Received 30 March 2023, accepted 23 April 2023, date of publication 27 April 2023, date of current version 5 May 2023.
Digital Object Identifier 10.1109/ACCESS.2023.3270927

RESEARCH ARTICLE

An Adaptive Masked Attention Mechanism to Act on the Local Text in a Global Context for Aspect-Based Sentiment Analysis

TE LIN AND INWHEE JOE
Department of Computer Science, Hanyang University, Seoul 04763, South Korea
Corresponding author: Inwhee Joe (iwjoe@hanyang.ac.kr)

This research was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2022. Project name: Customized tourism through AI-based tourist situation recognition and tourism information curation development of itinerary recommendation platform technology. Project Number: R2022020116.

ABSTRACT Aspect-based sentiment analysis (ABSA) is an important research area in natural language processing, which aims to analyze the sentiment polarity of the aspect terms present in the input sentences. In recent years, many models have focused on local text or local text-aspect relations by designing models that act directly on the local text and then fusing features of the global text. In fact, this ignores the role of the global text. This paper first proposes a masked attention mechanism that acts on the local embedding part of the global embedding, based on the global attention mechanism. Previous models use two methods, called Context-features Dynamic Mask (CDM) and Context-features Dynamic Weighted (CDW), to assign weights to text vectors based on the distance to the aspect term, these methods avoid information redundancy. In this paper, the proposed method uses this masked attention mechanism to intercept the local embedding in the global embedding and then calculate the position in the dimension of the aspect term, reorder the weights corresponding to the position, and assign them to the global embedding according to the corresponding subscripts, in this way, the proposed model not only takes into account noise reduction but can also pay more attention to the feature information of the global text. Compared with the previous embedding using two pre-training models for local and global text, the model proposed in this paper can learn features of both global and local text with only one pre-training model, so it can also improve the training efficiency of the model. The proposed model achieves good results on a total of eight datasets, including the triple-classified and quadruple-classified datasets of laptops and restaurants in SemEval2014, the restaurant dataset in SemEval2016, and the Multi-Aspect Multi-Sentiment (MAMS) dataset.

INDEX TERMS Aspect-based sentiment analysis, global context focus, masked attention, attention based context-featured dynamic inattention-based.

I. INTRODUCTION

With the recent development of e-commerce and tourism in these years, user reviews of a place or product play an increasingly important role, and companies or businesses can analyze user reviews to make informed decisions about the company's products to better recommend or experience them to consumers. Aspect-Based Sentiment Analysis (ABSA) [1], [2], [3] captures the sentiment of a particular entity by analyzing the sentiment polarity of the entity in a sentence to determine the consumer's attitude towards the entity in the current scenario.

ABSA tasks are divided into single ABSA tasks and compound ABSA tasks. Single ABSA tasks, which are intended to identify, extract or analyze only one sentiment element from the text, correspond to four tasks called Aspect Term Extraction (ATE), Aspect Category Detection (ACD), Opinion Term Extraction (OTE), and Aspect Sentiment Classification (ASC). While compound ABSA tasks can be considered as a composite task of single ABSA tasks, the goal of these

The associate editor coordinating the review of this manuscript and approving it for publication was Chuan Li.

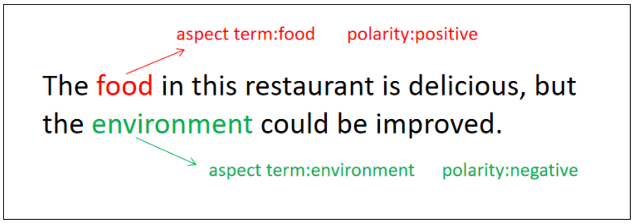

FIGURE 1. The arrows mark two different aspect terms in the same sentence and their emotional polarity.

composite tasks is not only to extract multiple sentiment elements, but also to couple them by predicting elements in pairs (two elements), triples (three elements), or even quadruples (four elements). And to improve the ASC task in single ABSA tasks, the model in this paper is proposed.

Aspect sentiment classification (ASC) [4], a subtask of aspect-based sentiment analysis (ABSA), is a fine-grained sentiment analysis task. It aims to identify the sentiment polarity of a given aspect in a sentence. A sentence may contain several different aspects. Each of these aspects may have a different sentiment polarity. For example, in Figure 1, ASC is to detect the sentiment polarity of the aspect term in the current sentence, if the user is reviewing the restaurant, and in the review, the user thinks that the food of the restaurant is good, but the environment of the restaurant is poor.

Since deep learning was applied to ABSA tasks and verified that deep learning can achieve good results on ASC tasks, subsequent studies started to adopt more deep learning models, such as LSTM [5], and TextCNN [6], etc. for the ASC task, which could effectively extract sequence or local information in sentences. With the birth of the attention model [7], the attention model became mainstream for ASC tasks. Then came the LSTM to extract sequence information and then connect attention to do weight calculations to optimize the task, but these traditional deep learning methods are based on using Word2Vec or Glove and other pre-trained word vectors to do word vector embedding. With the birth of large-scale pre-training models, such as BERT [8], Roberta [9], GPT [10], XLNet [11], etc., they use a large amount of data for learning, making pre-training models like this one equipped with rich textual representations and greatly improving the effectiveness of ASC task.

In 2018, Zeng et al. [12] proposed Context-features Dynamic Mask (CDM) and Context-features Dynamic Weighted (CDW) models to prevent redundant information and reduce noise, and modeled using pre-trained models that fused global text with added aspectual term features and local text features processed by CDM or CDW. Since then, various types of studies have been conducted around the relationship between global text and local text with good results, but their main operations on the model are focused on the processing of local text or only consider the distant and close relationship between token and aspect term in CDM or CDW. The model proposed in this paper improves on the weight assignment methods in the CDW and CDM models and proposes a novel method for fusing local text and global text features. The ability of global text features to learn aspect terms better than local text features has been demonstrated in previous studies [36], so in this paper, the proposed method focuses its work on the global text, and to obtain information about the local text, using a new masked attention mechanism. CDM or CDW assigns weights based on distance only, without considering the contextual relationship between the token and the contextual relationship between aspect terms, which lacks interpretability. AMA-GLCF model is also using the newly proposed attention model that works on the local text and reorders the weights calculated by the dimensionality of the aspect term and then assigns the weights. The proposed approach uses the basic attention structure for the local text in the global text, which saves the extra memory space needed to model the two parts separately and to learn information about the aspect term. When calculating the attention score, the weights of the dimensions corresponding to the aspect terms in the intercepted local text are re-ranked, and the distance weights are assigned according to the result of the ranking, which solves the problems of interpretability of weight assignment and missing information caused by setting thresholds. The proposed method in this paper achieves good results on the triple-classified and quadruple-classified datasets in SemEval-2014 [14], the Multi-Aspect Multi-Sentiment dataset [13], the restaurant dataset in SemEval-2016 [15], and two balanced samples processed according to under-sampling.

II. RELATED WORK

This section is divided into two parts. The first part is a summary of the models that have been used in ASC tasks in the past few years and of the recently popularized context focus mechanism, including the CDW and CDM methods in it.

A. METHOD IN ASC TASK

In the ASC task, the initial use of machine learning methods [16], similar to SVM [17] and other methods, achieved good results. However, it was too tedious for manual feature engineering and would waste a lot of time and resources. Therefore, in recent years, the use of deep learning or large-scale pre-trained models such as BERT [8], XLNet [11], Roberta [9], etc., which has just given a dramatic shock to the field of natural language processing, has become mainstream in ABSA tasks.

When using deep learning methods, pre-trained word embedding models such as Word2Vec [18] and Glove [19] are the methods that are used by most of the models to capture the syntactic and semantic features of the text. Convolutional neural networks have been more widely used in sentence-level sentiment analysis, but less widely used in fine-grained sentiment analysis tasks, in the deep learning approach. Wang et al. proposed a new model: the UP-CNN [20], which has an aspect detection network with prior knowledge and uses an aspect mask to build aspect-based contextual representations. Zhao et al. proposed a new model structure based on the idea that the sentiment polarity of sentences has a relatively significant correlation with the target aspect. It combines a convolutional neural network (CNN) and a gated recursive unit (GRU) [21]. The GCAE [22] model uses a gating mechanism to filter the aspect information in the information captured by two convolutional layers. It then performs sentiment analysis. The TD-LSTM [23] is used to make the model more focused on the content of the aspect term, which uses two LSTMs to obtain the relationship between the aspect term and the preceding and following text, and then extracts the information of the aspect term. The ATAE-LSTM [24] combines attention and LSTM and uses attention to obtain more important contextual information for different aspects to solve the aspect-level sentiment analysis problem. Ma et al. proposed a new attention network IAN [25], which uses two attention networks to model the aspect term and the context separately in an interactive way and was designed to address the problem of information loss when more than one aspect term appears in a sentence. Huang et al. [26] proposed an attention-over-attention (AOA) model that can automatically focus on the important parts of a sentence by jointly modeling and jointly learning the representation of sentences and aspect terms. Tang et al. [27] introduced deep memory networks to the sentiment classification task, followed by Chen et al. who proposed the recurrent attention network (RAM) by fusing recurrent neural networks and weighting mechanisms with multiple attention mechanisms [28] to further improve the classification effect of memory networks. Lin et al. [29] proposed a deep masked memory network, and the deep memory network added semantic information of aspects and inter-aspect relationship information, which allowed the network to learn the information of aspect terms more effectively. In recent years, with the popularity of graphical neural networks, graphical neural network methods have been applied to ASC tasks, where Sun et al. [30] and Zhang et al. [31] used graphical neural networks (GNN) [32] to model dependency trees to exploit syntactic information and word dependencies.

However, word vector models such as Word2Vec and Glove are limited by the length of the input text and cannot learn the semantic information of words in context more effectively. On the other hand, large-scale pre-training models can learn the semantic information of the context more effectively. As a result, in recent years, pre-training models such as BERT have been widely used for ASC tasks. BERT-SPC uses the addition of aspect terms to text to perform sentiment analysis on the stitched text, in addition to the direct sentiment analysis of text using BERT (BERT-FC). A very typical BERT-based classification model for ASC tasks was proposed in 2018: TD-BERT [33], which achieved very good results by directly classifying the features of the aspect term. Gao et al. proposed domain-adapted BERT (BERT-ADA) [33] for laptop and restaurant datasets. It achieved good performance. Later, BERT-PT [35] transformed the task of sentiment classification into a task of reading comprehension to perform the classification. Since then, the use of BERT to model text and aspect terms separately and then process the modeled content has emerged. For example, AEN-BERT [36] designed a lightweight multi-head self-attention to process and then classify the separately modeled content. Including the model AM-Weight-Bert [37], which was published just this year, which sets a threshold to filter the weights of the separately modeled contents after the attention calculation and then classify them.

B. CONTEXT FOCUS MECHANISM

Zeng et al. [12] proposed the concept of a local context focus mechanism in 2018, and they found that the relevance of each token in the text is gradually reduced according to the distance of the aspect term, so they proposed two kinds of context dynamical weighting and context dynamical masking mechanisms to attenuate the importance of the token far from the aspect term, and thus reduce the influence of noise on the model, and the experimental proved the effectiveness of this method. Based on this, Phan et al. [38] proposed a Local Context Focus on Syntax - ASC (LCFS-ASC) model based on syntactic structure and syntactic tree, which adds syntactic structure information to the text. It differs from LCF in that the distance between words used in performing CDM/CDW is no longer calculated by position, but by the distance between two words, The distance between words in the syntactic parse tree is no longer calculated by position, but by the distance between two words in the syntactic parse tree. It and LCF both introduce the multi-head self-attention (MHSA) [39] mechanism to capture global dependencies more accurately. Then, to capture the scope of the local context more comprehensively and exploit the non-equivalence of dependency relations, Xu et al. proposed the DFLCA-DCA [40] model. It can dynamically capture the scope of local contexts based on different maximum distances from the target aspect term to the context token. DCA allows the model to focus more on clustering. These models ignore the role that global text can play and focus more on the relationship between local text and aspect terms. In this year's proposed LGCF model [41], by using CNN and gating mechanism to fully capture the features in the global text, and combining the features processed by CDM/CDW on the local text, the model makes full use of the feature content of the global text and achieves good results.

To provide a more visual representation of the recently proposed models in the ABSA task, table 1 is presented. The four columns in the table represent the way text is modeled, the model name, the main method of the model, and the part of text modeled, where L denotes local text, G denotes global text, A denotes aspect term, MHSA denotes multi-head self-attention [39], and the last column is whether CDW/CDM distance weighting method is used. As shown in Table 1, most of the models use attention or attention-based models. The models with better recent results incorporate information

TABLE 1. Related Works.

Text Representation	Model	Main Method	Parts of Modeling	Using CDW/M Method
Embedding	UP-CNN	CNN/prior knowledge, aspect mask	L	–
	GCAE	CNN/gating mechanism	L	–
	TD-LSTM	RNN, extracts the information of the aspect term	L	–
	ATAE-LSTM	Combined aspect information, Attention, LSTM	L+A	–
	IAN	attention, interactive modeling	L+A	–
	AOA	joint modeling, joint learning	L	–
	RAM	multiple attention mechanisms, memory network	L	–
	MemNet	weighted attention mechanism	L	–
	GNN	use GNN to obtain syntactic tree information	L	–
BERT c	BERT-FC	BERT, fully connected layer	L	–
	BERT-SPC	combine information of aspect term	G	–
	BERT-ADA	add domain-specific knowledge base	L	–
	AEN-BERT	lightweight Multi-Head Attention, Attention	L	–
	AM-Weight-BERT	Attention/filtering information by threshold	L+A	–
	BERT-LGCF	MHSA, combine CNN and gating mechanism.	L+G	✓
	DFLCA-DCA	MHSA, dynamic capture of local contexts	L+G	✓
	BERT-LCF	MHSA, Separate modeling, CDW/M	L+G	✓
	BERT-LCFS	MHSA, CDW/M, add syntactic tree information	L+G	✓

from the both local and global text and use CDW/M for distance weighting. Based on these elements, the AMA-GLCF model is an improvement and innovation.

III. INTRODUCTION TO MODEL METHODS

Aspect-based sentiment analysis, as shown in Figure 1, is introduced in this section. And the Adaptive Masked Attention Mechanism Action of Local Text on Global Text (AMA-GLCF) which designed for this task, as shown in Figure 5. We will also propose an improved structure based on Context-features Dynamic Weighted (CDW) and Context-features Dynamic Mask (CDM), which is named Adaptive Masked Attention-based Context-features Dynamic Weighting (AMA-CDW), as shown in Figure 4 and Figure 6.

A. TASK DEFINITION

The purpose of aspect-based sentiment classification is to analyze the sentiment polarity of the specified entity in the input sentence, given the input sentence S:

$$S = \{w_0, w_1, \ldots w_i \ldots, w_n\}. \tag{1}$$

S consists of n words, which contain m words indicating the aspect term A for which need to perform the analysis.

$$A = \{w_i, w_{i+1}, \ldots, w_{i+m-1}\}. \tag{2}$$

Then our task is to analyze the sentiment polarity P of A in S. For the triple-classified task:

$$P_1 = \{Positive, Negative, Neutral\}. \tag{3}$$

For the four classification tasks:

$$P_2 = \{Positive, Negative, Neutral, Conflict\}. \tag{4}$$

For example, given the sentence in Figure 1, analyze the sentiment polarity of food and environment as positive and negative, respectively.

B. INPUT EMBEDDING LAYER

In the AMA-GLCF model, using a pre-trained BERT model to map the high-dimensional vectors to our input sentences, as introduced in Section II, to improve the performance of the task. In previous tasks for context, features focus, using BERT for local text ($[CLS] + sentence + [SEP]$) and global text ($[CLS] + sentence + [SEP] + aspect\ term + [SEP]$). But in the AMA-GLCF model, global text can learn aspect term information better than local text, only need to model global text once and process it later. As shown in the following two equations, in the obtained global context features E, reduce the length of aspect term greater than 1 in the global context to 1 by the *MaxPooling* method and then fill 0 at the end to facilitate later processing. Since the additional length is equal to the length reduced by the *MaxPooling* process, the complementary length is equal to the length of the aspect term. Finally, the resultant E' is transmitted to the network.

As shown in Figure 2:

$$E = \{e_0, e_1, \ldots, e_{a1}, \ldots, e_{a_i}, \ldots e_n\}. \tag{5}$$

$$E' = \{e'_0, e'_1, \ldots, e'_a, \ldots e'_n, \underbrace{0, \ldots, 0}_{i-1}\}. \tag{6}$$

C. ADAPTIVE MASKED ATTENTION MECHANISM ACTION OF LOCAL TEXT ON GLOBAL TEXT

This section will mainly introduce the AMA model method and the AMA-CDW method using the computed and ranked AMA model results, as shown in Figure 4 and Figure 6.

1) ADAPTIVE MASKED ATTENTION

The most basic attention method as shown in the following equation is used in the AMA-GLCF model to compute the input of sentence(global context) contextualized embedding.

$$att = softmax(\frac{Q \cdot E^T}{\sqrt{d}}). \tag{7}$$

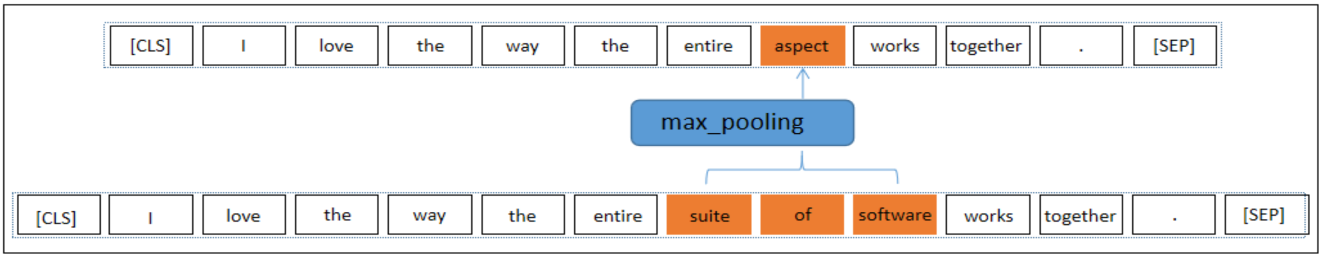

FIGURE 2. The process of context features when the length of aspect term is not 1.

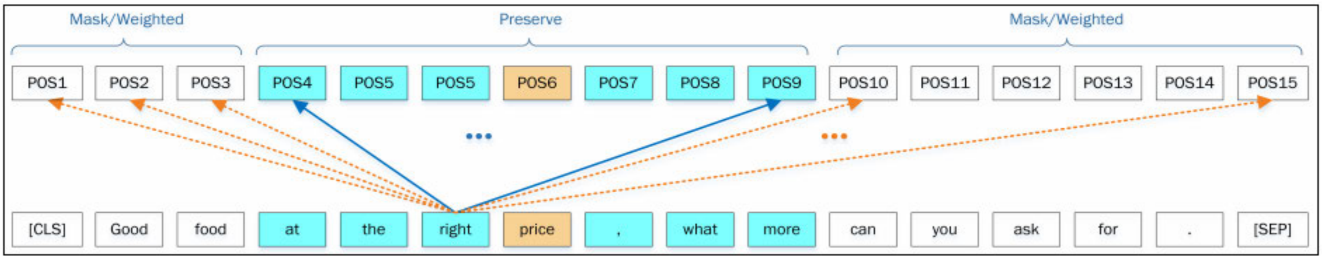

FIGURE 3. Simulation of context-features Dynamic Weighting(CDW) mechanism and context-features Dynamic Mask(CDM) mechanism.

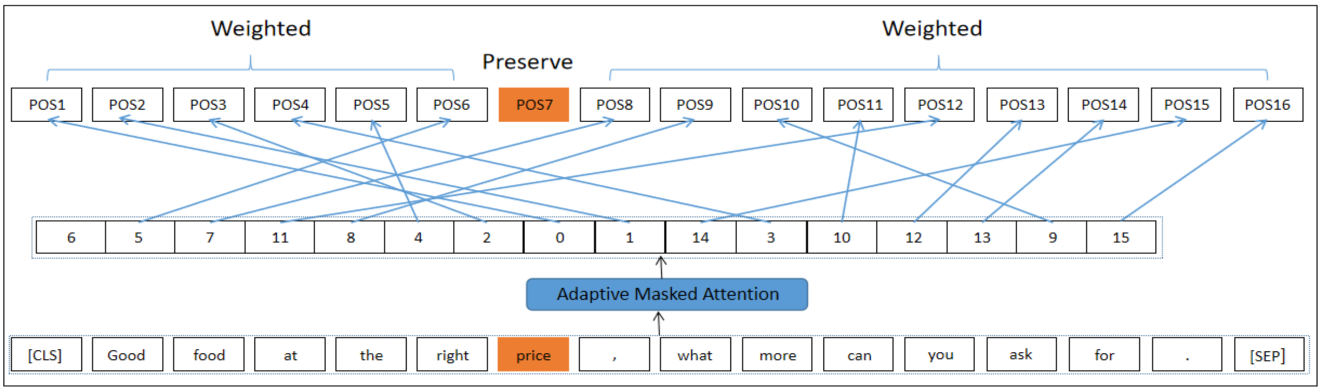

FIGURE 4. Simulation of adaptive masked attention based context-features Dynamic Weighting(AMA-CDW)(When the length of aspect term is 1).

The local contextualized embedding part in the AMA-GLCF model, where Q is obtained by global embedding through a linear layer, and E stands for global embedding because the global context can learn more information about the aspect term compared to the local context, as shown in the argument [33] and other experiments, BERT-SPC works better than BERT-BASE. The local embedding part can be truncated in the global embedding by saving the local text length and the aspect term length during data processing. $E_l = \{e_1, e_2, \ldots, e_n\}$, n represents the length of local text, and by equation (2), m represents the length of aspect term, because, in the previous subsection, using the *MaxPooling* method to reduce the length of aspect to 1, so here using the calculation of the basic attention model to act on the length of local text as n − m + 1. So the processed local embedding is $E'_l = \{e'_1, e'_2, \ldots, e'_{n-m-1}\}$., we process Q in the same way to obtain Q'. The attention score of the processed local embedding is calculated using the attention mechanism:

$$scores = \frac{Q' \cdot (E'_l)^T}{\sqrt{d}}. \tag{8}$$

The alpha is the result of the softmax calculation:

$$\alpha_{ij} = \frac{\exp scores_{ij}}{\sum_{z=1}^{n-m-1} \exp scores_{iz}}. \tag{9}$$

Since the last aspect term part of the global text after BERT modeling is redundant in the attention computation, the attention mechanism is used to process only the local embedding part of the global embedding, after obtaining the alpha value, its dimension is [n-m-1, n-m-1]. To compute with the original global embedding, we need to add 0 after it. So its dimension becomes [n, n]. The n in the first dimension represents the current sequence length. The n in the second dimension represents the weight of each

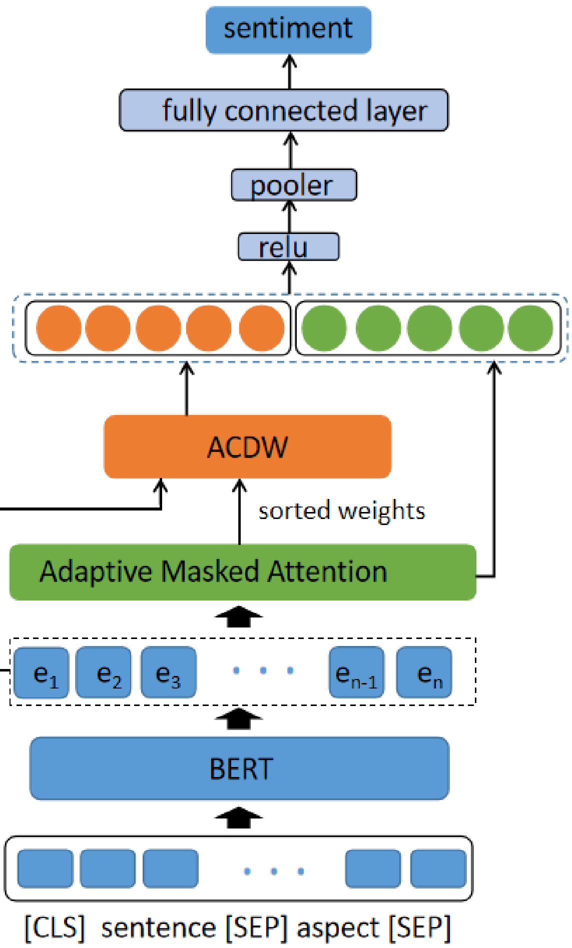

FIGURE 5. Overall network structure of AMA-GLCF.

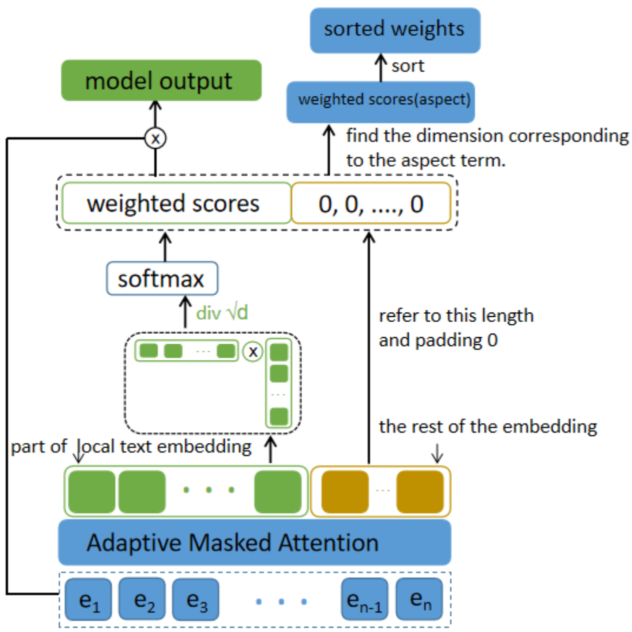

FIGURE 6. Simulation of Adaptive Masked Attention.

token assigned to the n tokens. So the vector of n in the first dimension that corresponds to the second dimension of the aspect vector, and sort the values in it from largest to smallest (removing the take one value assigned to the aspect term, since it will be set to 1 when the ACDW in the next subsection):

$$sorted\_\alpha_{ai} = \{\alpha_{ai_1}, \alpha_{ai_2}, \ldots, \alpha_{ai_{n-1}}\}. \tag{10}$$

$$sorted\_index = \{I_{\alpha_{ak_1}}, I_{\alpha_{ak_2}}, \ldots, I_{\alpha_{ak_{n-1}}}\}. \tag{11}$$

where $\alpha_{ak}(\alpha_{ak} \in sorted\_\alpha_{ai})$ denotes the k-th position on the aspect term dimension in the computed alpha obtained. We matrix multiply the computed alpha with the original global embedding to obtain the final result of the proposed new attention mechanism:

$$att\_out = \alpha \times E. \tag{12}$$

As shown in Figure 6, two results are obtained by the proposed new attention mechanism, one is the *sorted_index* and the other is the *att_out* that we obtained after calculation. *sorted_index* is used in the next subsection to compute the newly proposed ACDW module.

2) ATTENTION-BASED CONTEXT-FEATURE DYNAMIC WEIGHTING

LCF [12] A new CDW and CDM structure is proposed to notice the correlation between the emotional polarity of the aspect term and the local context, which is well demonstrated in papers such as LCF [12], and LCFS [38]. The results of experiments such as LCF, LCFS, and other papers [40], [41] fully show that the effect of CDW is better than that of CDM, so the proposed method in this paper focuses on improving CDW, and LCF by assigning weights to local text features according to the distance of token features for aspect term in turn, and assigning weights to tokens within a certain distance from the central word is 1, which is called Semantic Relative Distance (SRD), as shown in Figure 3.

$$SRD_i = |i - P_a| - |\frac{m}{2}|. \tag{13}$$

where i(1 < i < n) is the token position, Pa is the aspect term center, m is the aspect term length, and SRDi is the distance between the ith token and the aspect term.

The assigning weights based on distance alone is unjustified and does not fully account for the semantic relationship between the aspect term and the context, although it does account for the relationship between the aspect term and the local context. In the previous subsection, using the newly proposed attention mechanism to compute the weights assigned to other token features in this dimension of the aspect term, as shown in Figure 4, where we use the sorted weights and obtain the importance of the token relative to the aspect term according to the sorted index order, preserving the subscript. Each token feature in the global embedding is assigned a weight based on its distance to the aspect term. Because the weights are computed by the attention mechanism relative to the aspect term, SRD is not required.

$$V_i = \begin{cases} E & E_i' = aspect\ term, \\ \dfrac{n-i}{n} \cdot E & E_i' \neq aspect\ term. \end{cases} \tag{14}$$

$$M = \{V_0, V_1, \ldots, V_n\}. \tag{15}$$

$$O_{ACDW} = E' \cdot M. \tag{16}$$

where i($i \in sorted\_index$) denotes the subscript, E denotes the unit matrix, and E_i' in the condition denotes the i-th token feature in the global embedding. Equation 15 denotes the weights obtained after our computation, and E' in equation 16 denotes the E' in equation 6, which is the global embedding features after processing them by *MaxPooling*.

3) FEATURE CONCATENATION AND OUTPUT LAYER

As shown in Figure 5, the output of the Adaptive Masked Attention model *att_out* introduced in Section I) is spliced with the embedding features with assigned weights obtained after ACDW. The spliced results are passed through the *Relu* activation function and a *pooler* layer. The final sentiment classification result is obtained by the final computation of the full connectivity layer.

$$O_{concat} = [O_{ACDW}; att\_out]. \tag{17}$$

$$O_{Relu} = Relu(O_{concat}). \tag{18}$$

$$X_{pool} = POOLER(O_{Relu}). \tag{19}$$

$$Y = Softmax(X_{pool}). \tag{20}$$

where Y is the predicted sentiment classification result of the AMA-GLCF model.

IV. EXPERIMENTS

In this section, we will present the dataset used, as well as the experimental parameters set up, the experimental environment, the experimental results obtained, etc. The comparison models will also be part of this section.

A. LOSS FUNCTION

In experiments, using a cross-entropy loss function to calculate the error.

$$Loss = -\sum_{i=1}^{M} y_i \log \hat{y}_i. \tag{21}$$

B. DATASETS

A relatively large number of datasets were used in the experiments. We used SemEval-2014 Task 4, the restaurant dataset from SemEval-2016, and the Multi-Aspect Multi-Sentiment (MAMS) dataset. SemEval-2014 contains two datasets - laptop and restaurant - which contain a triple-classified dataset and a quadruple-classified dataset, respectively. Removing the category 'conflict' from the quadruple-classified dataset is a triple-classified dataset because the number of these instances is so small that keeping them in the training data would make the dataset very unbalanced, but to validate the validity of the AMA-GLCF model, also on the four unbalanced categories, The restaurant dataset from SemEval-2016 is a small number and can also fully represent the validity of the AMA-GLCF model with a small number of data. MAMS includes an ATSA dataset and an ACSA dataset, both of which are from restaurant reviews, and we use one of the ATSA datasets for our task. These datasets have data imbalance problems. This is true even for the triple-classified dataset with the 'conflict' category removed. Therefore, based on the idea of under-sampling, you should avoid the data imbalance problem by reducing the amount of data in other categories to the same amount as the category with the least amount of data. Table 2 shows the statistical results for all the triple-classified datasets, Table 3 shows the statistical results for the two quad-categorization datasets from SemEval2014 Task IV, and Table 4 shows the statistical results for the processed datasets where there are balanced samples. All three tables show the number of training, validation, and test samples for each dataset, as well as the classification of the labels.

TABLE 2. Statistics of the triple-classified experimental dataset.

Dataset	Positive	Negative	Neutral	Total
Laptop14-Train	987	870	464	2328
Laptop14-Test	341	128	169	638
Restaurant14-Train	2164	807	637	3608
Restaurant14-Test	728	196	196	1120
Restaurant16-Train	1240	437	69	1746
Restaurant16-Test	468	117	30	615
MAMS-Train	3380	2764	5042	11186
MAMS-Validation	403	325	604	1332
MAMS-Test	400	329	607	1335

TABLE 3. Statistics of the quadruple-classified experimental dataset.

Dataset	Positive	Negative	Neutral	Conflict	Total
Laptop14-Train	987	866	460	45	2358
Laptop14-Test	341	128	169	16	654
Restaurant14-Train	2164	805	633	91	3693
Restaurant14-Test	728	196	196	14	1134

TABLE 4. Statistics of the sample balanced dataset.

Dataset	Positive	Negative	Neutral	Total
Laptop14-Train	464	464	464	1392
Laptop14-Test	128	128	128	384
Restaurant14-Train	637	637	637	1911
Restaurant14-Test	196	196	196	688

C. EXPERIMENT SETTING

The PyTorch framework was used to build all of the AMA-GLCF model. Our experiments are tuned using the pre-trained model bert-uncased-base provided by the transformers library. The model was trained on the NVIDIA GeForce RTX 2080 Ti GPU. The Adam optimizer was used to tune the model. See Table 5 for other parameters in the experiments.

D. COMPARED MODELS

The compared models are modeled by word vectors, such as Word2Vec, Glove, etc., or pre-trained models BERT to

TABLE 5. Hyperparameters used in the experiment.

Parameter	Value
Dropout rate	0.1
Batch size	32
learning rate	2e-5
Max epoch	10
Max sequence length	128
Optimizer	Adam

model the text since the proposed AMA-GLCF model in this paper is based on BERT to model the text and assign weights to the local embedding part after modeling. Some comparison models also use CDM/CDW and methods that fuse the local text and global text information. Some popular baseline models of the last years or state-of-the-art models are compared in this subsection. F1 macro scores and accuracy rates are used as evaluation metrics.

IAN [25]: The text and the aspect term are modeled separately using LSTM. The semantic information of the text is then captured and learned using an interactive attention mechanism.

MemNet [27]: An end-to-end network that classifies the sentiment polarity of an aspect term using multiple attention layers to capture the relationship between features.

RAM [28]: The MemNet model is improved by fusing recurrent neural networks and weighting mechanisms, and combining multiple attention mechanisms to capture the relationship between distant features and reduce interference from irrelevant information.

BERT-FC: The text is modeled using BERT and the [CLS] content representing the whole sentence information is fed to the fully connected layer for classification and the same result is obtained for any token with sentiment polarity. Thus, this model is the most basic method when the BERT model is in use.

BERT-SPC: The global text ($[CLS] + text + [SEP] + aspect\ term + [SEP]$.) is modeled using BERT. The modeled content is fed to the fully connected layer for sentiment classification.

BERT-AOA [26]: A new Attention-over-Attention mechanism is used to more fully learn both the aspect term and the modeled text representation after the text has been modeled by the pre-training model BERT.

BERT-PT [35]: The text is modeled by BERT. Then, the sentiment classification problem is transformed into a reading comprehension problem to determine the sentiment polarity of the aspect term by reading comprehension.

AEN-BERT [36]: An attention coding network is proposed to model the text and aspect terms, and then a pre-training model BERT is combined. Label smoothing and regularization methods are used for classification.

LCF-BERT-CDM [12]: The local text and the global text are modeled separately using BERT, and a new weight assignment method, Context Dynamic Masking (CDM), is applied to the modeled local embedding to set the weight of a part of the modeled text to 0 for noise relative to the aspect term. As shown in Figure 3, the processed local embedding and global embedding are then merged and fed into the multi-head self-attention attention model. The results are then post-connected to the fully connected layer for classification.

LCF-BERT-CDW [12]: Similar to LCF-BERT-CDM, instead of setting the weights assigned to the FEATURES that are far away from the aspect terms to 0, different weights are assigned in turn according to the distance.

LCFS-BERT-CDM [38]: Similar to LCF-BERT-CDM, LCFS-BERT-CDM adds syntactic information embedding by introducing syntactic trees and then proposes syntactic relative distance to reduce the negative influence of irrelevant words with weak syntactic connections. This allows the content modeled by BERT to better combine sentences and attribute words and to better learn the semantic features of the context.

LCFS-BERT-CDW [38]: The weight assignment for local text embedding is changed based on LCFS-BERT-CDM.

AM-Weight-BERT [37] The text and the aspect term are modeled separately using BERT, and the two parts of the modeled content are spliced and then fed into the attention model. A threshold γ is set when calculating the attention score, and the score is set to 0 if the score is below the threshold, and the score is retained if it's above the threshold. After this attention structure, the original text vector is spliced and then classified, and the sentiment polarity of the aspect term is obtained.

E. RESULTS AND DISCUSSION

We conducted experiments with the newly proposed model on four triple-classified datasets, two quadruple-classified datasets, and two datasets with balanced samples, as shown in Tables 6, 7, and 8. It is possible to obtain high accuracy scores on the datasets with unbalanced samples, even if the model performance is poor, due to the difference in sample sizes. At this point, although the accuracy scores are high, they are not very meaningful. Therefore, when the data are unbalanced, the shortcomings of the accuracy evaluation method become obvious. Therefore, recall, accuracy, F1 score, and other evaluation methods should be introduced. However, since accuracy and recall are a pair of contradictory quantities when accuracy is high, recall tends to be relatively low, and when the recall is high, accuracy tends to be relatively low, so to better evaluate classifiers, F1-score is generally used as an evaluation criterion to measure the comprehensive performance of classifiers. In F1-score, the micro F1-score considers the number of samples from different classes, which is more likely to be influenced by common classes in the case of unbalanced samples. In contrast, the macro F1-score does not consider the number of samples from different classes, which is relatively easier to consider the influence of rare classes. Therefore, in the unbalanced sample, the macro F1-score is used as the only evaluation metric, while in the balanced sample, the accuracy evaluation metric is added. To demonstrate the robustness of the newly proposed model, five experiments were conducted. The average value was taken as the final result. Compared to the baseline model in

TABLE 6. These are the experimental results for all triple-classified datasets. In this experiment, "–" indicates that it did not appear from the previous experiments, and "*" indicates that we redid the experiments according to all hyperparameters provided in the previous paper to get the results. We run the program 5 times with random initialization and show "mean±std" as its performance. We have bolded the font of the best results. The datasets are derived from these papers: [33], [37], [41].

Model	Laptop14-3way	Restaurant14-3way	Restaurant16-3way	MAMS
	Macro-F1	Macro-F1	Macro-F1	Macro-F1
RAM	71.35	70.8	72.19	–
IAN	67.38	70.09	55.21	–
MemNet	64.09	65.83	65.99	–
BERT-FC	72.83±0.99	69.79±1.41	–	–
BERT-SPC	75.58±1.12	77.84±1.31	–	81.57±0.36
BERT-AOA	73.53±1.25	69.64±1.04	66.21	–
BERT-PT	75.08	76.96	–	–
AEN-BERT	73.86±1.19	71.73±1.12	69.58	–
TD-BERT	74.38±0.81	78.35±1.34	–	–
LCF-BERT-CDM*	75.77±1.77	78.61±1.12	73.56±3.41	82.80±0.10
LCF-BERT-CDW*	75.79±0.62	78.94±0.82	74.31±2.99	82.85±0.14
LCFS-BERT-CDM	75.35	78.99	74.49	82.77±0.15*
LCFS-BERT-CDW	75.35	78.16	70.48	82.91±0.34*
AM-Weight-BERT	76.20±0.71	78.92±0.97	–	82.56±0.59
OURS	**76.78±0.63**	**79.33±0.61**	**77.08±0.91**	**83.33±0.16**

TABLE 7. These are the experimental results for all four classification datasets. In this experiment, "–" indicates that it did not appear from the previous experiments, and "*" indicates that we redid the experiments according to all hyperparameters provided in the previous paper to get the results. We run the program 5 times with random initialization and show "mean±std" as its performance. We have bolded the font of the best results. The datasets are derived from these papers: [33].

Model	Laptop14-4way	Restaurant14-4way
	Macro-F1	Macro-F1
MemNet	–	–
RAM	–	–
IAN	–	–
BERT-FC	56.95±4.21	58.93±2.77
BERT-AOA	54.37±1.32	59.92±1.71
BERT-PT	–	–
AEN-BERT	–	–
TD-BERT	54.37±0.82	58.29±1.41
LCF-BERT-CDM*	56.84±1.21	58.44±2.28
LCF-BERT-CDW*	57.51±0.57	58.78±1.68
LCFS-BERT-CDM*	57.86±0.64	59.03±1.55
LCFS-BERT-CDW*	58.81±0.89	59.21±2.89
AM-Weight-BERT	–	–
OURS	**60.29±0.96**	**60.06±3.08**

TABLE 8. These are the experimental results for the balanced processed sample datasets. In this experiment, "*" indicates that we redid the experiments according to all hyperparameters provided in the previous paper to get the results. We run the program 5 times with random initialization and show "mean±std" as its performance. We have bolded the font of the best results.

Model	Laptop14-3way(balance		Restaurant14-3way(balance)	
	Accuracy	Macro-F1	Accuracy	Macro-F1
LCF-BERT-CDM*	79.91±0.43	79.69±0.52	80.20±0.42	80.03±0.63
LCF-BERT-CDW*	**80.05±0.45**	**79.85±0.29**	80.13±0.67	79.97±0.82
LCFS-BERT-CDM*	79.88±0.37	79.56±0.52	80.34±0.53	80.19±0.47
LCFS-BERT-CDW*	79.93±0.62	79.66±0.71	80.41±0.36	80.25±0.52
OURS	80.00±0.45	79.81±0.57	**80.59±0.43**	**80.39±0.35**

recent years, although we used only a simple attention structure and improved the weighting, the AMA-GLCF model still achieved good results for all datasets. By analyzing the datasets, we found that the advantage of the AMA-GLCF model over the LCF/LCFS model increases as the number of labeled categories in the dataset decreases. The AMA-GLCF model performs better in the quadruple-classified dataset than in the triple-classified dataset, with macro F1 scores essentially 1% to 2% higher. However, the experimental results show that the AMA-GLCF model did not achieve the best results in the sample-balanced dataset. We believe that this is due to the adoption of the idea of under-sampling, which led to the reduction of the dataset to a small size and resulted in the gap between the models not being very large, the difference is normal and acceptable when the training sample is not sufficient, however, we believe that the results of such a balanced sample are relevant when considering the use of accuracy and macro F1-score as evaluation metrics. And we think that although the weighting method has been improved, there is still the problem of information redundancy in weighting

TABLE 9. Comparison of the three weight assignment methods on the laptop dataset.

Method	Laptop14-3way Macro-F1	Restaurant14-3way Macro-F1
OURS	**76.78±0.63**	**79.33±0.61**
with CDM	76.13±0.55 ↓**0.65**	78.78±0.42 ↓**0.55**
with CDW	76.22±0.30 ↓**0.56**	78.86±0.65 ↓**0.47**

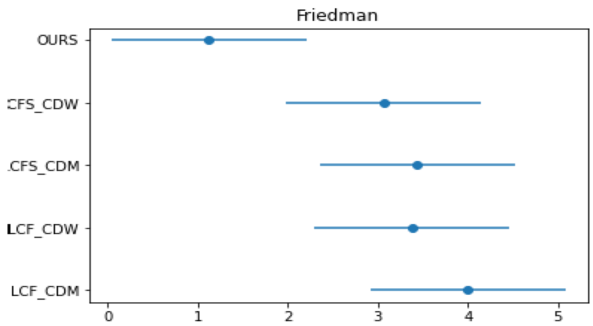

FIGURE 7. Friedman Test chart with CD(critical difference) represented by line segments.

by this method, since there may be nouns that have little relation to aspect words, which are also given some weights.

F. ANALYSIS OF WEIGHT ASSIGNMENT METHODS

In this subsection, we will introduce three methods of assigning weights, two new methods of assigning weights were proposed in the previous paper and they were verified to be very effective indeed, based on this, we propose a method of reassigning weights according to the results of calculating the attention score, in Section III, we present the method in detail, as can be seen from Table 9, the AMA-GLCF model is better than the two previously proposed methods when other conditions are held constant. In addition to the new way of assigning weights, we also propose a new attention mechanism. Comparing Table 9 with Table 6, the AMA-GLCF model still achieves better results compared to the previous LCF and LCFS methods when using the same CDM or CDW weighting method.

G. TEST OF EXPERIMENTAL RESULTS

In this subsection, we will use the Friedman test, Post-hoc Nimenyi test [42] to verify whether there is a significant difference between the measurements of any two models in experiments.

$$\chi_F^2 = \frac{12N}{k(k-1)}\left(\sum_{i=1}^{k} r_i^2 - \frac{k(k+1)^2}{4}\right). \tag{22}$$

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2}. \tag{23}$$

$$CD = q_\alpha\sqrt{\frac{k(k+1)}{6N}}. \tag{24}$$

Friedman's test is a rank-based statistical method that is used to compare whether there is a significant difference in the average performance of multiple models on multiple data sets, and its formula is shown in 22. Where r_i is the average ranking in the i-th algorithm, k is the number of algorithms, N is the number of data sets, assuming equal performance between the models, and the Friedman statistic follows a chi-squared distribution. Friedman's χ_F^2 is less conservative and yields a better statistic, which is distributed according to the F with degrees of freedom k-1 and (k-1)(N-1), as shown in Equation 23. See any textbook on statistics for the table of critical values. If the null hypothesis is rejected, we can perform a post hoc test called the Nemenyi Test [43]. The performance of two classifiers is significantly different if their average ranks differ by at least a critical value (Equation 24). The critical value q_α is based on the Studentized range statistic divided by $\sqrt{2}$. The Friedman statistic is 6.29 when calculated according to equations 22 and 23. In the case of 5 algorithms and 8 data sets, the distribution of F_F is F with 5-1=4 and (5-1)–(8-1)=28 degrees of freedom. The critical value of $F(4,28)$ with α=0.05 is 2.714, therefore, we can reject the null hypothesis and conclude that they differ significantly.

The CD(critical difference) value calculated by Nemenyi is 2.157, with this value, a Friedman test plot containing the CD values of the five algorithms can be obtained, as shown in Figure 7, and it can be seen that the AMA-GLCF model and the starting three models except the LCFS-CDW model have no overlaps, so there is a clear difference between them and the AMA-GLCF model, although there is an overlap with the LCFS-CDW, it is obvious that the AMA-GLCF model is better than the LCFS-CDW by the experiment data and the mean sequence value.

H. EFFICIENCY COMPARISON

The efficiency of the AMA-GLCF model in terms of training time and GPU memory usage is analyzed in this subsection.

We conducted experiments on the SemEval2014 Restaurant dataset, and we used 10 times results from 10 epochs

TABLE 10. Time and GPU Memory Usage statistics of different models.

Model	Time(s)	GPU MEMORY USAGE(Mib)
LCF-BERT-CDM	43.40	9923(90%)
LCF-BERT-CDW	43.03	
LCFS-BERT-CDM	44.15	9997(91%)
LCFS-BERT-CDW	44.04	
OURS	36.79	6227(60%)

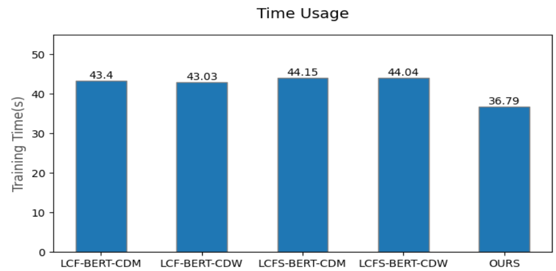

FIGURE 8. Training Time Usage.

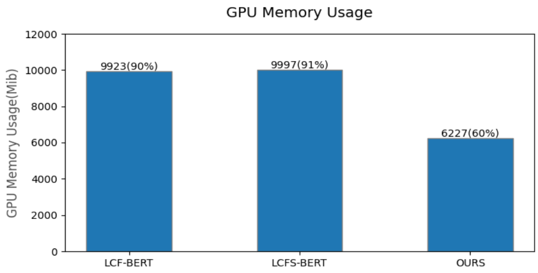

FIGURE 9. **GPU Memory Usage.**

of a training session and averaged them as the final time usage results. For the same model using different weighting methods, the GPU memory usage is the same. Table 10 shows the final results of our statistics. The AMA-GLCF model is the best in terms of training time and GPU memory usage.

V. CONCLUSION AND FUTURE WORK

The main contributions of this paper are shown below:

1) A new attention mechanism that acts on the local text of the global text is proposed, because the global text can better learn the content of the aspect terms, and acting only on the local text can prevent the redundancy of information and reduce the noise (previously added content of the aspect term). This attention mechanism can better calculate the weight of each token between local texts that have more fully learned the aspect term features.
2) According to the proposed new attention mechanism, the weights of each vector computed for the aspect term dimension are reassigned according to the index of the token by reordering them, and we believe that such a weight assignment is more convincing.
3) The AMA-GLCF model uses only one pre-training model compared to the previously proposed model, which improves resource utilization and reduces the model training time.

Most studies today focus on the relationship between local text and aspect terms. The role of global text has not received much attention. Even if the global text is used, it is modeled once by using BERT again. This is a great waste of training resources and efficiency. If the global text is used, this requires remodeling using BERT. Alternatively, the proposed method in this paper uses distance to assign weights to each token, which is effective in previous work. However, in our opinion, this type of weighting needs some explanation. A new model is proposed to solve these two problems: AMA-GLCF uses only the simplest attention structure and improves the previous weighting method by intercepting the local text in the global text. It can make full use of the global text and also combine the information of the local text. In the process of weight assignment, the weights are assigned according to the weights calculated by the corresponding dimension in the attention score of the aspect term. Thus, the weight assignment can be reasonably interpreted. In addition, since the structure of AMA-GLCF is simpler compared to the comparison model, and it only needs to model once to learn the information of local text and global text, it reduces the training time and GPU memory usage. However, even if using this new weight assignment method, there will still be content that is not very relevant to the aspect term to be weighted. But if using the full mask method, we are afraid that it will cause a loss of information. It is worth discussing how to reasonably assign weights without causing excessive information redundancy, or how to combine this with the masked attention method without causing information loss when the weighting method is used together with the attention model. In our future research, we will focus on this aspect.

REFERENCES

[1] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
[2] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 8, no. 4, 2018, Art. no. e1253.
[3] X. Li, L. Bing, P. Li, and W. Lam, "A unified model for opinion target extraction and target sentiment prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 6714–6721.
[4] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A survey on aspect-based sentiment analysis: Tasks, methods, and challenges," 2022, *arXiv:2203.01054*.
[5] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.
[6] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. EMNLP*, 2014, pp. 1746–1751.
[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
[9] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv:1907.11692*.
[10] A. Radford et al., "Improving language understanding by generative pre-training," OpenAI, San Francisco, CA, USA, 2018.
[11] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.
[12] B. Zeng, H. Yang, R. Xu, W. Zhou, and X. Han, "LCF: A local context focus mechanism for aspect-based sentiment classification," *Appl. Sci.*, vol. 9, no. 16, p. 3389, Aug. 2019.
[13] Q. Jiang, L. Chen, R. Xu, X. Ao, and M. Yang, "A challenge dataset and effective models for aspect-based sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 6280–6285.
[14] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*. Dublin, Ireland: Association for Computational Linguistics, 2014, pp. 27–35.
[15] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. D. Clercq, O. D. Clercq, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, and G. Eryiğit, "SemEval-2016 task 5: Aspect based sentiment analysis," in *Proc. Int. Workshop Semantic Eval.*, 2016, pp. 19–30.
[16] Z. Toh and J. Su, "NLANGP: Supervised machine learning system for aspect category classification and opinion target extraction," in *Proc. 9th Int. Workshop Semantic Eval. (SemEval)*, 2015, pp. 496–501.
[17] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," Microsoft, Redmond, WA, USA, Tech. Rep. MSR-TR-98-14, 1998.

[18] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*.
[19] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.
[20] X. Wang, F. Li, Z. Zhang, G. Xu, J. Zhang, and X. Sun, "A unified position-aware convolutional neural network for aspect based sentiment analysis," *Neurocomputing*, vol. 450, pp. 91–103, Aug. 2021.
[21] N. Zhao, H. Gao, X. Wen, and H. Li, "Combination of convolutional neural network and gated recurrent unit for aspect-based sentiment analysis," *IEEE Access*, vol. 9, pp. 15561–15569, 2021, doi: 10.1109/ACCESS.2021.3052937.
[22] W. Xue and T. Li, "Aspect based sentiment analysis with gated convolutional networks," 2018, *arXiv:1805.07043*.
[23] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective LSTMs for target-dependent sentiment classification," 2015, *arXiv:1512.01100*.
[24] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 606–615.
[25] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," 2017, *arXiv:1709.00893*.
[26] B. Huang, Y. Ou, and K. M. Carley, "Aspect level sentiment classification with attention-over-attention neural networks," in *Proc. Int. Conf. Social Comput., Behav.-Cultural Modeling Predict. Behav. Represent. Modeling Simulation*. Cham, Switzerland: Springer, 2018, pp. 197–206.
[27] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," 2016, *arXiv:1605.08900*.
[28] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 452–461.
[29] P. Lin, M. Yang, and J. Lai, "Deep mask memory network with semantic dependency and context moment for aspect level sentiment classification," in *Proc. IJCAI*, Aug. 2019, pp. 5088–5094.
[30] K. Sun, R. Zhang, S. Mensah, Y. Mao, and X. Liu, "Aspect-level sentiment analysis via convolution over dependency tree," in *Proc. EMNLP-IJCNLP*, 2019, pp. 5678–5687.
[31] C. Zhang, Q. Li, and D. Song, "Aspect-based sentiment classification with aspect-specific graph convolutional networks," in *Proc. EMNLP-IJCNLP*, 2019, pp. 4568–4578.
[32] Z. Wu and D. C. Ong, "Context-guided BERT for targeted aspect-based sentiment analysis," in *Proc. AAAI*, 2021, pp. 14094–14102.
[33] Z. Gao, A. Feng, X. Song, and X. Wu, "Target-dependent sentiment classification with BERT," *IEEE Access*, vol. 7, pp. 154290–154299, 2019, doi: 10.1109/ACCESS.2019.2946594.
[34] A. Rietzler, S. Stabinger, P. Opitz, and S. Engl, "Adapt or get left behind: Domain adaptation through BERT language model finetuning for aspect-target sentiment classification," 2019, *arXiv:1908.11860*.
[35] X. Li, L. Bing, W. Zhang, and W. Lam, "Exploiting BERT for end-to-end aspect-based sentiment analysis," 2019, *arXiv:1910.00883*.
[36] Y. Song, J. Wang, T. Jiang, Z. Liu, and Y. Rao, "Attentional encoder network for targeted sentiment classification," 2019, *arXiv:1902.09314*.
[37] A. Feng, X. Zhang, and X. Song, "Unrestricted attention may not be all you need–masked attention mechanism focuses better on relevant parts in aspect-based sentiment analysis," *IEEE Access*, vol. 10, pp. 8518–8528, 2022.
[38] M. H. Phan and P. O. Ogunbona, "Modelling context and syntactical features for aspect-based sentiment analysis," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3211–3220.
[39] Y. Lin, C. Wang, H. Song, and Y. Li, "Multi-head self-attention transformation networks for aspect-based sentiment analysis," *IEEE Access*, vol. 9, pp. 8762–8770, 2021.
[40] M. Xu, B. Zeng, H. Yang, J. Chi, J. Chen, and H. Liu, "Combining dynamic local context focus and dependency cluster attention for aspect-level sentiment classification," *Neurocomputing*, vol. 478, pp. 49–69, Mar. 2022.
[41] J. He, A. Wumaier, Z. Kadeer, W. Sun, X. Xin, and L. Zheng, "A local and global context focus multilingual learning model for aspect-based sentiment analysis," *IEEE Access*, vol. 10, pp. 84135–84146, 2022.
[42] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Dec. 2006.
[43] R. L. Iman and J. M. Davenport, "Approximations of the critical region of the fbietkan statistic," *Commun. Statist., Theory Methods*, vol. 9, no. 6, pp. 571–595, 1980.

TE LIN received the B.E. degree from Korea University, Seoul, South Korea, in 2021. He is currently pursuing the M.E. degree in computer technology with Hanyang University. His research interests include deep learning, natural language processing, and sentiment analysis.

INWHEE JOE received the B.S. degree in electronics engineering from Hanyang University, Seoul, South Korea, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1998. Since 2002, he has been a Faculty Member with the Division of Computer Science and Engineering, Hanyang University. His current research interests include the Internet of Things, cellular systems, wireless power communication networks, embedded systems, network security, machine learning, and performance evaluation.