and encoded with the Bi-LSTM encoder.

$$H^{ex} = [H^{LSTM}_{ex}; H^{sen}_{ex}] \tag{3}$$

where H_{ex}^{LSTM} stands for the hidden state vector of Bi-LSTM, and H_{ex}^{sen} is the corresponding sentiment vector. The aspect expanding vector is denoted as $H^{ex} \in \mathbb{R}^{2d_h+23}$.

Notably, since the word co-occurrence in the corpus has an impact on the word embedding of glove, to prevent the noise fusion, the aspect relative words are not pre-trained by glove. We take a $\langle unk \rangle$ for relative words that are absent from the given texts. Similarly, the absent-words of SenticNet are taken in place of zero.

3.3. Semantic Learning Module

Motivated by [10], most short sentences are of confused syntactic structure. That is, the rigid extraction of syntactic information can lead to the misinterpretation of the sentiment information. For this reason, a semantic learning module based on GCN is proposed to capture the semantic information among words. Both the enhanced sentiment vector and the aspect expanding vector are sent to the semantic learning module, which aims to further enrich the semantic information.

Node construction: Each word w_i from the sentence, together with each aspect relative word w_{exi}, is taken as a node. All nodes constitute a node set V.

Edge construction: The edge indicates the relationship between word nodes. Concretely, two semantic-related nodes are connected with an edge and vice versa. To capture the semantic relation of each word, we employ $K - heads$ multi-head self-attention mechanism to compute the attention weight, i.e.,

$$A_{ttn} = \frac{(H_{se}W_{se,k})(H_{se}W_{se,q})^T}{\sqrt{d_{head}}} \tag{4}$$

where

$$H_{se}^{(0)} = H^c \tag{5}$$

$$d_{head} = \frac{d_{lstm}}{k} \tag{6}$$

where $H_{se}^{(0)} \in \mathbb{R}^{2d_h+23}$ is the commonsense-enhanced hidden layer output; K is the head number of multi-head attention mechanism; $W_{se,k}$ and $W_{se,q} \in \mathbb{R}^{(2d_h+23) \times d_{head}}$ are trainable matrices. Subsequently, based on the top-k selecting approach, the largest k values of each dimension are selected and set to 1, while others are set to 0. Hence, the adjacency matrix A_{se} is obtained; see Equation (7). Corresponding to the edge construction principle, the adjacency matrix with value 1 denotes the semantic relevance between nodes. Notably, the A_{se} remains symmetric with the application of the top-k selector.

$$A_{se} = topk \sum_{i=0}^{k} A_{ttn} \tag{7}$$

Thereby, a graph $G_{sem} = (A_{se}, H^c)$ that concerns the node representations and the adjacency matrix is constructed. The graph is fed into the N-layer GCN to obtain the hidden layer state H_{se} :

$$H_{se}^{(l+1)} = GCN(A_{se}, H_{se}^{(l)}, W_{se}^{(l)}) \tag{8}$$

where $H_{se}^{(l)} \in \mathbb{R}^{(2d_h+23) \times d_{gcn}}$ stands for the parametric matrix of GCN. The mask operation is conducted on non-aspect words, following with the average pooling to compute semantic hidden layer output h_{se}, which is written as:

$$mask = \begin{cases} 0 & 1 \leq t < \tau+1, \tau+m < t < n \\ 1 & \tau+1 \leq t \leq \tau+m \end{cases} \tag{9}$$

$$h_{se} = f(mask(H_{se})) \tag{10}$$

where $\tau+1 \leq t \leq \tau+m$ indicates the aspect index and $f(\cdot)$ is the average pooling function.

3.4. Syntax Aware Module

The syntax aware module is devised by modifying the method proposed by Zhang et al. [7]. The sentence syntax is characterized by the syntax dependency tree. Note that not all context words are syntactically related to the aspect—an aspect-related selection approach is taken to reshape the syntax dependency tree. Only if a context word reaches the aspect within n hops can the dependency edge between nodes be kept. We can thus revise the adjacency matrix A_0 to A_{sy}. In this way, the revised graph is written as $G_{sy} = (A_{sy}, H^{LSTM})$, where H^{LSTM} is the current node representation. Before sending G_{sy} to GCN, the position-aware transformation is performed [7]:

$$q_i = \begin{cases} 1 - \frac{\tau+1-i}{n} & 1 \leq i \leq i+1 \\ 0 & \tau+1 \leq i \leq \tau+m \\ 1 - \frac{\tau+1-i}{n} & \tau+m < i \leq n \end{cases} \tag{11}$$

with

$$\mathcal{F}(h_i) = q_i h_i \tag{12}$$

where $q_i \in \mathbb{R}$ the position weight of the i-th token and $\mathcal{F}(\cdot)$ is the function for position weight assignment. The syntactic information is learned by using graph convolution. The syntactic hidden layer output is expressed as:

$$H_{sy}^{(l)} = \mathcal{F}(H_{sy}^{(l-1)}) \tag{13}$$

$$H_{sy}^{(l+1)} = GCN(A_{sy}, H_{sy}^{(l)}, W_{sy}^{(l)}) \tag{14}$$

$$H_{sy}^{(0)} = \mathcal{F}(H^{LSTM}) \tag{15}$$

where $H^{(l)} \in \mathbb{R}^{2d_h \times d_{gcn}}$ is a trainable parametric matrix. Similar to the semantic-based GCN, the syntactic hidden state representation W_{sy} is revised via masking (Equation (16)). The

$$H^t = mask(H_{sy}) \tag{16}$$

where $H^t = \{h_1^t, h_2^t, \ldots, h_j^t\}$. The outcome hidden layer state from Equation (16) concentrates more on the aspect words. In addition, to further detect the significant semantic feature concealed within the syntax structure, the attention weight of each context word is assigned. The dot product of h_i^t and h_i are obtained to denote the syntactic representation, i.e.,

$$h_{sy} = \sum_{j=1}^{n} a_j h_j^t \tag{17}$$

$$a_j = \frac{exp(\beta_j)}{\sum_{i=1}^{n} exp(\beta_j)} \tag{18}$$

$$\beta_t = \sum_{i=1}^{n} h_j^t h_i = \sum_{i=\tau+1}^{\tau+m} h_j^t h_i \tag{19}$$

3.5. Sentiment Classifier

Both the semantic representation and the syntactic representation are so far computed. We shall thus concatenate $h_s e$ and $h_s y$ to obtain the final representation h_a (Equation (20)). The sentiment polarity of the given aspect is classified by sending h_a to the Softmax classifier, which is:

$$H_a = [H_{se}; H_{sy}] \tag{20}$$

$$y = softmax(h_a) \tag{21}$$

3.6. Model Training

The training process is performed by using the categorical cross entropy and L_2 regularization as the loss function:

$$Loss = -\sum_i \sum_j y_i^j log(p_i^j) \tag{22}$$

where *i* is the index of the ABSA sample and *j* is the corresponding sentiment polarity.

4. Experiment

In this section, we designed the main experiment and attention visualization to verify the effectiveness of our model on the ABSA task. Specifically, we first introduce the benchmark datasets used in our experiment, and then briefly introduce the details of the experiment and the selected baseline. Then, we carried out the main experiment and analyzed the experimental results. In addition, in order to explore the contribution of each module to the model, we designed ablation experiments and analyzed the mechanism of knowledge enhancement in attention visualization.

4.1. Dataset

To verify the working performance of the proposed model, experiments were carried out on four publicly available benchmark datasets, i.e., Rest14 and Lap14 from SemEval 2014 [36], Rest15 from SemEval 2015 [37] and Rest16 from SemEval 2016 [1], containing reviews of restaurant and laptop domains.

Every single sentence from the datasets contains at least one aspect. The sentiment polarity of each aspect is given as well, including: positive, negative and neutral. For example, in the sentence "*Great food but the service was dreamful!*", there are two aspect terms, *'food'* and *'service'*, and their sentiment polarity are positive and negative, respectively. The details of each dataset are presented in Table 1.

Table 1. Statistics of datasets.

Dataset	Positive		Neutral		Negative	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Rest14	2164	728	637	196	807	196
Lap14	994	341	464	169	870	128
Rest15	1178	439	1c50	35	382	328
Rest16	1620	597	88	38	709	190

4.2. Implementation Details

The best test result of each method was taken for evaluation. For the proposed model, the initialization of word embeddings was conducted using Glove [38] and uncased BERT [35], respectively. The pretrained Glove provides a 300-dimensional word vector, with a learning rate of 0.001 and a batch size of 64. Moreover, the dimension of Bert-based word embeddings was 768, with a learning rate of 0.00002 and a batch size of 32. The head number of multi-head attention network was set to 1. The value of top-k selection was 2. Besides, the Adam optimizer was employed. The L_2 regularization weight was 0.0001. The value of dropout was determined within the interval of [0.4, 0.6] using grid searching. With respect to the GCN in our model, the number of layers and the dimension of hidden layers ranged within [1,4] and [100, 200], respectively, which were also selected via grid searching.

4.3. Baseline

For the purpose of validating the effectiveness of our model, twelve state-of-the-art methods were taken for comparison, which are presented as follows:

- **CDT** [8]: GCN is taken to deal with the syntax dependency tree, which aims to learn the sentence syntactic information. Specifically, it exploits a GCN to model the structure of a sentence through its dependency tree, where node (word) embeddings of the tree are initialized by means of a Bi-LSTM network.
- **ASGCN** [7]: On the task of ABSA, GCN is applied to learn the aspect-specific representation for the first time. Specifically, it starts with a LSTM layer to encode the sentence, and a multi-layered graph convolution structure is implemented on top of the LSTM output to obtain aspect-specific features.
- **SK-GCN** [14]: A syntax-based GCN and a knowledge-based GCN are designed to model the syntax dependency tree and knowledge graph, respectively. Specifically, it obtains the sentiment information from the SenticNet to enrich the representation of a sentence toward a given aspect.
- **R-GAT** [4]: It reshapes and prunes an ordinary dependency parse tree to obtain an aspect-oriented dependency tree structure rooted at a target aspect. Then, a relational graph attention network (R-GAT) is introduced to encode the new tree structure for sentiment prediction.
- **DualGCN** [5]: Considering the complementarity of syntax structures and semantic correlations, a dual graph convolutional network is proposed to tackle both the syntactic information and semantic information.
- **DMGCN** [11]: A multi-channel GCN-based method is developed to exploit not only the syntax and the semantics, but also the correlated information from the generated graph.
- **BERT** [35]: The basic BERT model is established based on a bidirectional transformer. With the concatenation of sentence and the corresponding aspect, BERT can be applied to ABSA.
- **SK-GCN+BERT** [14], **R-GAT+BERT** [9], **DualGCN+BERT** [10], **DMGCN+BERT** [11] : The pre-trained BERT is integrated with SK-GCN, R-GAT, DualGCN and DMGCN, respectively, where BERT is used for sentence encoding.
- **TGCN+BERT** [39]: The dependency type is identified with type-aware graph convolutional networks, while the relation is distinguished with attention mechanism. The pre-trained BERT is used for sentence encoding.

4.4. Experimental Results

Experimental results on all datasets are exhibited in Table 2. In this experiment, we took accuracy and macro-F1 as the method evaluation metrics. Comparing with the baseline models, **KDGCN** generally obtained the best and most consistent results in all evaluation settings. However, our model with the Bert encoder was less competitive than **DMGCN+BERT** on the dataset of Rest14. A possible explanation is that the pre-trained Bert contains a wealth of semantic information. The semantic enhancement via SenticNet is not that distinctive. With respect to the Glove-based word embeddings, the performance of **KDGCN** was 0.93% and 2.89% higher than **DMGCN** in accuracy and Macro-F1, respectively.

Comprehensively, current GCN-based models focus on encoding either the syntactic information (e.g., **ASGCN**, **CDT**, **R-GAT** and **TGCN+BERT**) or the semantic-integrated syntactic information (e.g., **DualGCN** and **DMGCN**). The performance of these methods largely depends on their fitting capabilities. By contrast, the proposed model adopted the aspect-related selection approach to prune the edges of the syntax dependency tree, based on which the unrelated information to the aspect was eliminated. On the other hand, the commonsense knowledge was introduced to enhance the semantic information and the sentiment of the aspect. In this way, the results of ABSA can be improved.

Furthermore, **SK-GCN** also uses the external knowledge derived from SenticNet to construct the syntax-based GCN and semantic-based GCN. In comparison with **SKGCN**, our model performs significantly better on all datasets. Clearly, **KDGCN** is capable of exploiting the commonsense knowledge in ABSA tasks. As such, it is rational to expect the

integration of external knowledge into the given sentence and thus improved sentiment classification results.

Table 2. Experimental results on four public datasets. The results of **R-GAT** and **R-GAT+BERT** are retrieved from [40], and others are retrieved from the original papers.

Models	Rest14		Lap14		Rest15		Rest16	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
CDT [8]	74.66	73.66	77.19	72.99	-	-	85.58	69.93
ASGCN [7]	80.77	72.02	75.55	71.05	79.89	61.89	88.99	67.48
SK-GCN [14]	80.36	70.43	73.20	69.18	80.12	60.70	85.17	68.08
R-GAT [9]	83.30	76.08	77.42	73.76	80.83	64.17	88.92	70.89
DualGCN [10]	84.27	78.08	78.48	74.74	-	-	-	-
DMGCN [11]	83.98	75.59	78.48	74.90	-	-	-	-
Our KDGCN	**84.91**	**78.48**	**79.00**	**75.03**	**82.10**	**67.13**	**90.74**	**73.46**
BERT [35]	85.62	78.28	77.58	72.38	83.48	66.18	90.10	74.16
SK-GCN+BERT [14]	83.48	75.19	79.00	75.57	83.20	66.78	87.19	72.02
R-GAT+BERT [9]	86.60	81.35	78.21	74.07	83.22	69.73	89.71	76.62
DualGCN+BERT [10]	87.13	81.16	81.80	78.10	-	-	-	-
DMGCN+BERT [11]	**87.66**	**82.79**	80.22	77.28	-	-	-	-
TGCN+BERT [39]	86.16	79.95	80.88	77.03	85.26	71.69	92.32	77.29
Our KDGCN+BERT	87.23	81.69	**82.60**	**79.55**	**85.98**	**72.40**	**93.66**	**82.49**

4.5. Ablation Study

An ablation study was conducted to quantitively investigate the importance of different modules in the proposed model. The results of the ablation study are given in Table 3 and Figure 4. We took the basic KDGCN as the baseline and ablated the knowledge enhancement module, semantic learning module, syntax aware module and the aspect-related select procedure. According to Table 3, the most important component for the proposed model is the syntax aware module. The accuracy drop on four datasets were 6.78%, 6.12%, 4.61% and 3.08%, which are significant. Obviously, the use of syntactic information plays a pivotal role in ABSA. Moreover, the contributions of the semantic learning module and the knowledge enhancement module are comparable. The integration of commonsense knowledge into the semantic learning process gives an improvement of the sentiment classification performance. Lastly, withdrawal of the aspect-related selection also caused a minor decrease of the working performance.

Table 3. Results of the ablation study.

| Model	Rest14		Lap14		Rest15		Rest16	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
KDGCN *w/o* aspect-related select	84.11	77.02	76.96	73.14	80.10	66.48	89.61	70.94
KDGCN *w/o* syntax aware module	78.13	68.34	72.88	68.13	77.49	52.19	87.66	66.74
KDGCN *w/o* knowleage enhancement	83.13	76.01	76.49	72.38	81.55	58.71	89.61	72.01
KDGCN *w/o* semantic learning module	83.22	75.35	76.33	73.27	79.89	60.87	89.28	71.96
KDGCN	**84.91**	**78.48**	**79.00**	**75.03**	**82.10**	**67.13**	**90.74**	**73.46**
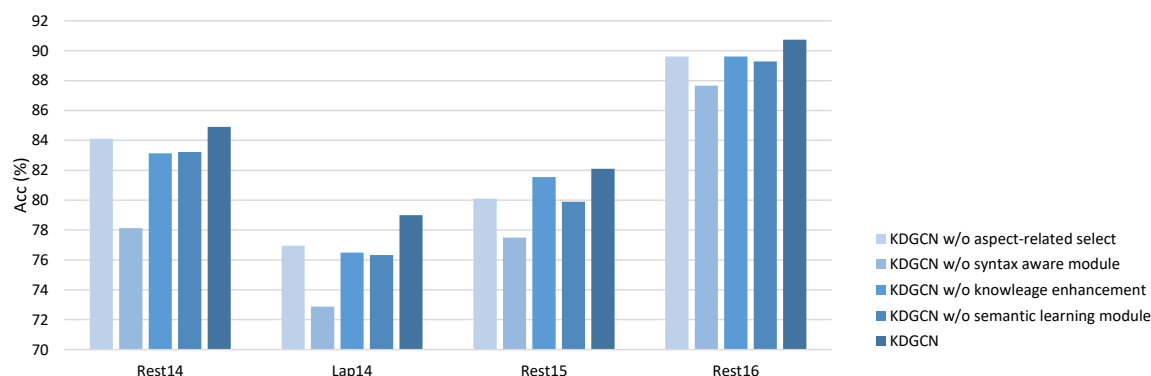
Figure 4. Results of the ablation study. Different columns show the performance of different models on different datasets.

4.6. Attention Visualization

To investigate the effectiveness of the knowledge enhancement, we visualized the attention matrix. In our model, the semantics enhancement is carried out by using the commonsense from SenticNet. The connection between the aspect and its opinion word is established and enhanced. The syntax-based GCN also removes the irrelevant information by encoding the syntax dependency tree. Cases are presented to demonstrate the attention weight distribution. In the first line of Figure 5, the attentive weights are assigned based on a basic multi-head attention mechanism. One can easily see that the minor attention was given to the opinion word **'excellent'** of the aspect **'food'**. Likewise, the attention weight of **'food'** toward **'excellent'** was also weakened. With the integration of commonsense knowledge, the relationships of both **'food'** and **'excellent'** to the context word **'meal'** were established. That is, the **'food-meal'** edge and the **'excellent-meal'** edge can be constructed by using a top-k selection. As a result, the sentiment information of **'excellent'** can be aggregated on the aspect word **'food'** with the encoding of GCN. Besides, the syntactic-based GCN, which deals with the syntactic relation among words, also facilitates the determination of aspect sentiment polarity.

Similarly, from the two figures in the second line, we can see that the aspect word **'waiter'** established a direct connection with the opinion word **'helpful'** after knowledge enhancement. Additionally, from the two figures in the last line, the aspect word **'sauce'** and the opinion word **'flavorful'** are connected through the path **'sauce-dough-flavorful'** after knowledge enhancement, so that the sentiment polarity of the aspect words can be better predicted after the subsequent network structure.
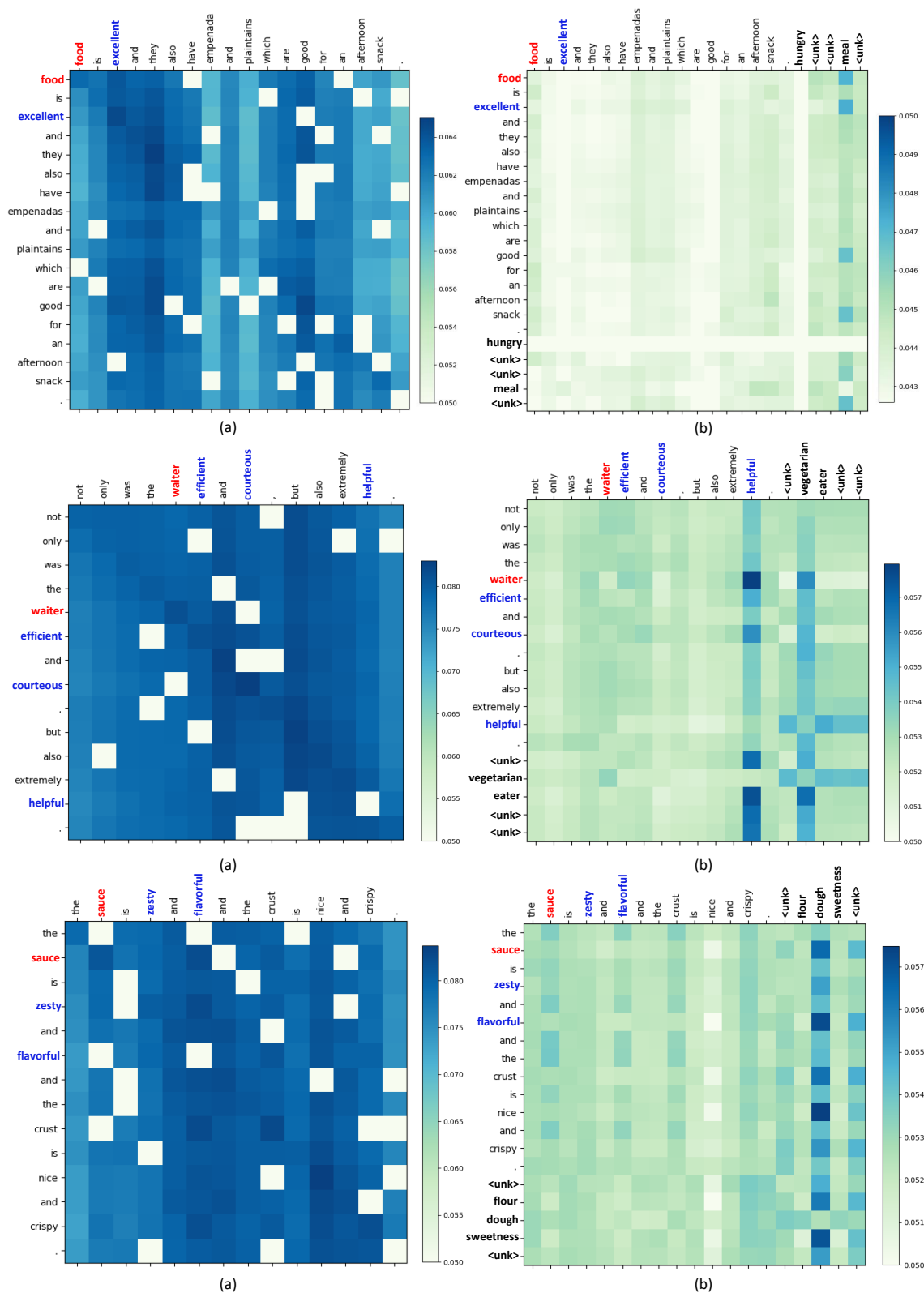
Figure 5. An illustration on knowledge-enhancement. (**a**) Basic attention matrix of the sentence. (**b**) Knowledge-enhanced attention matrix of the sentence. The red words are aspect words, the blue words are opinion words and the black bold words are aspect-expansion words.

5. Discussion

Through a series of experiments, we can see that our KDGCN performs well on the ABSA task. Specifically, in the main experiment part (Section 4.4), the accuracy and F1-score of our model on the four datasets are generally higher than baselines, especially compared with SK-GCN [14], which also uses SenticNet for knowledge enhancement; our improvement was 2–5%. In the ablation study, we removed the semantic learning module, the syntax aware module and so on, which proves that semantics and syntax are both important for ABSA tasks. In addition, after removing the knowledge enhancement module, the model performance also decreased significantly on the four datasets, indicating that our knowledge enhancement facilitates ABSA tasks.

Moreover, we also found the limitations of our model. Take DMGCN [11] and the use of the glove encoder as an example—KDGCN's improvement on Lap14 was not as big as that on rest14 (0.52% and 0.93%, respectively). This may be because most of the Lap14 datasets are proper nouns (such as *Windows 7* and *Microsoft*), and they do not have obvious emotional clues. Different from it, most of the words in Rest14 are daily words, so the sentiment information is rich and can be further enhanced through SenticNet. In order to obtain more semantic information and deeper connections, large-scale knowledge graphs can be introduced into the ABSA task in future work.

6. Conclusions

In this work, we propose a knowledge-enhanced dual-channel graph convolutional network to deal with the ABSA tasks. A semantic-based GCN and a syntactic-based GCN are devised to encode both the sentence semantics and the syntax. On the one hand, the external commonsense knowledge is introduced to enhance the semantics, based on which more attention is assigned to the aspect and its relevant words. On the other hand, the syntactic-based GCN processing on the syntax dependency tree further filters the low-dependency words. We demonstrate the effectiveness of our method on four benchmark datasets, obtaining state-of-the-art results on both accuracy and macro-F1. Comparing with the baseline models, the proposed method is the best alternative that produces results considerably better than the widely-applied approaches in ABSA. In the ablation experiment, we tested the contribution of each module to the model and verified that our innovation is effective. In addition, we also carried out a case analysis to further intuitively demonstrate the role of knowledge enhancement in promoting our task.

However, SenticNet is a small-scale knowledge base with shallow and limited semantics, which limits the performance of the model. Therefore, future work can consider exploring the use of a larger scale knowledge graph (such as Wikipedia) to enhance the knowledge of ABSA tasks, which can provide more clues to predict the sentiment polarity of the aspect.

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

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. 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, San Diego, CA, USA, 16–17 June 2016.
2. Li, H.; Xue, Y.; Zhao, H.; Hu, X.; Peng, S. Co-attention networks for aspect-level sentiment analysis. In Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, Guilin, China, 24–25 September 2019; Springer: Cham, Switzerland, 2019.
3. Schouten, K.; Frasincar, F. Survey on aspect-level sentiment analysis. *IEEE Trans. Knowl. Data Eng.* **2015**, *28*, 813–830. [CrossRef]
4. 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, Austin, TX, USA, 1–5 November 2016.
5. 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, 9–11 September 2017.
6. Fan, F.; Feng, Y.; Zhao, D. Multi-grained attention network for aspect-level sentiment classification. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018.
7. 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), Hong Kong, China, 3–7 November 2019; pp. 4567–4577.
8. 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
9. 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.
10. 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), Virtual Event, 1–6 August 2021.
11. Pang, S.; Xue, Y.; Yan, Z.; Huang, W.; Feng, J. Dynamic and multi-channel graph convolutional networks for aspect-based sentiment analysis. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, Virtual, 1–6 August 2021.
12. Dai, A.; Hu, X.; Nie, J.; Chen, J. Learning from word semantics to sentence syntax by graph convolutional networks for aspect-based sentiment analysis. *Int. J. Data Sci. Anal.* **2022**, *14*, 17–26. [CrossRef]
13. Yang, P.; Li, L.; Luo, F.; Liu, T.; Sun, X. Enhancing topic-to-essay generation with external commonsense knowledge. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy 28 July–2 August 2019; pp. 2002–2012.
14. Zhou, J.; Huang, J.X.; Hu, Q.V.; He, L. Sk-gcn: Modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification. *Knowl.-Based Syst.* **2020**, *205*, 106292. [CrossRef]
15. Speer, R.; Chin, J.; Havasi, C. Conceptnet 5.5: An open multilingual graph ofgeneral knowledge. In Proceedings of the AAAI, San Francisco, CA, USA, 4–9 February 2017.
16. Cambria, E.; Poria, S.; Hazarika, D.; Kwok, K. SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In Proceedings of the AAAI, Edmonton, AB, Canada, 13–17 November 2018.
17. Miller, G.A. WordNet: A lexical database for English. *Commun. ACM* **1995**, *38*, 39–41. [CrossRef]
18. Jiang, L.; Yu, M.; Zhou, M.; Liu, X.; Zhao, T. Target-dependent twitter sentiment classification. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, ON, USA, 19–24 June 2011.
19. 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.
20. Ding, X.; Zhang, Y.; Liu, T.; Duan, J. Deep learning for event-driven stock prediction. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.
21. Tang, D.; Qin, B.; Feng, X.; Liu, T. Effective LSTMs for target-dependent sentiment classification. In Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 11–16 December 2016; pp. 3298–3307.
22. Majumder, N.; Poria, S.; Gelbukh, A.; Akhtar, M.S.; Cambria, E.; Ekbal, A. IARM:Inter-aspect relation modeling with memory networks in aspect-based sentiment analysis. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 3402–3411.
23. Tang, D.; Qin, B.; Liu, T. Aspect level sentiment classification with deep memory network. *arXiv* **2016**, arXiv:1605.08900.
24. Ma, D.; Li, S.; Zhang, X.; Wang, H. Interactive attention networks for aspect-level sentiment classification. *arXiv* **2017**, arXiv:1709.00893.
25. Huang, B.; Ou, Y.; Carley, K.M. Aspect level sentiment classification with attention-over-attention neural networks. In Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, Stockholm, Sweden, 10–15 July 2018; Springer: Cham, Switzerland, 2018.
26. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. In Proceedings of the ICLR, Toulon, France, 24–26 April 2017.

27. Yao, L.; Mao, C.; Luo, Y. Graph convolutional networks for text classification. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 7370–7377.
28. Huang, L.; Ma, D.; Li, S.; Zhang, X.; Wang, H. Text level graph neural network for text classification. 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. 3444–3450.
29. Zhang, Y.; Qi, P.; Manning, C.D. Graph convolution over pruned dependency trees improves relation extraction. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 2205–2215.
30. Sun, K.; Zhang, R.; Mao, Y.; Mensah, S.; Liu, X. Relation extraction with convolutional network over learnable syntax-transport graph. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34.
31. Gou, J.; Yu, B.; Maybank, S.J.; Tao, D. Knowledge distillation: A survey. *Int. J. Comput. Vis.* **2021**, *129*, 1789–1819. [CrossRef]
32. Bastings, J.; Titov, I.; Aziz, W.; Marcheggiani, D.; Sima'an, K. Graph convolutional encoders for syntax-aware neural machine translation. In Proceedings of the EMNLP, Copenhagen, Denmark, 9–11 September 2017; pp. 1957–1967.
33. Li, Y.; Pan, Q.; Yang, T.; Wang, S.; Tang, J.; Cambria, E. Learning word representations for sentiment analysis. *Cogn. Comput.* **2017**, *9*, 843–851. [CrossRef]
34. Bengio, Y.; Ducharme, R.; Vincent, P.; Jauvin, C. A neural probabilistic language model. *J. Mach. Learn. Res.* **2003**, *3*, 1137–1155.
35. 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, Volume 1 (Long and Short Papers), Minneapolis, MI, USA, 2–7 June 2019; pp. 4171–4186.
36. Suresh, M. 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.
37. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Man, har, S.; Androutsopoulos, I. Semeval-2015 task 12: Aspect based sentiment analysis. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, CL, USA, 4–5 June 2015.
38. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conferenceon Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
39. 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, Mexico City , Mexico, 6–11 June 2021.
40. 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]