Article

Attention-Enhanced Graph Convolutional Networks for Aspect-Based Sentiment Classification with Multi-Head Attention

Guangtao Xu ¹, Peiyu Liu ^{1,*}, Zhenfang Zhu ² , Jie Liu ¹ and Fuyong Xu ¹
¹ School of Information Science and Engineering, Shandong Normal University, Jinan 250358, China; 2019020878@stu.sdnu.edu.cn (G.X.); 2019020851@stu.sdnu.edu.cn (J.L.); 2020020961@stu.sdnu.edu.cn (F.X.)

² School of Information Science and Electrical Engineering, Shandong Jiaotong University, Jinan 250357, China; zhuzf@sdjtu.edu.cn

* Correspondence: liupy@sdnu.edu.cn; Tel.: +86-131-8889-9297

Abstract: The purpose of aspect-based sentiment classification is to identify the sentiment polarity of each aspect in a sentence. Recently, due to the introduction of Graph Convolutional Networks (GCN), more and more studies have used sentence structure information to establish the connection between aspects and opinion words. However, the accuracy of these methods is limited by noise information and dependency tree parsing performance. To solve this problem, we proposed an attention-enhanced graph convolutional network (AEGCN) for aspect-based sentiment classification with multi-head attention (MHA). Our proposed method can better combine semantic and syntactic information by introducing MHA and GCN. We also added an attention mechanism to GCN to enhance its performance. In order to verify the effectiveness of our proposed method, we conducted a lot of experiments on five benchmark datasets. The experimental results show that our proposed method can make more reasonable use of semantic and syntactic information, and further improve the performance of GCN.

Keywords: aspect-based sentiment classification; attention mechanism; multi-head attention; graph convolutional network

Citation: Xu, G.; Liu, P.; Zhu, Z.; Liu, J.; Xu, F. Attention-Enhanced Graph Convolutional Networks for Aspect-Based Sentiment Classification with Multi-Head Attention. *Appl. Sci.* **2021**, *11*, 3640. <https://doi.org/10.3390/app11083640>

Academic Editor: Giancarlo Mauri

Received: 1 March 2021

Accepted: 16 April 2021

Published: 18 April 2021

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

Copyright: © 2021 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-based sentiment classification (ABSC) [1] is a fine-grained subtask in the field of sentiment analysis. Its purpose is to identify the sentiment polarity of the aspects that clearly appear in the sentence. For example, in a restaurant review: “This restaurant has a good environment, but the price is a bit expensive”, the sentiment polarity of the two aspects “environment” and “price” are positive and negative, respectively. In our research, aspects are usually noun or noun phrases. The difficulty of aspect-based sentiment classification task is how to accurately find out the opinion words related to aspects. For example, in the above example, the opinion words corresponding to environment and price are “good” and “expensive”, respectively.

Early work is mainly based on neural networks, like Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) [2,3]. Since Tang et al. pointed out the importance of modeling the semantic relationship between context and aspects [4], more and more research consider introducing an attention mechanism on the basis of RNN or CNN to establish connections between various aspects and context words [5–10]. However, due to the complexity of sentences, the attention mechanism cannot accurately capture the relationship between aspects and context words. For example, in the sentence “so delicious was the food but terrible servers”, for the aspect “food”, the attention mechanism may assign a higher weight to the word “terrible” that is closer to it.

Other works consider using sentence structure information to establish connections between aspects and opinion words. The main idea is to construct a dependency tree

based on the syntactic structure of the sentence, use the dependency tree, and Graph Convolutional Networks (GCN) to update the representation of the sentence [11–14]. There is no doubt that dependency trees can establish long-term dependencies between aspects and opinion words. However, due to the limitations of the dependency tree itself, when the sentence structure is more complex or the sentence expression is more colloquial, the model usually cannot correctly predict the sentiment polarity of the aspect. Specifically, when the dependency tree cannot correctly establish the connection between the aspect and the opinion word, the model will not be able to accurately predict the sentiment polarity of the aspect. In addition, in the process of using the dependency tree and GCN to update the word representation, noise information is usually integrated into the new representation, causing the model to learn wrong parameters.

To solve the above two problems, we proposed an attention-enhanced graph convolutional network for aspect-based sentiment classification with multi-head attention. For the first problem, we combine the graph convolutional network with the multi-head attention, using the advantages of the multi-head attention mechanism to capture contextual semantic information to alleviate the defects of the graph convolution network in processing data with unobvious syntactic features. For the second problem, we introduce the attention mechanism into the traditional graph convolutional network, alleviating the problem of introducing too much noise when updating node information by assigning appropriate attention weights to each adjacent node.

In this paper, we divide multi-head attention into multi-head self-attention (MHSA) and multi-head interactive attention (MHIA). The model is divided into two parts, and both parts use the same input. The first part captures contextual semantic information through the attention coding layer (ACL), and the second part integrates syntactic information through the attention-enhanced graph convolutional network. Finally, we use multi-head interactive attention (MHIA) to integrate the above information to obtain the final feature representation. We conducted a lot of experiments on five benchmark datasets. The experimental results show that our proposed method can effectively utilize syntactic and semantic information, and further improve the performance of graph convolutional networks.

Our contributions are as follows:

1. We introduced an attention mechanism into the graph convolutional network to enhance its performance.
2. We introduced multi-head self-attention to capture contextual semantic information, used multi-head interactive attention to interact semantic and syntactic information to obtain a more complete feature representation.
3. In order to better match the dependency tree, we applied the pre-trained Bidirectional Encoder Representations from Transformers (BERT) model of the whole word masking version to the task and achieved better performance.
4. The experimental results on five benchmark datasets prove that our proposed model is effective compared with other mainstream models.

2. Related Work

A lot of research work has proved that the method of using neural network model to model the feature representation of word sequence in ABSC can well capture the context information. For example, Convolutional Neural Networks (CNNs) [15,16], Recurrent Neural Networks (RNNs) [17], or a combination of both Convolutional Recurrent Neural Networks (CRNNs) [18]. Recently, in view of the good performance of the attention mechanism in modeling contextual information, more and more works have begun to consider the use of attention-based neural network models for ABSC. The main idea is to capture and establish the connection between aspects and opinion words through the attention mechanism. Strictly speaking, attention-based neural network models can also be used as a method of using sentence structure information, because the distance between aspects and opinion word is generally not too far. Wang et al. [5] proposed using attention-based to identify important sentiment information related to aspects. Li et al. [19]

introduced a multi-layer attention mechanism to capture the long-distance opinion words of the aspect. For similar purposes, Tang et al. [20] proposed a deep memory network with multi-hop attention and explicit memory. Fan et al. [21] proposed a multi-granularity attention network. In addition, in view of the advantages of the multi-head attention mechanism in modeling contextual semantic relations, Song et al. [22] proposed an attention encoder network to draw the hidden states and semantic interactions between target and context words. Zhu et al. [23] proposed a novel Interactive Dual Attention Network (IDAN) model that aims to interactively learn the representation between contextual semantics and sentimental tendency information.

Sun et al. [24] utilizes the transformer module to learn the word-level representations of aspects and context, respectively, and further utilizes the tree transformer module to obtain the phrase-level representations of contexts. In addition, they adopt dual-pooling method and multi-grained attention network to extract high quality aspect–context interactive representations. Zhang et al. [25] introduced multi-head interactive attention based on the work of Song et al. [22] to enhance the interaction between aspect items and context.

Other works consider using sentence structure information for ABSC. Aspects are usually the core of this task. Therefore, using sentence structure information to establish connections between aspects and opinion words can improve the performance of sentiment classification models. Since the graph convolutional network [26] was first introduced, due to its excellent performance in processing graph structure information, it has been quickly applied to various tasks in the field of natural language processing (NLP) and achieved good results. Marcheggiani and Titov [27] proposed a GCN-based semantic role annotation model. Huang and Carley [28] propose a novel target-dependent graph attention network, which explicitly utilizes the dependency relationship among words. Wang et al. [29] defined a unified aspect-oriented dependency tree structure rooted at a target aspect, and proposed a relational graph attention network (GAT) to encode the new tree structure for sentiment prediction. Tang et al. [30] proposed a dependency graph enhanced dual-transformer network by jointly considering the flat representations learnt from transformer and graph-based representations learnt from the corresponding dependency graph.

In addition, Gao et al. [31] constructed three target-dependent variants of the BERT model for target-dependent sentiment classification. Chen and Qian [32] proposed a transfer capsule network model to transfer sentence-level semantic knowledge from document-level sentiment classification to ABSC. Tang et al. [33] proposed a gradual self-supervised attention learning method to strengthen the performance of the attention mechanism. Sun et al. [34] transformed the ABSC task into a sentence pair classification task by constructing auxiliary sentences.

3. Methodology

Given a sentence sequence $W^c = \{w_1^c, w_2^c, \dots, w_n^c\}$ composed of n words and an aspect sequence $W^a = \{w_1^a, w_2^a, \dots, w_m^a\}$ composed of m words, the goal of this model is to predict the sentiment polarity of sentence W^c on aspect W^a . Figure 1 shows the network architecture of our proposed attention-enhanced graph convolutional network (AEGCN) model. We use attention coding layer to capture semantic information (it contains a multi-headed self-attention and a point-wise convolution transformation), use the syntactic dependency tree and attention-enhanced graph convolutional network to capture syntactic information, and use the multi-head interact attention mechanism to interact the two parts of information. Finally, we use the output of multi-head interactive attention to do a pooling splicing operation as the feature vector for sentiment prediction. Next, we will introduce the various components of the AEGCN model.

Figure 1. Overview of the proposed model for aspect-based sentiment classification.

3.1. Input Layer

We use two methods to obtain embedding vector and contextualized representation.

The first method is pre-trained GloVe static embedding and BiLSTM. GloVe is a popular embedding method, and we use it to embed each word token into a low-dimensional real-valued vector space. Through the pre-trained GloVe embedding matrix $L \in \mathbb{R}^{d_m \times |V|}$, each word is matched to the corresponding embedding vector $e_i \in \mathbb{R}^{d_m \times 1}$, where d_m represents the embedding dimension of the word vector, and $|V|$ represents the size of the vocabulary. Then, we send the word embedding matrix to the BiLSTM to obtain the hidden state output of the input layer. BiLSTM is an extension of RNN. It solves the gradients vanishing or exploding problem in standard RNN, makes the hidden state output of any time step contain previous and subsequent timing information.

The second method is pre-trained BERT. It is worth noting that we are using the Whole Word Masking variant of BERT-Large. The reason is that the tokenizer used by BERT is Wordpiece Tokenizer, which cuts certain words into a collection of several subwords, and then randomly selects subwords to mask for prediction training. In order to match our dependency tree, we need to add these subwords first. Therefore, compared with the word

vector obtained by the random sub-word masking method, the word vector obtained by the whole word masking method is more consistent with our model.

We use the output $H^c = \{h_1^c, h_2^c, \dots, h_n^c\} \in \mathbb{R}^{d_{hid} \times n}$ of BERT or BiLSTM as the contextual representation of the input text.

3.2. Attention Coding Layer

As shown in Figure 2, Attention Coding Layer (ACL) includes a Multi-Head Attention (MHA) and a Point-wise Convolution Transformation (PCT). We use MHA to capture the semantic information of the sentence, obtain the hidden layer state based on the contextual semantic information, and further transform the semantic information extracted by MHA through PCT.

Figure 2. The structure of the Attention Coding Layer.

3.2.1. Multi-Head Self-Attention

Multi-head Attention (MHA) uses multiple heads to capture the semantic information of the context in parallel, each attention head focuses on different aspects, and finally, the information of each attention head is combined to obtain the semantic representation of the input sentence. According to whether the two inputs of MHA are the same or not, we divide it into multi-head self-attention (MHSA) and multi-head interactive attention (MHIA). In this layer, we use MHSA to capture contextual semantic information. Formally, given two identical inputs $H^c = \{h_1^c, h_2^c, \dots, h_n^c\}$, MHSA is defined as:

$$MHSA(H^c, H^c) = (head_1 \oplus head_2 \oplus \dots \oplus head_h) \cdot W^O \quad (1)$$

$$head_i = Attention_i(H^c, H^c) \quad (2)$$

$$Attention_i(H^c, H^c) = Softmax\left(\frac{H^c H^{c^T}}{\sqrt{d_k}}\right) H^c \quad (3)$$

Among them, h represents the number of attention heads in multi-head attention, $\oplus$ represents vector connection, $W^O \in \mathbb{R}^{d_{hid} \times d_{hid}}$ is a parameter matrix, $head_i$ represents the output of the i -th attention head, d_k is the dimension size of h_i^c .

3.2.2. Point-Wise Convolution Transformation

We perform two convolution operations on the output of MHSA, and the size of the convolution kernel is 1. The activation function of the first convolution operation is ReLU, and the second convolution operation uses the linear activation function. Formally, given the input sequence h , the PCT is defined as:

$$PCT(h) = \text{ReLU}(h * W_c^1 + b_c^1) * W_c^2 + b_c^2 \quad (4)$$

Where $*$ represents the convolution operation. $W_c^1 \in \mathbb{R}^{d_{hid} \times d_{hid}}$ and $W_c^2 \in \mathbb{R}^{d_{hid} \times d_{hid}}$, respectively, represent the weight of the two convolution kernels. b_c^1 and b_c^2 are the bias of the two convolution kernels. We denote the output of PCT as $H^A = \{h_1^A, h_2^A, \dots, h_n^A\}$.

3.3. AEGCN Layer

In order to use the syntactic information of the sentence when predicting the sentiment polarity of the sentence, we constructed the L-layer AEGCN to capture the syntactic information. First, we use spaCy toolkit to construct a dependency tree for each sentence, and then we use these dependency trees to get the corresponding adjacency matrix $A \in \mathbb{R}^{n \times n}$. Among them, n represents the length of the sentence. Each element in the adjacency matrix represents whether the two word-nodes at the corresponding position are adjacent in the sentence structure. If they are adjacent, the value is 1, and if they are not adjacent, the value is set to 0. For example, element A_{ij} in the i -th row and j -th column of the matrix represents whether the i -th word and the j -th word in the sentence are adjacent in the dependency tree. If it is 1, it means adjacent, and if it is 0, it means not adjacent. In particular, the diagonal elements of the adjacency matrix are all 1, that is, each word is adjacent to itself. After getting the adjacency matrix A , we can use it to capture the syntactic information of the sentence.

Figure 3 shows an example of an AEGCN layer. We denote the output of each layer in AEGCN as $H^l = \{h_1^l, h_2^l, \dots, h_n^l\}$, $l \in [1, L]$. If all adjacent nodes of node i are represented as $\mathcal{N}_i$, the output of the i -th node in l -th layer AEGCN can be expressed as:

$$h_i^l = \text{ReLU}\left(\sum_{j=1}^n A_{ij} g_j^{l-1} W^l + b^l\right) \quad (5)$$

$$g_j^{l-1} = e_{ij}^l h_j^{l-1} \quad (6)$$

$$e_{ij}^l = \text{attention}(h_i^{l-1}, h_j^{l-1}), j \in \mathcal{N}_i \quad (7)$$

$$\text{attention}(h_i^{l-1}, h_j^{l-1}) = \text{softmax}(h_i^{l-1} h_j^{l-1}) \quad (8)$$

Among them, the weight W^l and the bias b^l are parameters that need to be learned, A_{ij} represents the adjacency coefficient, and e_{ij}^l represents the normalized attention coefficient of node i and j in the l -th AEGCN. The output of the last layer of AEGCN is $H^L = \{h_1^L, h_2^L, \dots, h_n^L\} \in \mathbb{R}^{d_{hid} \times n}$.

3.4. Interaction Layer

In order to realize the interaction between syntactic information and semantic information, we added an MHIA after ACL and AEGCN, respectively. Expressing the output of ACL and AEGCN as H^A and H^L , respectively, the output H^{AI} and H^{LI} of the interactive layer are calculated as follows:

$$H^{AI} = \text{MHIA}(H^{Aa}, H^L) \quad (9)$$

$$H^{LI} = \text{MHIA}(H^{La}, H^A) \quad (10)$$

where H^{Aa} and H^{La} denote the aspects in H^A and H^L , respectively.

Figure 3. An example of attention-enhanced graph convolutional network (AEGCN) layer.

3.5. Output Layer

As shown in Figure 1, we first perform the average pooling operation on the two outputs of the interactive layer, and then connect these average pooling outputs as the final feature representation. The final feature representation h^o is computed as follow:

$$h^o = h^1 \oplus h^2 \quad (11)$$

$$h^1 = \sum_{i=1}^m h_i^{AI} / m \quad (12)$$

$$h^2 = \sum_{i=1}^m h_i^{LI} / m \quad (13)$$

Finally, we send the feature representation h^o into the fully connected softmax layer to obtain the probability distribution p of the sentiment polarity.

$$p = \text{softmax}(W_p h^o + b_p) \quad (14)$$

where W_p and b_p are the learnable parameters, and d_p represents the number of categories of sentiment polarity.

3.6. Training

The model is trained by standard gradient descent algorithm, and the objective function is defined as minimizing the cross-entropy loss with L_2 regularization:

$$\text{Loss} = - \sum_{(d, \tilde{p}) \in D} \log(P_{\tilde{p}}) + \lambda \|\theta\|_2 \quad (15)$$

4. Experiments

In this chapter, we first introduced five datasets and experimental parameter settings. Secondly, we compared our proposed model with other popular models and analyzed the comparison results. Finally, we conducted experimental analysis on our proposed model from multiple perspectives.

4.1. Datasets and Experimental Settings

In order to make a comprehensive comparison with other baseline models and the most advanced models, we conducted experiments on five datasets. Among them, Twitter is composed of Twitter posts collected by Dong et al. [2], the other four (Lap14, Rest14, Rest15, Rest16) are, respectively, from SemEval 2014 task 4 [35], SemEval 2015 task 12 [36], SemEval 2016 Task 5 [37]. Among them, SemEval 2014 task 4 contains two datasets, Lap14 and Rest14. The detailed statistical results of each dataset are shown in Table 1.

Table 1. Detailed statistical of five datasets in our experiments.

Dataset	Category	Positive	Neural	Negative
Twitter	Train	1561	3127	1560
	Test	173	346	173
Lap14	Train	994	464	870
	Test	341	169	128
Rest14	Train	2164	637	807
	Test	728	196	196
Rest15	Train	912	36	256
	Test	326	34	182
Rest16	Train	1240	69	439
	Test	469	30	117

In the experiment, we used two different input methods. In AEGCN-GloVe, we use a 300-dimensional pre-trained GloVe vector as a static embedding, and the vector dimension of the hidden state is also set to 300. In AEGCN-BERT, we use pre-trained BERT as the embedding layer and fine-tune it on our task. Both the embedding dimension and the hidden state dimension are 768. All weight parameters in the model (except BERT) are initialized with uniform distribution. We use Adam optimizer [38] in our model, using different learning rates for GloVe static embedding and BERT embedding: 1×10^{-3} and 3×10^{-5} . The coefficient of the L_2 regularization term is 1×10^{-5} . The dropout rate and batch size are 0.1 and 64, respectively. In addition, according to the optimal experimental results of the model, the number of AEGCN layers is set to 2. We use Accuracy and Macro-F1 indicators as the criteria for evaluating model performance. The experimental results are the average of three random initialization runs.

4.2. Model Comparisons

In order to comprehensively evaluate and analyze the performance of our proposed models, we compared them with a series of baselines and state-of-the-art models. According to their method types, we divide these models into attention-based and syntactic-based models.

4.2.1. Attention-Based Models

ATAE-LSTM: They proposed to use LSTM and attention mechanism to obtain a vector representation for sentiment prediction, and add aspect embedding to each context word embedding.

MemNet: They proposed to use external memory to model the context representation, using a multi-hop attention architecture.

IAN [39]: An interactive modeling model of aspects and context is designed, using BiRNN and attention mechanism to interactively learn aspects and context representation.

AOA: They proposed an attention-over-attention neural network to model aspects and sentences in a joint way and explicitly capture the interaction between aspects and context sentences.

T-MGAN: They proposed a transformer-based, multi-granularity attention network (T-MGAN), which uses a tree transformer module to obtain a phrase-level representation, using a dual-pool operation and a multi-granularity attention network to extract high-quality feature representations.

IMAN: They made improvements based on the AEN model, adding a multi-head interactive attention mechanism to the last layer to interact with context information and aspect information to obtain the final feature representation.

AEN-Glove: They proposed an attention encoder network to model the relationship between context and specific aspects, and the embedding layer uses Glove static embedding.

AEN-BERT: Different from the AEN-Glove model, a pre-trained BERT-base model is used in the embedding layer.

4.2.2. Syntactic-Based Models

LSTM + SynATT [40]: They proposed a method that can better capture the semantic meaning of aspects, proposing an attention model to integrate syntactic information into the attention mechanism.

CDT: They propose to use BiLSTM to obtain the feature representation of the sentence, and to further enhance the embedding by directly performing convolution operations on the dependency tree.

ASGCN: It is proposed to learn feature representations of specific aspects through GCN and dependency trees to solve the long-distance multi-word dependency problem.

BiGCN [41]: They built a concept hierarchy on both the syntactic and lexical graphs for differentiating various types of dependency relations or lexical word pairs, designing a bi-level interactive graph convolution network to fully exploit these two graphs.

4.3. Results and Analysis

Table 2 shows the experimental results of our proposed model and other comparison models. According to the data in the table, we can get the following conclusions. The performance of our proposed model is stronger than all comparison models on most datasets, and the improvement is particularly obvious when using pre-trained BERT as embedding (using pre-trained BERT as embedding has a huge improvement in the performance of the model). The experimental results prove the effectiveness of our model. Models based on graph convolutional networks and dependency trees are significantly better than models based on attention in capturing long-distance dependent information. This result reflects the superiority of graph convolutional network in ABSC task.

Table 2. Model comparison results of accuracy and macro-F1(%) on five datasets. The best results of different categories in each dataset are shown in bold. The best results in all models are bolded and underlined. “-” means not reported.

Category	Model	Twitter		Lap14		Rest14		Rest15		Rest16	
		Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Att	ATAE-LSTM	-	-	68.70	-	77.20	-	-	-	-	-
	MenNet	71.48	69.90	70.64	65.17	79.61	69.64	77.31	58.28	85.44	65.99
	IAN	72.50	70.81	72.05	67.38	79.26	70.09	78.54	52.65	84.74	55.21
	AOA	72.30	70.20	72.62	67.52	79.97	70.42	78.17	57.02	87.50	66.21
	T-MGAN	71.23	70.63	76.38	73.02	82.06	72.65	-	-	-	-
	AEN	72.83	69.81	73.51	69.04	80.98	72.14	-	-	-	-
	AEN-BERT	74.71	73.13	79.93	76.31	83.12	73.76	-	-	-	-
	IMAN	75.72	74.50	80.53	76.91	83.95	75.63	-	-	-	-
Syn	LSTM + SynATT	-	-	72.57	69.13	80.45	71.26	80.28	65.46	83.39	66.83
	CDT	74.66	73.66	77.19	72.99	82.30	74.02	-	-	85.58	69.93
	ASGCN	72.15	70.40	75.55	71.05	80.77	72.02	79.89	61.89	88.99	67.48
	BiGCN	74.16	73.35	74.59	71.84	81.97	73.48	81.16	64.79	88.96	70.84
Ours	AEGCN	73.86	71.59	75.91	71.88	81.43	73.66	80.85	63.96	88.76	68.73
	AEGCN-BERT	75.99	75.01	80.37	76.68	84.46	76.33	83.92	67.08	89.61	70.71

Compared with the attention-based model, our model can capture long-distance dependence information through the attention-enhanced graph convolutional network and dependency tree, obtaining better feature representation; compared with the model based on the graph convolutional network, our model can effectively integrate semantic

information and syntactic information through the multi-head attention mechanism, and alleviate the impact of the limitations of the dependency tree.

Compared with the AEN model, our model has a significant improvement effect. The AEN model models the context and aspect words separately, extracts semantic features through a multi-head attention mechanism, and interacts context information and aspect information to obtain feature representations. The effect of the model highly depends on whether the multi-head attention mechanism can accurately establish the connection between the aspect words and the context. However, due to the defects of the attention mechanism in capturing long-distance dependent information and the complexity of the sentence structure itself, simply using the attention mechanism cannot accurately model the relationship between context and aspect words.

The IMAN model is an improvement based on the AEN model. They added a multi-head interactive attention mechanism to the last layer, interacting context information and aspect information, and obtained good results. Our model has achieved better results on all datasets except Lap14, which shows that using sentence structure information as a supplement to the model to determine the polarity of emotions can further improve the performance of the model. However, our model achieved sub-optimal results on the Lap14 dataset, and we suspect that this dataset may be insensitive to syntactic information.

The T-MGAN model also uses sentence structure information. They used the tree transformer module to capture phrase grammatical information and obtained phrase-level feature representation. However, they only considered the phrase information in the sentence, not the global structure information. The experimental results of our model are better than the T-MGAN model, which shows that the global structure information of sentences is positively helpful for aspect-based sentiment analysis.

Compared with the GCN-based models, our model has a certain improvement on all datasets. Through our analysis, the GCN-based models have a significant improvement in the construction of long-distance multi-word dependence compared with the traditional attention-based neural network model. However, the prerequisite is that the dependency tree must be complete and effective. When the sentence is too complex, the dependency tree cannot accurately establish the relationship between the aspects and the opinion words, which will lead to the degradation of model performance. Secondly, in the process of using dependency tree and graph convolutional network to introduce syntactic information, noise information will also be introduced into it. This problem becomes more obvious when the graph convolutional network layer is deeper. Our model combines the multi-head attention mechanism with the graph convolutional network, adds semantic information on the basis of syntactic information, and interacts with the two parts of information to obtain a more complete feature representation, thereby enhancing the accuracy of the model.

4.4. Ablation Study

In order to further study the influence of each component of AEGCN on performance improvement, we designed several ablation experiments. The experimental results are shown in Table 3 (using Accuracy as the evaluation indicator).

First, we removed the attention mechanism (w/o att) in the graph convolutional network, and the experimental results dropped slightly, indicating that using the attention mechanism to assign weights to the syntactic-related items of each node can improve the performance of the model. We removed the MHIA behind ACL and AEGCN, respectively (w/o MHIA¹ and w/o MHIA²). It can be seen from the experimental results that these two components play a positive role in the process of learning semantic information. Compared with MHIA¹, MHIA² has a greater impact on the model. We believe that the reason is that the graph convolutional network will also capture noise information in the process of capturing syntactic information, which leads to a decrease in the accuracy of the model when judging sentiment polarity.

Table 3. Ablation study results of accuracy (%) on five datasets.

Model	Twitter	Lap14	Rest14	Rest15	Rest16
	Acc	Acc	Acc	Acc	Acc
AEGCN-BERT	75.99	80.37	84.46	83.92	89.61
w/o att	75.12	79.13	83.45	83.14	88.36
w/o MHIA ¹	74.77	78.91	83.11	82.73	88.13
w/o MHIA ²	73.43	78.31	82.11	81.52	86.93

4.5. Case Study

In order to better understand our model, we compared it with the AEN and ASGCN models in several test examples. The experimental results are shown in Table 4. The attention visualization column in the table shows the attention scores of each model, and we use darker to lighter colors according to their scores. In the first example, “delicious food but terrible environment”, because the sentence contains two entities “food” and “environment” and two opinion words “delicious” and “terrible”. The AEN model based on the attention mechanism cannot capture the connection between the two very well, leading to incorrect prediction results. Due to the complexity of the sentence, in the second example, the AEN model is still unable to correctly model the connection between aspect words and emotional words, and the attention mechanism focuses on the wrong point. In the ASGCN model, since “lovely” is also close to “son” in the sentence structure, it is also updated to the sentence representation when the node information is updated, which causes the model to also assign a high weight to “lovely” when calculating the attention score. When the dependency tree contains noisy information or syntactic information is not obvious, the ASGCN model cannot correctly model the relationship between the aspect words and the opinion words, which leads to prediction errors.

Table 4. Visual analysis cases of ASGCN, AEN, and AEGCN.

Model	Aspect	Attention Visualization					Prediction	Label
AEGCN	food	Delicious	food	but	terrible	environment	Positive	Positive
	son	His	lovely	son	is	always	lazy	Negative
AEN	food	Delicious	food	but	terrible	environment	Neutral	Positive
	son	His	lovely	son	is	always	lazy	Negative
ASGCN	food	Delicious	food	but	terrible	environment	Positive	Positive
	son	His	lovely	son	is	always	lazy	Negative

Our model correctly predicted the sentiment polarity of the two samples, which means that our model can make better use of the semantic and syntactic information of the sentence when processing complex sentences, thereby improving the model’s performance to a certain extent, and can maintain good stability on different datasets.

4.6. Impact of the AEGCN Layers

In order to verify the influence of the number of AEGCN layers on the model, we used different layers of AEGCN on the Lap14 dataset to compare its effects. Accuracy is still used as the evaluation indicator, and the experimental results are shown in Figure 4.

Figure 4. Impact of the AEGCN layers.

It can be seen from the above figure that when the number of AEGCN layers exceeds two, the model performance begins to decline. Due to the limitations of the dependency tree itself, when the number of AEGCN layers is too large, a lot of noise is also updated to the representation of the last layer, which has a negative impact on the performance of the model.

5. Conclusions

Recently, neural network models based on dependency trees have attracted widespread attention in ABSC. However, due to the imperfect parsing performance of the dependency tree, the noise information will be updated to the sentence representation during the process of introducing syntactic information. Therefore, we propose an AEGCN model with multi-head attention. This model uses a multi-head self-attention mechanism to obtain input semantic information, using attention-enhanced graph convolutional network and dependency tree to obtain input syntactic information. The multi-head interactive attention mechanism integrates semantic and syntactic information to obtain the final feature vector to predict the sentiment polarity. The experimental results on five datasets show that the interactive integration of sentence syntactic information and semantic information can indeed effectively improve the performance of the model.

This paper aims to strengthen the interaction between syntactic information and semantic information, make rational use of the advantages of graph convolutional network and attention mechanism, and use the semantic information and syntactic information of sentences for aspect-based sentiment prediction. However, we only interacted semantic information and syntactic information in the last layer of the model. Future research can consider building a multi-layer network architecture, interacting semantic information and syntactic information at each layer.

Author Contributions: Conceptualization, G.X. and P.L.; funding acquisition, P.L.; methodology, G.X.; project administration, Z.Z.; resources, P.L. and Z.Z. Writing—original draft, G.X.; writing—review and editing, J.L., F.X.; visualization, G.X.; investigation, G.X.; validation, G.X. All authors have read and agreed to the published version of the manuscript.

Funding: Our work was supported by the Science Foundation of the Ministry of Education of China (no.14YJC860042), the National Social Science Fund (19BYY076), and the Shandong Provincial Social Science Planning Project (no.19BJCJ51/18CXWJ01/18BJYJ04/16CFXJ18).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to all commenters for their valuable and constructive comments.

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

References

1. Liu, B. Sentiment Analysis and Opinion Mining. *Synth. Lect. Hum. Lang. Technol.* **2012**, *5*. [[CrossRef](#)]
2. 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, Baltimore, MD, USA, 22–27 June 2014; pp. 49–54.
3. Vo, D.T.; Zhang, Y. Target-dependent twitter sentiment classification with rich automatic features. In Proceedings of the Twenty-Fourth International Joint Conference in Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.
4. 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.
5. 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, 1–4 November 2016; pp. 606–615.
6. Xue, W.; Li, T. Aspect Based Sentiment Analysis with Gated Convolutional Networks. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; pp. 2514–2523.
7. Li, X.; Bing, L.; Lam, W.; Shi, B. Transformation Networks for Target-Oriented Sentiment Classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; pp. 946–956.
8. Chen, P.; Sun, Z.; Bing, L.; Yang, W. Recurrent Attention Network on Memory for Aspect Sentiment Analysis. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9–11 September 2017; pp. 452–461.
9. Huang, B.; Ou, Y.; Carley, K.M. Aspect level sentiment classification with attention-over-attention neural networks. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*; Springer: Cham, Switzerland, 2018; pp. 197–206.
10. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 2048–2057.
11. Beck, D.; Haffari, G.; Cohn, T. Graph-to-Sequence Learning Using Gated Graph Neural Networks. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; pp. 273–283.
12. Huang, B.; Carley, K.M. Parameterized Convolutional Neural Networks for Aspect Level Sentiment Classification. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 1091–1096.
13. 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. 4560–4570.
14. 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. 5683–5692.
15. Yoon, K. Convolutional neural networks for sentence classification. In Proceedings of the Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1746–1751.
16. Johnson, R.; Zhang, T. Semi-supervised convolutional neural networks for text categorization via region embedding. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 919. [[PubMed](#)]
17. Castellucci, G.; Filice, S.; Croce, D.; Basili, R. UNITOR: Aspect Based Sentiment Analysis with Structured Learning. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 23–24 August 2014; pp. 761–767.
18. Lai, S.; Xu, L.; Liu, K.; Zhao, J. Recurrent convolutional neural networks for text classification. In Proceedings of the AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; Volume 29.
19. Li, C.; Guo, X.; Mei, Q. Deep Memory Networks for Attitude Identification. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, Cambridge, UK, 6–10 February 2017; pp. 671–680.
20. Tang, D.; Qin, B.; Liu, T. Aspect Level Sentiment Classification with Deep Memory Network. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing; Association for Computational Linguistics (ACL), Berlin, Germany, 7–12 August 2016; pp. 214–224.
21. Fan, F.; Feng, Y.; Zhao, D. Multi-grained Attention Network for Aspect-Level Sentiment Classification. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics (ACL), Brussels, Belgium, 31 October–4 November 2018; pp. 3433–3442.
22. Song, Y.; Wang, J.; Jiang, T.; Liu, Z.; Rao, Y. Attentional encoder network for targeted sentiment classification. *arXiv* **2019**, arXiv:1902.09314.
23. Zhu, Y.; Zheng, W.; Tang, H. Interactive Dual Attention Network for Text Sentiment Classification. *Comput. Intell. Neurosci.* **2020**, *2020*, 8858717. [[CrossRef](#)] [[PubMed](#)]

24. Sun, J.; Han, P.; Cheng, Z.; Wu, E.; Wang, W. Transformer Based Multi-Grained Attention Network for Aspect-Based Sentiment Analysis. *IEEE Access* **2020**, *8*, 211152–211163. [[CrossRef](#)]
25. Zhang, Q.; Lu, R.; Wang, Q.; Zhu, Z.; Liu, P. Interactive Multi-Head Attention Networks for Aspect-Level Sentiment Classification. *IEEE Access* **2019**, *7*, 160017–160028. [[CrossRef](#)]
26. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In Proceedings of the 2017 International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017.
27. Marcheggiani, D.; Titov, I. Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9–11 September 2017; pp. 1506–1515.
28. Huang, B.; Carley, K.M. Syntax-Aware Aspect Level Sentiment Classification with Graph Attention 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. 5472–5480.
29. Wang, K.; Shen, W.; Yang, Y.; Quan, X.; Wang, R. Relational Graph Attention Network for Aspect-based Sentiment Analysis. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 3229–3238.
30. Tang, H.; Ji, D.; Li, C.; Zhou, Q. Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 6578–6588.
31. Gao, Z.; Feng, A.; Song, X.; Wu, X. Target-Dependent Sentiment Classification With BERT. *IEEE Access* **2019**, *7*, 154290–154299. [[CrossRef](#)]
32. Chen, Z.; Qian, T. Transfer Capsule Network for Aspect Level Sentiment Classification. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; Association for Computational Linguistics (ACL), Florence, Italy, 28 July–2 August 2019; pp. 547–556.
33. Tang, J.; Lu, Z.; Su, J.; Ge, Y.; Song, L.; Sun, L.; Luo, J. Progressive Self-Supervised Attention Learning for Aspect-Level Sentiment Analysis. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), Florence, Italy, 28 July–2 August 2019; pp. 557–566.
34. Sun, C.; Huang, L.; Qiu, X. Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Volume 1, pp. 380–385.
35. 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; pp. 27–35.
36. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; Androutsopoulos, I. Semeval-2015 task 12: Aspect based sentiment analysis. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, CO, USA, 4–5 June 2015; pp. 486–495.
37. 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 International Workshop on Semantic Evaluation (SemEval 2015), San Diego, CA, USA, 16–17 June 2016; pp. 19–30.
38. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, Chia Laguna Resort, Italy, 13–15 May 2010; pp. 249–256.
39. Ma, D.; Li, S.; Zhang, X.; Wang, H. Interactive Attention Networks for Aspect-Level Sentiment Classification. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 4068–4074.
40. He, R.; Lee, W.S.; Ng, H.T.; Dahlmeier, D. Effective attention modeling for aspect-level sentiment classification. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August 2018; pp. 1121–1131.
41. 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), Online, 16–20 November 2020; pp. 3540–3549.