Article

Adaptive Local Context and Syntactic Feature Modeling for Aspect-Based Sentiment Analysis

Jie Huang [1,2] , **Yunpeng Cui** [1,2,*] and **Shuo Wang** [3]

1   Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural, Beijing 100081, China
2   Agriculture Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China
3   Commonwealth Scientific and Industrial Research Organisation, Sydney 2122, Australia
*   Correspondence: cuiyunpeng@caas.cn

Abstract: Aspect-based sentiment analysis is a fine-grained sentiment analysis task that consists of two types of subtasks: aspect term extraction and aspect sentiment classification. In the aspect term extraction task, current methods suffer from the lack of fine-grained information in aspect term extraction and difficulty in identifying aspect term boundaries. In the aspect sentiment classification task, the current aspect sentiment classifier cannot adapt itself to the text and determine the local context. To address these two challenges, this work proposes an adaptive semantic relative distance approach based on dependent syntactic analysis, which uses adaptive semantic relative distance to determine the appropriate local context for each text and increase the accuracy of sentiment analysis. Meanwhile, the study also predicts the current word labels by combining local information features extracted by local convolutional neural networks and global information features to precisely locate the word labels. In two subtasks, our proposed model improves accuracy and F1 scores on the SemEval-2014 Task 4 Restaurant and Laptop datasets compared to the state-to-the-art approaches, especially in the aspect sentiment classification subtask.

Keywords: aspect term extraction; aspect sentiment classification; adaptive distance; local context

1. Introduction

Product reviews contain information about consumers' emotions regarding the product. By performing fine-grained recognition and sentiment analysis on the review text, it facilitates potential consumers to evaluate the product or service to help them make purchase decisions.

Sentiment analysis is a fundamental task in natural language processing [1]. Coarse-grained sentiment analysis [2] is targeted at the chapter level or utterance level, which is difficult to meet people's needs and obtain useful sentiment information. Fine-grained sentiment analysis targets aspects of a product or service in the text and analyzes the aspects of sentiment expressed by the user. Aspect refers to the attributes of the product. Aspect-based sentiment analysis (ABSA) [3] is a fine-grained sentiment analysis task that aims to identify aspects of a text and judge their sentiment polarity. An aspect-based sentiment analysis task consists of two subtasks, aspect term extraction (ATE) and aspect sentiment classification (ASC). Aspect-based sentiment analysis can provide a more comprehensive and in-depth analysis than chapter-level or utterance-level sentiment analysis. For example, given the text "While the ambiance was great, the food and service could have been a lot better." Aspect terms are "ambiance", "food", and "service". The aspect-based sentiment is positive, negative, and negative, respectively.

The text is "quality of food needs to be improved". The aspect term is "quality of food" and the aspect sentiment is negative. The following problems exist in the aspect-based sentiment analysis task, which is explored based on the above text. There are some problems in the aspect-based sentiment analysis task and the following questions are explored based on the above text.

C1: the low level of interaction between ATE and ASC tasks to fully federate multitasks. Most previous aspect-based sentiment analysis models focus only on the accuracy of aspect sentiment classification while neglecting the study of aspect term extraction and lacking the interaction between tasks. The current aspect-based sentiment analysis model cannot take full advantage of the relationship between multiple tasks to achieve mutual regulation between tasks. For example, in the above text, the aspect term "quality of food" is first determined in order to accurately determine the aspect sentiment, and the aspect term can be adjusted according to whether the aspect sentiment is wrong or not. In this paper, we provide a more efficient end-to-end aspect-based sentiment analysis solution relative to the model mentioned in Section 4.2, implementing both aspect term extraction and aspect sentiment classification. The flow of the aspect-based sentiment analysis task is shown in Figure 1.
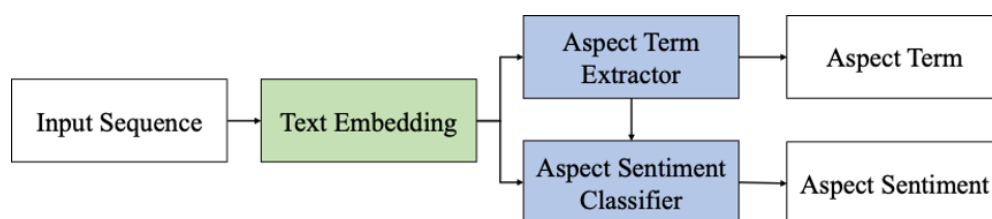

Figure 1. ABSA task procedure.

C2: the word vector lacks local information features and the term boundaries are blurred in terms of multiple words. Aspect term extraction task can be viewed as a sequence labeling task, using the BIO tag to annotate textual data. When the current word label is determined according to the global sequence features, the connection between the context cannot be effectively combined, and the phenomenon of aspect term splitting and label confusion occurs. For example, in the above text, the aspect term is "quality of food" instead of "quality of" and "food", and the error phenomenon of "I" in front of "B" may occur.

C3: the scope of the local context is difficult to define. For the solution of aspect sentiment classification, Heng et al. [4] proposed a local context focus mechanism to reduce the interference of non-local context on aspect sentiment analysis. Later, Phan et al. [5] applied syntactic dependencies to determine the semantic relative distance and, thus, the local context. However, due to the varying length of the text, setting a fixed threshold of syntactic relative distance will easily introduce other negative words, which will affect the judgment of the aspect emotion.

To solve the above aspect-based sentiment analysis problem, this paper proposes the adaptive aspect-based sentiment analysis model (A-ABSA), which combines local information and adaptive local context methods to achieve the best performance on the commonly used SemEval-2014 Task 4 dataset [3].

The main contributions of this paper are highlighted as follows:

(1) For the C1 problem, we propose the A-ABSA model to gradually train the two tasks of aspect term extraction and aspect sentiment classification, to enhance the interaction between tasks, to improve the model performance simultaneously, and to solve the aspect level sentiment analysis problem end-to-end.

(2) For the C2 problem, our aspect term extraction model integrates local context information by the equal-width convolutional neural network, aggregates global information and local information by the gating unit, enriches word vector information, and adds constraints to determine aspect term labels by Bi-LSTM and CRF to ensure the validity of terms.

(3) For the C3 problem, the adaptive semantic relative distance is introduced in the aspect sentiment classification task to determine the appropriate local context for aspect terms in each text, accurately analyze the sentiment polarity of aspect terms, and exclude the influence of irrelevant words on the sentiment judgment of aspect terms.

(4) In the SemEval-2014 Task 4 restaurant and laptop dataset, the proposed A-ABSB model improved by 5.12% on the laptop domain dataset F1, with a comparable performance on the restaurant domain dataset regarding the aspect term extraction compared to other advanced methods. For the aspect sentiment polarity classification on the laptop domain dataset, the accuracy is improved by 1.11% and F1 is improved by 1.06%, the accuracy on the restaurant domain dataset is improved by 0.73% and F1 is improved by 1.11%.

2. Related Works

Most aspect-based sentiment analysis methods are oriented toward aspect term extraction and aspect sentiment classification as separate task studies. Recently, joint approaches for multiple tasks have also emerged. In this section, relevant research developments in aspect term extraction and aspect sentiment classification are presented.

2.1. Aspect Term Extraction

Earlier, the aspect term extraction task was mainly based on a rule and lexicon approach, proposing a series of unsupervised models. Later supervised statistical models were used to extract aspect terms, such as conditional random fields(CRF), hidden Markov model(HMM), etc.

With the development of deep learning techniques, most researchers worked on developing various types of neural network models. The aspect term extraction task is similar to the named entity recognition task. Since the bi-directional long short-term memory (Bi-LSTM) model is a common model in named entity recognition tasks, it can be applied to the aspect term extraction task as well. Since words in a sentence have dependency information between them, recurrent neural networks are sequential models that cannot effectively capture the tree-based dependency information of a sentence. In order to effectively utilize the dependency information of sentences, Ye et al. [6] proposed the DTBCSNN model to capture syntactic features by introducing a multilayer convolutional neural network based on the dependency tree. Xu et al. [7] proposed a DE-CNN model to obtain word embeddings using generic embeddings and domain-specific embeddings and then used convolutional neural networks for aspectual term extraction. The DE-CNN model makes full use of domain knowledge to make the word vector more accurate, but cannot dynamically acquire contextual semantic information. Prior to 2018, word embeddings used Word2Vec [8] or GloVe [9] models, which provide only context-independent word-level features, but this is insufficient to capture complex semantic dependencies in sentences. The advent of pre-trained models, such as BERT [10] allows for improved model performance by adding even simple linear classification layers to the BERT model structure. Li et al. [11] study the ability of the BERT model to model contextual embedding, which is validated in aspect-based sentiment analysis. Since words play different roles in different sentences, there are dependencies between different tags. Zhang et al. [12] proposed the BERT-GLCLD model to construct a global–local context representation using the BERT model and location-aware attention, and a label-dependent module based on recurrent neural networks and conditional random fields to constrain the boundaries of aspect terms. The BERT-GLCLD model focuses on both fine-grained and coarse-grained information, combining label information for prediction, but does not fully utilize contextual information, and the current word prediction relies not only on the previous word but also uses the latter word to assist in prediction. To enrich the word vector information, Phan et al. [5] proposed the CASE model, which combines dependency-based embedding, contextual embedding, and lexical embedding to enhance the performance of the model. The CASE model introduces external knowledge and makes full use of various information, which increases the complexity of the model.

Inspired by Zhang et al. [12], it is believed that global–local contextual information can understand different semantic meanings of words in different sentences. In this study, local context information is integrated using convolutional neural networks to extract features further.

This study is the first to integrate local contextual information with the help of an equal-width convolutional neural network and to explore the potential of the new network structure for the aspect term extraction task.

2.2. Aspect Sentiment Classification

Aspect sentiment classification is a multi-classification task and has been studied more extensively in relation to aspect term extraction. The research on aspect sentiment classification is divided into traditional machine learning-based methods and deep learning methods. The most successful of the traditional machine learning approaches is a feature-based support vector machine (SVM), which engineers features through experts and uses external resources, such as parsers and sentiment dictionaries.

Current research on aspect sentiment classification is mainly based on deep learning techniques. Commonly used deep neural networks are recurrent neural networks (RNNs), convolutional neural networks (CNNs), etc. Neural network models can learn continuous text representation from data without any feature engineering to obtain rich text information. Tang et al. [13] proposed the TD-LSTM model to model the correlation between target words and their context for target sentiment classification. Meanwhile, in order to better utilize the target information and capture the relationship between the target word and its context, the TC-LSTM model was proposed, which makes each word cascade with the target word on the basis of the TD-LSTM model and improves the accuracy of the target sentiment classification. Aspect sentiment is determined by the sentiment words in the text. Wang et al. [14] proposed the ATAE-LSTM model, which enables the model to focus on different parts of the sentence when different aspects are involved. Ma et al. [15] proposed the IAN model, which models the target and the context separately and learns them interactively, both focusing on different parts of the sentence and supervising the modeling of the target. To further focus on the degree of influence of different parts of the sentence on the target, Chen et al. [16] proposed the RAM model, which uses multiple attention mechanisms to synthesize important features of complex sentence structures. Huang et al. [17] proposed the AOA model, similar to the IAN model, which can adequately capture the interactions between aspects and contexts and focus on the focused parts of sentences, but cannot handle complex emotional expressions. In aspect sentiment classification tasks, coarse-grained attention mechanisms are prone to information loss. To address this situation, Fan et al. [18] propose the multi-grained attention network model (MGAN) to implement word-level interactions between aspects and contexts. The model utilizes both fine-grained and coarse-grained attention mechanisms to form a multi-grained attention network structure and implements aspect alignment loss to bring additional useful information for aspect sentiment classification, which further improves the performance of the model. The application of attentional mechanisms in aspect sentiment classification tasks may incorrectly identify syntactically irrelevant contextual words as cues for determining aspect sentiment. To solve this problem, Zhang et al. [19] proposed the ASGCN model, which uses graph convolutional networks, syntactic information, and remote word dependencies to predict aspect sentiment. The above model does not make full use of contextual semantic information for word embedding based on a static word vector model.

After the pre-trained models came out, most researchers started to use pre-trained models, such as BERT, Roberta, etc., for word embedding to obtain more informative word vectors. Song et al. [20] proposed an attention encoder network (AEN) for modeling and semantic interaction of target words and contexts to address the problem that RNNs model contexts and target words that are difficult to parallelize. Rietzler et al. [21] argued that fine-tuning the BERT model using a domain-specific corpus and then fine-tuning it in a supervised manner for downstream tasks would improve the performance of the model.

Zeng et al. [4] proposed the LCF-BERT model, which is the first local context-focused mechanism to focus more on contextual words using the context-feature dynamic mask (CDM) or context-feature dynamic weighting (CDW). This demonstrates the importance of

local context for predicting aspect sentiment. This study uses one of the context-feature dynamic mask strategies. Phan et al. [5] propose the LCFS-ASC model based on the LCF-BERT model, which uses dependent syntactic analysis to determine the local context and addresses the use of token-based semantic relative distance to cover negative words or exclude emotive words. Yang et al. [22] proposed the LCF-ATEPC model based on LCF-BERT, which can perform aspect term extraction and aspect sentiment classification on Chinese reviews. Later, Yang et al. [23] argued that using syntactic dependency trees would take up more resources, so they proposed the LSA mechanism to learn the aspect sentiment dependency in sentiment clusters and constructed a differential weighting strategy to enhance the importance of sentiment dependency. The above model learns local text information, focuses on fine-grained information, fuses fine-grained information with coarse-grained information, and even uses syntactic relations to determine local text information, which is more accurate. However, the local text length varies and should be adaptively adjusted, otherwise using a fixed length will result in errors, which can be followed in Section 3.2.2.

Based on previous studies, we propose adaptive semantic relative distance as a starting point for our research.

3. Methodology

This study proposes a multi-task learning aspect-based sentiment analysis model that uses adaptive semantic relative distance to integrate local contextual information into the model.

Given a text $S = \{w_1, w_2, ..., w_n\}$, where n is the length of the text. ABSA model is able to extract the aspect terms in the text, $A = \{a_1, a_2, ..., a_m\}$, where m is the length of the aspect terms, and determine their sentiment polarity, $y_p \in \{Positive,\ Negative,\ Neutral\}$.

This section introduces the structure of the A-ABSA model and the methods involved in it. The ATE module and the ASC module are introduced in order from bottom to top according to the order of the network hierarchy. The overall structure of the ABSA model is shown in Figure 2.
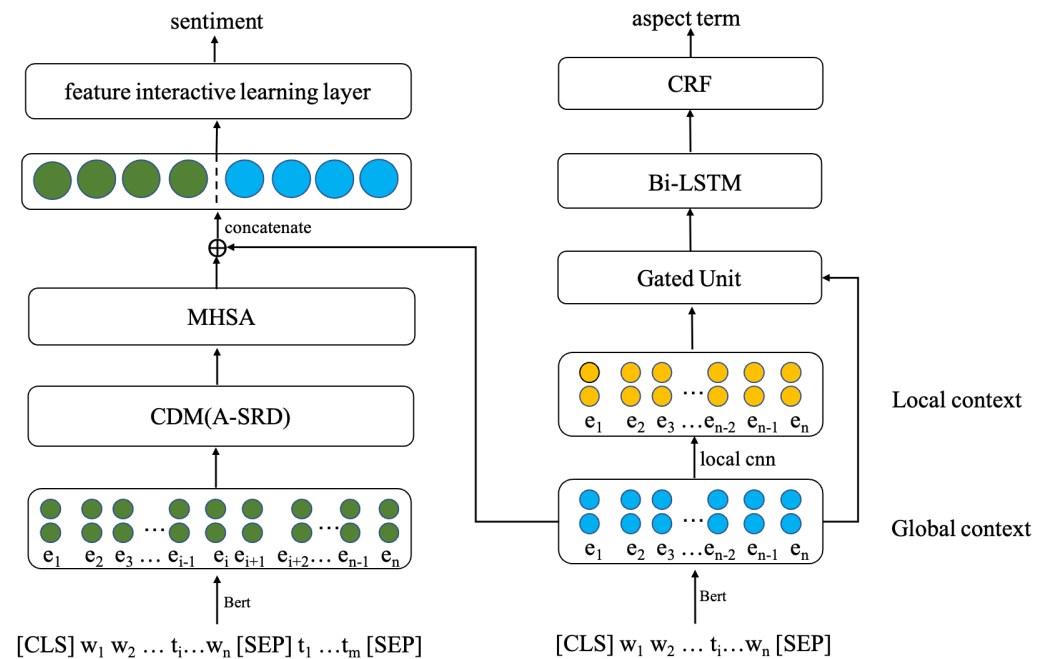

Figure 2. Overall structure of ABSA.

A-ABSA uses BERT as the word embedding layer, which can effectively obtain contextual information. The aspect sentiment classification model structure refers to that proposed by Yang et al. [22] on which adaptive semantic relative distance is added to solve

the problem of different sizes of semantic relative distance, see Section 3.2.2 for details. Aspect term extraction model structure classical model is BiLSTM+CRF, based on which CNN and gate mechanism are introduced to further extract word vector information, see Section 3.1 for details.

A-ABSA model workflow: (i) input text and convert it into model input format; (ii) aspect term extraction, consisting of word embedding layer, convolutional layer, gated unit layer, Bi-LSTM layer, and CRF layer; (iii) output aspect term; (iv) transform aspect sentiment classification input format; (v) aspect sentiment classification, consisting of word embedding layer, CDM layer, MHSA layer, and interactive learning layer; (vi) determine sentiment polarity.

3.1. Aspect Term Extraction Structure

Aspect term extraction can be viewed as a sequence labeling problem. The study uses BIO labels, namely Begin, Inside, and Outside, which are the beginning of the aspect term, the inside of the aspect term, and not the aspect term, respectively. For example, this text "Good spreads, great beverage selections, and bagels really tasty.", is marked as {O, B-asp, O, O, B-asp, I-asp, O, B-asp, O, O, O, O}.

The right structure of Figure 2 shows the structure of the aspect term extraction model, including the BERT pre-training model, convolutional layer, gating unit, Bi-LSTM and CRF, which further enriches the word vector information by combining global and local information, and adds constraints for determining the aspect term boundaries.

3.1.1. Input Representation

The input format of the BERT model requires the addition of special tokens "[CLS]" and "[SEP]", the [CLS] token is mainly used for the classification task, and the [SEP] token is used to distinguish between two sentences. When the input is only one sentence, it is simply added to the end of the sentence. The input text format of the BERT pre-training model is "[CLS]" + Input Sequence + "[SEP]". For each word in the text, its representation consists of three components pre-trained word vector, position vector, and segment vector.

3.1.2. BERT Embedding Layer

The emergence of transformer-based pre-training models, represented by GPT and BERT, has brought great benefits to the field of natural language processing. The previous word vector generation is based on word vector models such as word2vec and glove, which are static vectors and cannot solve the problem of multiple meanings of words. In contrast, pre-trained models, such as BERT, can generate contextual word vectors based on contextual information, providing a more informative word vector for many downstream tasks and improving the performance of downstream tasks. In this study, we use the BERT pre-trained basic model as the word vector model.

3.1.3. Global–Local Context Modeling

Aspect terms consist of consecutive words in the text. Each word has the problem of having multiple semantics. To solve the problem, the BERT model is used to generate contextual word vectors. We construct global word vectors and local contextual word vector representations and design the integration of word vector information by combining both types of word vector information using a gate cell structure.

The representation of words in the global sequence refers mainly to the dependencies between words at the sentence level. We use the BERT model to mine words for global sequence features in sentences. First, the input text $S = \{w_1, w_2, \ldots, w_n\}$, where n is the length of the text, is transformed into the input format of the BERT model, $S = ([CLS], w_1, w_2, \ldots, w_n, [SEP])$.

$$g_i = BERT(w_i) \tag{1}$$

where g_i is the global sequence representation of the word w_i and i indicates the position of the word in the sentence.

In the ATE task, the neighboring words of each word will have a significant impact on predicting its label, and the more distant words will have less impact, requiring a focus on local contextual information. We propose to use equal-width convolutional neural networks to mine the contextual information around words and extract local features.

In this study, the local contextual features of each word are obtained using an equal-width convolutional layer with a convolutional kernel size of K and a step size $S = 1$, with P zeros complemented at both ends of the input.

$$P = \frac{K-1}{2} \tag{2}$$

$$l_i = Conv(g_i) \tag{3}$$

where l_i is the local contextual feature of the word w_i.

To enhance the word vector information, the global sequence features and local sequence features are combined through the gating unit f_i.

$$f_i = \sigma(W_f g_i + U_f l_i + b_f) \tag{4}$$

where W_f, U_f, and b_f are the training parameters of the gating unit, g_i is the global sequence feature, and l_i is the local sequence feature.

The gating unit can combine global and local information, keep the important information, and remove the redundant information. Finally, we obtain the global–local gl_i.

$$gl_i = f_i g_i + (1 - f_i)l_i \tag{5}$$

The parameter f_i controls the filtering of information. When $f_i > 0.5$, it means the global information is more important; when $f_i < 0.5$, it means the local information is more important; when $f_i = 0.5$, it means both parts of information are equally important.

3.1.4. Label Constraint Module

To make full use of the information, the global–local information is modeled with the help of a bidirectional long short-term memory network (Bi-LSTM). The dependencies of the labels are then constrained using conditional random fields (CRF), which are automatically learned by the CRF layer from the training data during the training process. The illegal case of the label "I" preceding label "B" is avoided. When the model predicts the aspect label of the current word, it is important to understand the semantics of the previous word as well as to pay attention to the information of the latter word. After feeding the global–local contextual word vectors to the BiLSTM module and semantic modeling of each word, we obtain the final sequence $S = \{s_1, s_2, \ldots, s_n\}$. For the real label sequence $Y = \{y_1, y_2, \ldots, y_n\}$, the output graph of CRF is connected by undirected edges to predict the probability of the correct label for each word based on the state features and transfer features. The conditional probability $p(Y|S)$ formula is shown below.

$$p(Y \mid S) = \frac{\prod_{i=1}^{n} \psi_i(y_{i-1}, y_i, S)}{\sum_{y' \in \mathfrak{y}(S)} \prod_i^n \psi_i\left(y'_{i-1}, y'_i, S\right)} \tag{6}$$

where $\mathfrak{y}(S)$ denotes all possible labeled sequences of the observed sequence and the conditional probability $p(Y|S)$ is the score of the labeled sequence on a given observed sequence.

3.2. Aspect Sentiment Classification Structure

Given a text $S = \{w_i | i \in [1, n]\}$ and aspect terms in the text $A = \{a_i | i \in [1, m]\}$, we need to determine the sentiment (positive, neutral, or negative) of the aspect terms in the text.

The left structure of Figure 2 shows the architecture of aspect sentiment classification, which consists of the BERT embedding layer, MHSA, and interactive learning layer. It deter-

mines the local context of aspect terms based on the dependent syntactic tree to eliminate the influence of irrelevant words and determine their sentiment polarity.

3.2.1. Input Representation

The previous study verified that separate modeling and interaction between aspect terms and contexts facilitated the judgment of aspect term sentiment polarity. So the input format taken in the construction of local context features was "[CLS] + S + [SEP] + A + [SEP]". The input format of global context features was the same as that of aspect term extraction.

3.2.2. Local Context Focus

The effective information of aspect terms exists in the local context of the text. The determination of local context includes semantic relative distance methods and methods based on syntactic dependencies. Zeng et al. [4] first proposed to apply local context to sentiment classification by using semantic relative distance (SRD) to determine local context. The context-feature dynamic mask strategy and context-feature dynamic weighting strategy are also proposed. Then Phan [5] pointed out that using semantic relative distance to determine local context may make the sentiment words not present in the local context and have a smaller effect on judgmental aspect sentiment. Thus Phan et al. proposed to use the shortest distance between pairs of nodes in the dependent syntactic tree to determine the local context, which can effectively solve this problem. In our study, we found that the length of the review text varied, and the length of each aspect of the local context varied. To address this point, we propose adaptive semantic relative distances (ASRD), which assigns semantic relative distances to each text and determines the local context. In the study, we used the spacy toolkit to generate dependency syntax trees as a way to calculate the semantic relative distances between different words. If the aspect term is a single word, the semantic relative distance is the shortest semantic distance between the two words. If the aspect terms are composed of multiple words, the semantic distance between the input word and the multi-word aspect item is calculated as the average distance between each aspect sub-word and the input word. Figure 3 shows the dependency syntactic tree of the review text.
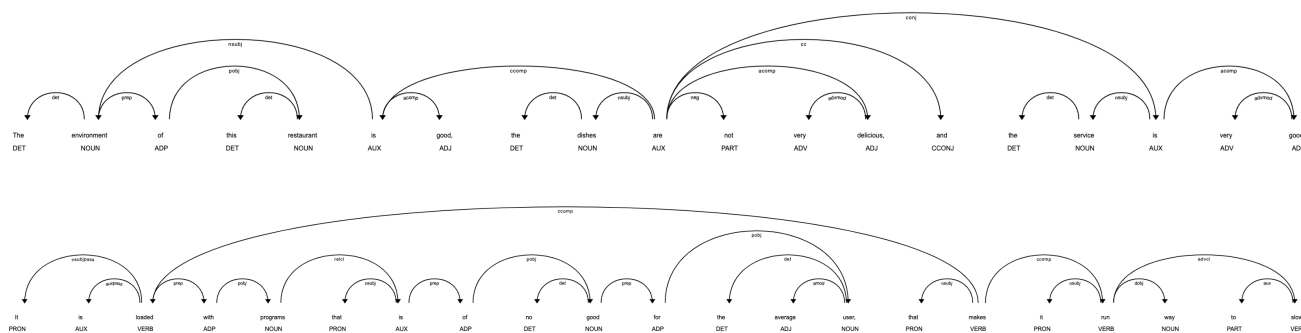

Figure 3. Dependency syntactic tree of the review text.

The first review text in Figure 3 is "The environment of this restaurant is good, the dishes are not very delicious, and the service is very good." The aspect terms in the review text are "environment", "dishes" and "service". Take "environment" as an example to calculate the semantic relative distance.

SRD(environment, good) = 2

The second review text in Figure 3 is "It is loaded with programs that are of no good for the average user, that makes it runway to slow." The aspect terms in the review text are "programs" and "run". Take "programs" as an example to calculate the semantic relative distance. The corresponding emotional word for "programs" is "no good".

SRD(programs, no) = 4

SRD(programs, good) = 3
SRD(programs, no good) = 4

If the threshold of SRD takes a smaller value, it may cause the local context of other aspect words to not include their sentiment words. If the threshold of SRD takes a larger value, it may cause other aspect words to contain irrelevant negative sentiment words.

To solve this problem, this study proposes adaptive semantic relative distance to select the appropriate SRD for each text, satisfying as much as possible that the local context of each aspect contains its sentiment words and not irrelevant negative sentiment words.

In this study, it is observed from the dataset that the length of aspect words is about 3 and the length of text containing a single aspect is 7. Therefore, the original text will be directly used as the local context for the text of length 7. If the text length is more than 7, it is possible that there may be more than one aspect in a text, and therefore the local context needs to be determined.

Therefore, two calculation methods are proposed in this study. The first method obtains the maximum shortest distance from each word to the aspect word by depending on the syntactic tree and determines the local context by setting a threshold value.

$$bdist = max(D) \tag{7}$$

$$SRD = bdist * \alpha \tag{8}$$

$$ASRD = max\{SRD, STHD\} \tag{9}$$

where D is the set of shortest distances from each word to the aspect word, bdist is the maximum value in D, α is the hyperparameter to regulate the size of *SRD*, and *STHD* denotes the minimum threshold to prevent the calculated *SRD* from being too small, resulting in the local context not containing the sentiment words.

The second method is to determine the local context by the length of the text, where n is the input text length. The formula for the second calculation method is shown below.

$$num = max\{n * \beta, Snum\} \tag{10}$$

where β is a hyperparameter to set the number of local context words, Snum denotes the minimum number and serves a similar purpose as STHD to prevent the calculated threshold from being too small, resulting in too little information being contained in the local context.

After obtaining the shortest distance set D from each word to the aspect word, the top num words that are closer to the aspect word are selected as local contexts.

After determining the local context, we refer to the context-feature dynamic mask strategy proposed by Zeng et al. [4] to mask the non-local context and preserve the local context semantics. Suppose the set of local context words is LCW, the input text $S = \{w_i | i \in [1, n]\}$, and the local context is V^l.

$$v_i^m = \begin{cases} O, & w_i \in LCW \\ I, & w_i \notin LCW \end{cases} \tag{11}$$

$$M = [v_1^m, v_2^m, \ldots, v_n^m] \tag{12}$$

$$V^{CDM} = V^l \odot M \tag{13}$$

where O and I denote the zero and one vectors, respectively, and v_i^m denotes the masking vector, which determines whether each word m_i belongs to a local context word according to the adaptive semantic relative distance calculation.

In this study, only the CDM strategy is used instead of the CDW strategy. Because the CDW strategy will indirectly introduce the influence of other irrelevant words. To avoid this part of the impact, only the CDM strategy is used. Meanwhile, to avoid information loss, global contextual sequence features are used to supplement the information.

3.2.3. Multi-Head Self-Attention and Feature Interactive Learning Layer

Ma et al. [15] proposed an interactive attention network model (IAN) based on long and short-term memory networks and attention mechanisms, which verified the interaction modeling of context and target and helped to determine emotions. We used a multi-head self-attention mechanism to extract information features from multiple dimensions. After the local information is modeled, the global and local information are stitched together and learned from each other by linear functions for the input vector of the softmax function.

$$O^l = MHSA(V^{CMD}) \tag{14}$$

$$O^{lg} = [O^l : O^g] \tag{15}$$

$$O^{lg}_{dense} = W^{lg} \cdot O^{lg} + b^{lg} \tag{16}$$

$$O^{lg}_{FIL} = MHSA(O^{lg}_{dense}) \tag{17}$$

$$sentiment_{label} = softmax(O^{lg}_{FIL}) \tag{18}$$

where O_l is the local context vector, O_g is the global context vector, and O^{lg}_{FIL} is the final word vector input to softmax.

4. Experiments

4.1. Datasets and Hyperparameters Setting

4.1.1. Dataset Details

To evaluate the performance of our proposed model, we evaluate and compare the ATE model and the ASC model on two baseline datasets. The two baseline datasets are the laptop domain dataset and the restaurant domain dataset from the SemEval-2014 Task 4 challenge [3]. Each example text in the dataset is labeled with aspect terms and sentiment polarity. The original dataset was stored in XML format. The data were preprocessed in the experiment to reformat the original dataset. An example of the raw data is shown in Figure 4. The data pre-processing process is shown in Figure 5. Detailed information on the dataset is shown in Table 1 below.

Table 1. Restaurant and Laptop dataset details.

Dataset	Train			Test		
	Pos	**Neg**	**Neu**	**Pos**	**Neg**	**Neu**
Restaurant	1315	462	368	426	143	146
Laptop	602	514	260	201	197	94

```xml
<sentence id="2339">
    <text>I charge it at night and skip taking the cord with me because of the good battery life.</text>
    <aspectTerms>
        <aspectTerm term="cord" polarity="neutral" from="41" to="45"/>
        <aspectTerm term="battery life" polarity="positive" from="74" to="86"/>
    </aspectTerms>
    <aspectCategories>
    </aspectCategories>
</sentence>
```

Figure 4. Example of benchmark dataset.

```
I 0 -1
charge 0 -1
it 0 -1
at 0 -1
night 0 -1
and 0 -1
skip 0 -1
taking 0 -1
the 0 -1
cord B-ASP 1
with 0 -1
me 0 -1
because 0 -1
of 0 -1
the 0 -1
good 0 -1
battery B-ASP -1
life I-ASP -1
. 0 -1
```
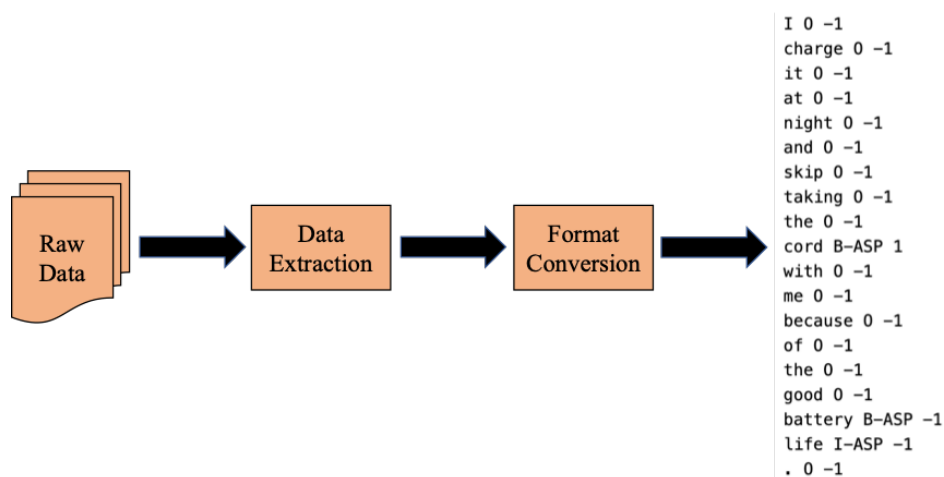
Figure 5. Processing of data flow.

4.1.2. Hyperparameter Setting

Based on the hyperparameter settings with reference to previous studies, some hyperparameters are optimized through continuous experiments to achieve the best performance of the model. Some of the important hyperparameter settings for the ATE and ASC models are listed in Table 2.

Table 2. Some important hyperparameter settings for ATE and ASC models.

Hyperparameters	Setting
batch size	16
learning rate	3×10^{-5}
max length	120
rate	0.4
dropout	0.1
optimizer	AdamW

Note: the rate parameter determines the magnitude of the semantic relative distance.

4.2. Baseline Models

4.2.1. ATE Models

We compare recent models in the ABSA task to demonstrate the effectiveness of the proposed ATE architecture and the ASC architecture. Since the performance of the multitasking model is lower than that of the single-tasking model, the model comparison takes place between independent single-tasking models.

The group of models compares the performance of aspect term extraction models and trains single-task models independently. In order to compare the effectiveness of local information extraction based on an equal-width convolutional neural network, a model that integrates local contextual information (in a way involving attention mechanisms and other convolutions) is chosen to verify the effectiveness of the method.

BiLSTM and MNA models perform aspect term extraction with the help of recurrent neural networks. DTBCSNN, DE-CNN, and STC models apply convolutional neural networks for the aspect term extraction. BERT-AE and CSAE perform aspect term extraction by using BERT as a word embedding layer for word vectorization. The following is a description of each ATE model.

BiLSTM [13] is a commonly used named entity recognition model using bidirectional LSTM for word embedding representation.

DTBCSNN [6] is a dependency tree-based convolutional stacking neural network to extract aspect terms without any artificial feature engineering.

DE-CNN [7] is based on domain embedding and generic embedding using a multilayer convolutional neural network model.

BERT-AE [10] uses the BERT model and softmax for the aspect term extraction task.

CSAE [5] is an aspect term extraction model that combines contextual features, dependent syntactic relations, and lexical properties.

STC [24] is an aspect term extraction model that aggregates local information using graph convolutional neural networks.

MNA [25] is based on an improved multi-head attention mechanism for aspect term extraction.

4.2.2. ASC Models

The group of models compares the performance of aspect sentiment classification models. We choose models based on different word embeddings and models with semantic relative distances or fixed syntactic relative distances, respectively, to verify the importance of the word embedding approach and the effectiveness of adaptive semantic relative distances.

ASC models are divided into three main categories, which are LSTM, GCN, and BERT-based models for aspect sentiment classification. The LSTM-based models include TD-LSTM, IAN, RAM, AOA, and MGAN. The GCN-based model is ASGCN. The BERT-based models include BERT-SPC, AEN-BERT, BERT-PT, LCF-BERT, and LCFS-ASC. The following is an introduction to each ASC model.

TD-LSTM [13] is a model based on two LSTMs capturing contextual information related to the target.

IAN [15] is an interactive attentional network model based on LSTM and attentional mechanisms that consider the interactive learning of target words and contexts.

RAM [16] is a model based on BiLSTM and multiple attention mechanisms that focus on important features in complex sentences.

AOA [17] models aspects and sentences in a federated manner, focusing on the important parts of the sentence.

MGAN [18] is a new multi-granularity attention network that captures word-level interactions between aspects and contexts.

ASGCN [19] builds a graph convolutional network on a sentence dependency tree, exploiting syntactic information and word dependencies.

BERT-SPC [20] is a pre-trained BERT model designed for sentence pair classification tasks.

AEN-BERT [20] is an attentional encoder network that models between context and target based on attentional encoders.

BERT-PT [26] is inspired by the reading comprehension task and is suitable for the aspect sentiment classification task.

LCF-BERT [4] uses a semantic relative distance local focus mechanism to determine the local context to eliminate the influence of irrelevant words.

LCFS-ASC [5] is a sentiment classification model that uses dependent syntactic trees to determine local context.

4.3. Model Variations

To evaluate the compositional structure of our proposed model, we perform a series of experiments in different settings.

For the ATE task model structure, we remove certain modules from the model to show their impact on the final model performance.

Ours-model-ATE-Conv removes the equal-width convolution layer to check the impact of the equal-width convolution layer on aspect extraction.

Ours-model-ATE-BiLSTM removes the BiLSTM layer to check the importance of BiLSTM.

Ours-model-ATE is our proposed model of ATE, which consists of BERT, convolutional neural network, Bi-LSTM, and CRF.

For the ASC task, we compare with the model using fixed semantic relative distances to verify the effectiveness of adaptive semantic relative distances. The validity of LCF has been verified in previous studies, so no comparison is made.

Ours-model-ASC-M1 is a model that uses the first type of calculating adaptive semantic relative distances.

Ours-model-ASC-M2 is a model that uses the second type of calculating the adaptive semantic relative distance.

The BERT models used in the experiments are all basic models to ensure a fair comparison with other models.

For ATE tasks, we use F1 scores as evaluation metrics. For ASC tasks, we use accuracy and F1 scores as evaluation metrics.

5. Results & Analysis

5.1. Aspect Term Extraction Result

5.1.1. ATE Main Results

Table 3 summarizes the results of our proposed model in comparison with other ATE models. Compared with the word embedding-based model, our model outperforms Bi-LSTM, DTBCSNN, and DE-CNN models by 9.05%, 7.11%, 1.18%, and 4.65%, 2.1%, and 11.7% on the laptop dataset and restaurant dataset, respectively. Our model approaches the CSAE model performance in the restaurant domain and exceeds the STC advanced model in the laptop domain.

Based on the model comparison, it is seen that our model is valid in both domain datasets, indicating that our proposed model has some generality.

Table 3. Comparing our ATE model with other advanced methods based on F1 scores (%).

Baseline	Domain	Laptop	Restaurant
	Model	**F1**	**F1**
RNN	BiLSTM	73.72	81.42
	MNA	78.63	85.71
CNN	DTBCSNN	75.66	83.97
	DE-CNN	81.59	74.37
	STC	82.34	-
BERT	BERT-AE	73.92	82.56
	CSAE	77.65	**86.65**
Our	Ours-model-ATE-Conv	81.84	85.17
	Ours-model-ATE-BiLSTM	81.42	85.02
	Ours-model-ATE	**82.77**	**86.07**

Note: The best result in each dataset is highlighted in bold. '-' means no result.

5.1.2. Ablation Study

To study the effects of different components in the ATE model, we remove the components to be studied and study the effects of the remaining structures on both datasets. The results in Table 3 show that removing the convolutional neural network reduces the remaining model structure by 0.93% and 0.9% on the Laptop and Restaurant datasets, respectively, and removing the Bi-LSTM reduces the remaining model structure by 1.35% and 1.05% on the Laptop and Restaurant datasets, respectively. Convolutional neural networks extract contextual features at the local level and BiLSTM relates contextual information at the global level, both of which further enrich the word vector information. The model does not make further use of external information such as syntactic information and lexical information, and only uses the basic BERT model to obtain contextual information, so the model has much room for improvement.

5.2. Aspect Sentiment Classification Result

5.2.1. ASC Main Results

Table 4 shows that our proposed ASC model using adaptive semantic relative distances performs well on both laptop and restaurant datasets. Our model does not utilize domain datasets to obtain additional knowledge for domain-specific embedding, and by comparing with the LCFS-ASC model and LCF-BERT model, it demonstrates that the adaptive syntactic relative distance is improved while maintaining the advantages of the old method, and also demonstrates again the effectiveness of syntactic relative distance in determining local contextual information.

Table 4. Comparison of our ASC model with state-of-the-art methods in terms of accuracy and F1 score (%).

Baseline	Domain	Laptop		Restaurant	
	Model	F1	Acc	F1	Acc
LSTM	TD-LSTM	68.43	71.83	66.73	78
	IAN	70.81	72.5	67.38	72.05
	RAM	71.35	74.49	70.8	80.23
	AOA	67.52	72.62	70.42	79.97
	MGAN	72.47	75.39	71.94	81.25
GCN	ASGCN	71.05	75.55	72.19	80.86
BERT	BERT-SPC	75.55	79.4	80.52	86.77
	AEN-BERT	76.31	79.93	73.76	83.12
	BERT-PT	75.08	78.07	76.96	84.95
	LCF-BERT-CDW	76.2	80.21	79.12	85.91
	LCF-BERT-CDM	76.2	79.65	78.74	85.73
	LCFS-ASC-CDW	77.13	80.52	80.31	86.71
	LCFS-ASC-CDM	76.45	80.34	80.10	86.13
Our	Ours-model-ASC-M1	**77.51**	**81.45**	80.64	86.6
	Ours-model-ASC-M2	77.3	80.35	**81.21**	**86.86**

Note: The best result in each dataset is highlighted in bold.

5.2.2. Analyze the Advantages of ASRD

Since aspect sentiment is determined by sentiment words, which exist in context, determining the syntactically local context of an aspect can accurately capture aspect sentiment information. Phan et al. [5] pointed out that determining the context based on positional relative distance does not include the affective words in the local context, so a dependent syntactic tree is used to determine the syntactic relative distance. For different texts, the semantic relative distance thresholds are different, so it is necessary to give the adapted semantic relative distance thresholds according to the characteristics of each text. According to this situation, we propose an adaptive semantic relative distance, which is determined by the text length. If there is only one aspect in the text and there are no other superfluous emotional words, the local context is the whole text. If there are irrelevant negative words in the text, the threshold of semantic relative distance will be adjusted to exclude them. If the text contains multiple data items, different semantic relative distance thresholds are determined for each data item.

After using adaptive semantic relative distance, accuracy and F1 improved by 1.11% and 1.06% on the Laptop dataset, and accuracy and F1 improved by 0.73% and 1.11% on the Restaurant dataset, respectively.

6. Conclusions and Future Work

We propose an end-to-end ABSA solution, which reflects the importance of local context information in different tasks. Therefore, fine-grained information is essential in task learning. Our proposed adaptive semantic relative distance approach relies on text-dependent syntactic structures. A suitable local context is determined based on the

text length. In aspect term extraction tasks, equip-width convolutional neural networks can efficiently aggregate local information. Other convolutional neural networks will be explored later to apply them to this task. From the experimental results, A-ABSA uses adaptive semantic relative distance to improve the accuracy and F1 by 1.24%, 1.31%, and 0.95%, 2.09%, on the laptop dataset and restaurant data, respectively, compared to LCF-BERT. This validates the effectiveness of the adaptive relative distance method. Our proposed solution is applicable to long-text comment information and cannot effectively model contextual syntactic relationships for short-text information. Our proposed model does not make effective use of labeling information, using different classification models to solve different aspects of the sentiment analysis subproblem, and never being able to link the two more closely. The use of generative models to solve the ABSA multitasking problem has been proposed as a preliminary exploration. It would be worthwhile to investigate whether the combination of existing methodological theories with it would further promote the development of ABSA.

References

1. Bakshi, R.K.; Kaur, N.; Kaur, R.; Kaur, G. Opinion mining and sentiment analysis. In Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016; pp. 452–455.
2. McDonald, R.; Hannan, K.; Neylon, T.; Wells, M.; Reynar, J. Structured models for fine-to-coarse sentiment analysis. In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, 23–30 June 2007; pp. 432–439.
3. Maria Pontiki, D.; John Pavlopoulos, H.; Ion Androutsopoulos, 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. 23–24.
4. Zeng, B.; Yang, H.; Xu, R.; Zhou, W.; Han, X. Lcf: A local context focus mechanism for aspect-based sentiment classification. *Appl. Sci.* **2019**, *9*, 3389. [CrossRef]
5. Phan, M.H.; Ogunbona, P.O. Modelling context and syntactical features for aspect-based sentiment analysis. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Virtual Event, 5–10 July 2020; pp. 3211–3220.
6. Ye, H.; Yan, Z.; Luo, Z.; Chao, W. Dependency-tree based convolutional neural networks for aspect term extraction. In *Advances in Knowledge Discovery and Data Mining. PAKDD 2017*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2017; pp. 350–362.
7. Xu, H.; Liu, B.; Shu, L.; Philip, S.Y. Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; pp. 592–598.
8. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, arXiv:1301.3781.
9. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
10. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.
11. Li, X.; Bing, L.; Zhang, W.; Lam, W. Exploiting BERT for End-to-End Aspect-based Sentiment Analysis. *arXiv* **2019**, arXiv:1910.00883.

12. Zhang, Q.; Shi, C. Exploiting BERT with global-local context and label dependency for aspect term extraction. In Proceedings of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), Sydney, Australia, 6–9 October 2020; pp. 354–362.

13. 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, Osaka, Japan, 11–16 December 2016; pp. 3298–3307.

14. Wang, Y.; Huang, M.; Zhu, X.; Zhao, L. Attention-based LSTM for aspect-level sentiment classification. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 606–615.

15. Ma, D.; Li, S.; Zhang, X.; Wang, H. Interactive Attention Networks for Aspect-Level Sentiment Classification. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 4068–4074.

16. Chen, P.; Sun, Z.; Bing, L.; Yang, W. Recurrent attention network on memory for aspect sentiment analysis. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 7–11 September 2017; pp. 452–461.

17. Huang, B.; Ou, Y.; Carley, K.M. Aspect level sentiment classification with attention-over-attention neural networks. In *Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2018*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; pp. 197–206.

18. Fan, F.; Feng, Y.; Zhao, D. Multi-grained attention network for aspect-level sentiment classification. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 3433–3442.

19. 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. 4568–4578.

20. Song, Y.; Wang, J.; Jiang, T.; Liu, Z.; Rao, Y. Attentional encoder network for targeted sentiment classification. *arXiv* **2019**, arXiv:1902.09314.

21. Rietzler, A.; Stabinger, S.; Opitz, P.; Engl, S. Adapt or Get Left Behind: Domain Adaptation through BERT Language Model Finetuning for Aspect-Target Sentiment Classification. *arXiv* **2019**, arXiv:1908.11860.

22. Yang, H.; Zeng, B.; Yang, J.; Song, Y.; Xu, R. A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction. *Neurocomputing* **2021**, *419*, 344–356. [CrossRef]

23. Yang, H.; Zeng, B.; Xu, M.; Wang, T. Back to Reality: Leveraging Pattern-driven Modeling to Enable Affordable Sentiment Dependency Learning. *arXiv* **2021**, arXiv:2110.08604.

24. Wang, R.; Liu, S.; Wang, B.; Xing, S. STC: Stacked Two-stage Convolution for Aspect Term Extraction. In Proceedings of the 2021 International Symposium on Electrical, Electronics and Information Engineering, Seoul, Republic of Korea, 19–21 February 2021; pp. 464–470.

25. Dong, Y.; Wang, J.; Wang, J. Multi-task Learning Network based on Attention for Aspect-Based Sentiment Analysis. In Proceedings of the 6th International Conference on Electronic Technology and Information Science (ICETIS 2021), Harbin, China, 8–10 January 2021; Volume 1827, p. 012173.

26. Xu, H.; Liu, B.; Shu, L.; Yu, P. BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis. 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.