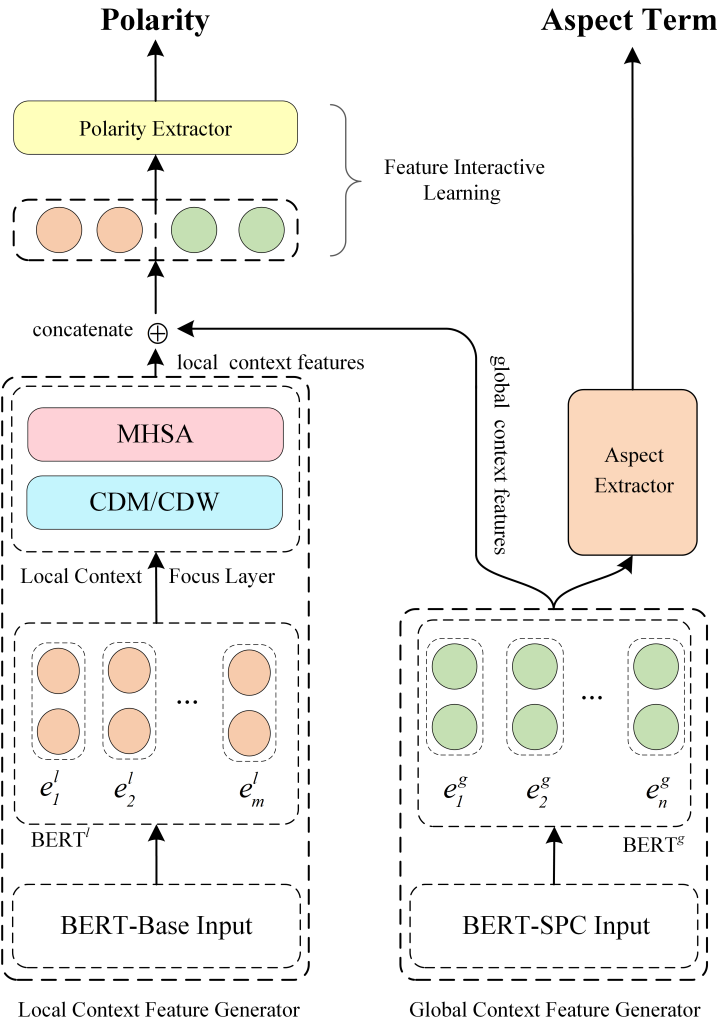# Graphical Abstract

A Multi-task Learning Model for Chinese-oriented Aspect Polarity Classification and Aspect Term Extraction

Heng Yang,Biqing Zeng,JianHao Yang,Youwei Song,Ruyang Xu

Highlights

A Multi-task Learning Model for Chinese-oriented Aspect Polarity Classification and Aspect Term Extraction

Heng Yang,Biqing Zeng,JianHao Yang,Youwei Song,Ruyang Xu

- Proposing a model for the joint task of aspect term extraction and aspect polarity classification.

- The model proposed is Chinese language-oriented and applicable to the English language, with the ability to handle both Chinese and English reviews.

- The model also integrates the domain-adapted BERT model for enhancement.

- The model achieves state-of-the-art performance on seven ABSA datasets.

A Multi-task Learning Model for Chinese-oriented Aspect Polarity Classification and Aspect Term Extraction

Heng Yang[a,*], Biqing Zeng[b], JianHao Yang[b], Youwei Song[c] and Ruyang Xu[a]

[a]*School of Computer, South China Normal University, Guangzhou 510631, China*
[b]*School of Software, South China Normal University, Foshan 528225, China*
[c]*Baidu Inc., Beijing 100085, China*

ARTICLE INFO

ABSTRACT

Aspect-based sentiment analysis (ABSA) task is a multi-grained task of natural language processing and consists of two subtasks: aspect term extraction (ATE) and aspect polarity classification (APC). Most of the existing work focuses on the subtask of aspect term polarity inferring and ignores the significance of aspect term extraction. Besides, the existing researches do not pay attention to the research of the Chinese-oriented ABSA task. Based on the local context focus (LCF) mechanism, this paper firstly proposes a multi-task learning model for Chinese-oriented aspect-based sentiment analysis, namely LCF-ATEPC. Compared with existing models, this model equips the capability of extracting aspect term and inferring aspect term polarity synchronously, moreover, this model is effective to analyze both Chinese and English comments simultaneously and the experiment on a multilingual mixed dataset proved its availability. By integrating the domain-adapted BERT model, the LCF-ATEPC model achieved the state-of-the-art performance of aspect term extraction and aspect polarity classification in four Chinese review datasets. Besides, the experimental results on the most commonly used SemEval-2014 task4 Restaurant and Laptop datasets outperform the state-of-the-art performance on the ATE subtask.

1. Introduction

Aspect-based sentiment analysis Pontiki, Galanis, Pavlopoulos, Papageorgiou, Androutsopoulos and Manandhar (2014); Pontiki, Galanis, Papageorgiou, Manandhar and Androutsopoulos (2015); Pontiki, Galanis, Papageorgiou, Androutsopoulos, Manandhar, AL-Smadi, Al-Ayyoub, Zhao, Qin, De Clercq, Hoste, Apidianaki, Tannier, Loukachevitch, Kotelnikov, Bel, Jiménez-Zafra and Eryiğit (2016) (ABSA) is a fine-grained task compared with traditional sentiment analysis, which requires the model to be able to automatic extract the aspects and predict the polarities of all the aspects. For example, given a restaurant review: "The dessert at this restaurant is delicious but the service is poor," the full-designed model for ABSA needs to extract the aspects "dessert" and "service" and correctly reason about their polarity. In this review, the consumers' opinions on "dessert" and "service" are not consistent, with positive and negative sentiment polarity respectively.

Generally, aspects and their polarity need to be manually labeled before running the aspect polarity classification procedure in the supervised deep learning models. However, most of the proposed models for aspect-based sentiment analysis tasks only focus on improving the classification accuracy of aspect polarity and ignore the research of aspect term extraction. Therefore, when conducting transfer learning on aspect-based sentiment analysis, those proposed models often fall into the dilemma of lacking aspect extraction method on targeted tasks because there is not enough research support.

The APC task is a kind of classification problem. The researches concerning APC tasks is more abundant than the ATE task, and a large number of deep learning-based models have been proposed to solve APC problems, such as the models Vo and Zhang (2015); Wagner, Arora, Cortes, Barman, Bogdanova, Foster and Tounsi (2014); Tang, Qin, Feng and Liu (2016); Wang, Huang, Zhu and Zhao (2016); Ma, Li, Zhang and Wang (2017); Fan, Feng and Zhao (2018)

---

based on long short-term memory (LSTM) and the methodologies Song, Wang, Jiang, Liu and Rao (2019); Zeng, Yang, Xu, Zhou and Han (2019a) based on transformer Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser and Polosukhin (2017). The purpose of the APC task is to predict the exact sentiment polarity of different aspects in their context, rather than to fuzzily analyze the overall sentiment polarity on the sentence-level or document-level. In the APC task, the polarities are most usually classified into three categories: positive, negative, and neutral. It is obvious that the sentiment polarity classified based on aspects can better mine the fine-grained emotional tendency in reviews or tweets, thus providing a more accurate reference for decision-makers.

Similar to the named entity recognition Sang and De Meulder (2003) (NER) task, the ATE task is a sequence labeling task, which aims to extract aspects from the reviews or tweet. In most researches Chen, Sun, Bing and Yang (2017a); Xue and Li (2018); Chen, Xu, He and Wang (2017b), the ATE task is studied independently, away from the APC task. The ATE task first segments a review into separate tokens and then infers whether the tokens belong to any aspect. The tokens may be labeled in different forms in different studies, but most of the studies have adopted the IOB[2] label to annotate tokens.

Dataset	Sample	Apsect	Polarity
Laptop	Great laptop that offers many great **features** !	features	positive
Restaurant	The **seats** are uncomfortable if you are sitting against the wall on wooden benches.	seats	negative
Twitter	how do you like settlers of catan for the **xbox** ?	xbox	neural
Car	**车内顶灯**我觉得灯光效果一般	车内顶灯	negative
Phone	高**分辨率**决定了其拥有比以往采用分辨率的娱乐机型更精细逼真的画面效果	分辨率	positive
Notebook	其致命弱点是**寿命**短而且具有记忆效应	寿命	negative
Camera	出色的**制造工艺**加上强金属质感的拉丝工艺面板	制造工艺	positive

Figure 1: Several samples of the seven ATEPC datasets. All the datasets are domain-specific.

Aiming to automatically extract aspects from the text efficiently and analyze the sentiment polarity of aspects simultaneously, this paper proposes a multi-task learning model for aspect-based sentiment analysis. Multilingual processing is an important research orientation of natural language processing. The LCF-ATEPC[3] model proposed in this paper is a novel multilingual and multi-task-oriented model. Apart from achieving state-of-the-art performance in commonly used SemEval-2014 task4 datasets, the experimental results in four Chinese review datasets also validate that this model has a strong ability to expand and adapt to the needs of multilingual task. The proposed model is based on multi-head self-attention (MHSA) and integrates the pre-trained BERT Devlin, Chang, Lee and Toutanova (2019) and the local context focus mechanism, namely LCF-ATEPC. By training on a small amount of annotated data of aspect and their polarity, the model can be adapted to a large-scale dataset, automatically extracting the aspects and predicting the sentiment polarities. In this way, the model can discover the unknown aspects and avoids the tedious and huge cost of manually annotating all aspects and polarities. It is of great significance for the field-specific aspect-based sentiment analysis.

The main contributions of this article are as follows:

1. For the first time, this paper studies the multi-task model of APC subtask and ATE subtask for multilingual reviews, which provides a new idea for the research of Chinese aspect extraction.
2. This paper firstly applies self-attention and local context focus techniques to aspect word extraction task, and fully explore their potential in aspect term extraction task.
3. The LCF-ATEPC model proposed in this paper integrates the pre-trained BERT model, significantly improves both the performance of ATE task and APC subtask, and achieves new state-of-the-art performance especially the F1 score of ATE task. Besides, we adopted the domain-adapted BERT model trained on the domain-related

---

[2]The labels adopted in this paper are: B_{asp}, I_{asp}, O
[3]The codes for this paper are available at https://github.com/yangheng95/LCF-ATEPC

corpus to the ABSA joint-task learning model. The experimental results show that the domain-adapted BERT model significantly promotes the performance of APC tasks on the three datasets, especially the Restaurant dataset.

4. We designed and applied dual labels for the input sequence applicable for the SemEval-2014 and Chinese review datasets of ABSA joint-task, the aspect term label, and the sentiment polarity label, respectively. The dual label improves the learning efficiency of the proposed model.

2. Related Works

Most ABSA-oriented methodologies regard the ATE and the APC as independent tasks and major in one of them. Accordingly, this section will introduce the related works of ATE and APC in two parts.

2.1. Aspect Term Extraction

The approaches to ATE tasks are classified into two categories: the early dictionary-based or rule-based approaches, and methodologies based on machine-learning or deep learning. Poria, Cambria, Ku, Gui and Gelbukh (2014) proposed a new rule-based approach to extracting aspects from product reviews using common sense and sentence dependency trees to detect explicit and implicit aspects. Liu, Gao, Liu and Zhang (2015) adopts an unsupervised and domain-independent aspect extraction method that relies on syntactic dependency rules and can selects rules automatically.

Compared with manually annotating all aspects in the dataset, the models for ATE can learn the features of aspects and automatically extract aspects in the text, which greatly saves labor and time. Mukherjee and Liu (2012) proposed a model that can extract and cluster aspects simultaneously according to the seed words provided by users for several aspect categories. By classification, synonymous aspects can be grouped into the same category. Poria, Cambria and Gelbukh (2016) proposed the first aspect-oriented deep learning model in opinion mining, which deploys a 7-layer deep convolutional neural network to mark each word in the sentences with opinions as an aspect or non-aspect word. He, Lee, Ng and Dahlmeier (2017) proposed a new method for aspect term extraction, which utilizes word embedding to explore the co-occurrence distribution of words and applies the attention mechanism to weaken the irrelevant words and further improves the coherence of all aspects. Wang, Pan, Dahlmeier and Xiao (2017) proposed a deep neural network-based model namely coupled multilevel attention, which does not require any parser or other linguistic resources to be pre-processed and provides an end-to-end solution. Besides, the proposed model is a multi-layer attention network, where each layer deploys a pair of attentions. This model allows the aspect terms and opinion terms learned interactively and dual propagate during the training process.

For the Chinese-oriented ATE task, a multi-aspect bootstrapping (MAB) method Zhu, Wang, Zhu, Tsou and Ma (2011) is proposed to extract the aspects of Chinese restaurant reviews. Zhao, Dong and Yang (2015a) introduced machine learning methods to explore and extract aspect terms from Chinese hotel reviews. they chose the optimal feature-dimension, feature representation, and maximum entropy (ME) classifier according to the empirical results, and studied the integral effect of aspect extraction.

Up to now, the MHSA and pre-trained model has not been applied in the ATE task. This paper explores the potential of the new techniques of deep learning and new network architecture in the ATE task.

2.2. Aspect Polarity Classification

Aspect polarity classification is another important subtask of ABSA. The approaches designed for the APC task can be categorized into traditional machine learning and recent deep learning methods.The APC task has been comprehensively turned to the the deep neural networks. Therefore, this section mainly introduces approaches based on deep learning techniques.

The most commonly applied deep neural network architectures for APC task are recurrent neural networks Tang et al. (2016); Wang et al. (2016); Ma et al. (2017); Li, Bing, Lam and Shi (2018); Huang, Ou and Carley (2018) (RNNs) and convolutional neural networks (CNNs) Xue and Li (2018); Chen et al. (2017b); Zhang, Zou and Gan (2018). TD-LSTM Tang et al. (2016) first divides the context of aspects into the left and right parts and modeling for them independently. Attention mechanism Bahdanau, Cho and Bengio (2014) has been adapted to APC task in the last few years. ATAE-LSTM takes the feature representation of aspects and context words as the input of the model and applies an attention mechanism to dynamically calculate the attention weight according to the relationship between aspects and context words, and finally predicts the polarity of aspects according to the weighted context

features. Another LSTM-based model IAN Ma et al. (2017) deployed with attention mechanism equips two independent LSTM networks to capture the features of the context and aspect, with interactively integrating and learning the inner correlation of the features of context and targeted aspects. The RAM Chen et al. (2017a) is a bi-directional LSTM-based architecture deploys a multi-layer deep neural network with dedicated memory layers. The multi-layer network utilizes the token features learned based on the attention mechanism and GRUs to finally obtain the global semantic features of the text to predict the sentiment polarities of targeted aspects. In order to retard the loss of context features during the training process, TNet Li et al. (2018) introduced a conventional transformation architecture based on context-preserving transformation (CPT) units. TNet integrates the bidirectional LSTM network and convolutional neural network and significantly improves the accuracy of sentiment polarity prediction. Multi-grained attention network Fan et al. (2018) (MGAN) is a new deep neural network model, which equips with a variety of fine-grained attention mechanisms, and applies the fine-grained attention mechanisms to interactively learn the token-level features between aspects and context, making great use of the inherent semantic correlation of aspects and context.

Peng, Ma, Li and Cambria (2018) proposed the methods for the Chinese language APC task, which conducted the APC task at the aspect level via three granularities. Two fusion methods for the granularities in the Chinese APC task are introduced and applied. Empirical results show that the proposed methods achieved promising performance on the most commonly used ABSA datasets and four Chinese review datasets. Meanwhile, a joint framework aimed to aspect sentiment classification subtask and aspect-opinion pair identification subtask is proposedby Chen and Huang (2019), in which the external knowledge are considered and put into the network to alleviate the problem of insufficient train data. The gated alternate neural network (GANN) Liu and Shen (2019) proposed for APC task aimed to solve the shortcomings of traditional RNNs and CNNs. The GANN applied the gate truncation RNN (GTR) to learn the aspect-dependent sentiment clue representations. Zeng, Ma, Chen and Li (2019b) proposed an end-to-end neural network model for the ABSA task based on joint learning, and the experimental results on a Chinese review show that the proposed model works fine while conducting ATE and APC subtask simultaneously.

BERT-SPC is the BERT text pair classification model, it is a variation model of Bert and is adapted to solve the ABSA task in Song et al. (2019) and achieve high performance. LCF-Bert Zeng et al. (2019a) proposed a feature-level local context focus mechanism based on self-attention, which can be applied to aspect level emotion analysis and many other fine-grained natural language processing tasks. BERT-ADA Rietzler, Stabinger, Opitz and Engl (2019) shows that although the pre-trained model based on a large universal corpus, and is easy to be applied to most tasks and improve performance. Still, it is not task-specific. For specific tasks, if the pre-trained BERT is adapted to specific tasks through the fine-tuning process on a task-related corpus, the task performance can be further improved.

3. Methodology

Aspect-based sentiment analysis relies on the targeted aspects, and most existing studies focus on the classification of aspect polarity, leaving the problem of aspect term extraction. To propose an effective aspect-based sentiment analysis model based on multi-task learning, we adopted domain-adapted BERT model from BERT-ADA and integrated the local context focus mechanism into the proposed model. This section introduces the architecture and methodology of LCF-ATEPC.

This section introduces the methodology of the APC module and the ATE module, respectively. and the contents are organized by order of the network layer hierarchy.

3.1. Task Definition
3.1.1. Aspect Term Extraction
Similar to name entity recognition (NER) task, the ATE task is a kind of sequence labeling task, and prepare the input based on IOB labels. We design the IOB labels as B_{asp}, I_{asp}, O, and the labels indicate the beginning, inside and outside of the aspect terms, respectively. For ATE task, the input of the example review "The price is reasonable although the service is poor." will be prepared as $S = \{w_1, w_2 \cdots w_n\}$, and w stands for a token after tokenization, $n = 10$ is the total number of tokens. The example will be labeled in $Y = \{O, B_{asp}, O, O, O, O, B_{asp}, O, O, O\}$.

3.1.2. Aspect Polarity Classification
Aspect polarity classification is a multi-grained sub-task of sentiment analysis, aiming at predicting the aspect polarity for targeted aspects. Suppose that "The price is reasonable although the service is poor . " is the input for APC task, consistently with ATE task, $S = \{w_1, w_2 \cdots w_n\}$ stands for all the token of the review, and $S^t =$
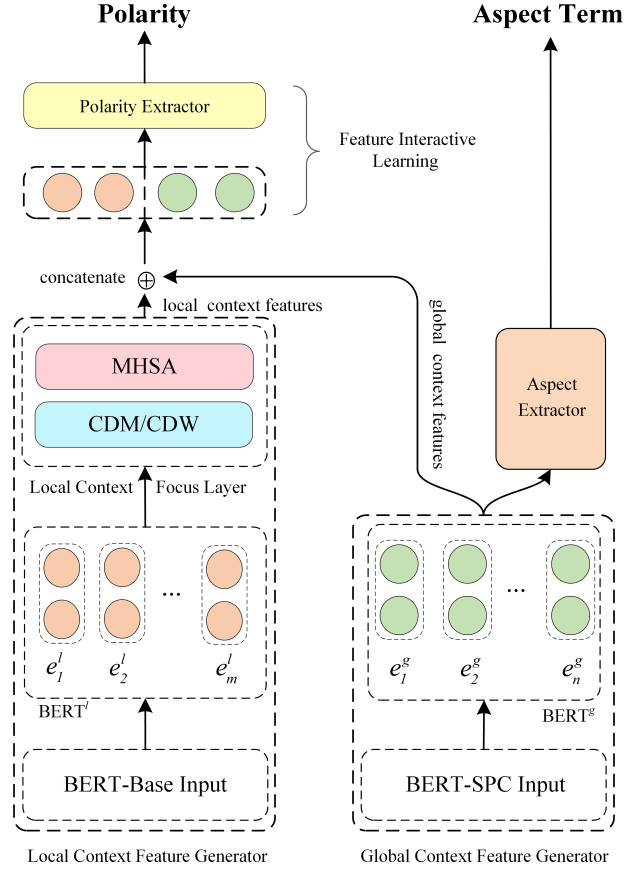
Figure 2: Network architecture of LCF-ATEPC

$\{w_i, w_{i+1} \cdots w_j\}(1 <= i < j <= n)$ is the aspect sequence within S, i and j are the beginning and end positions in S respectively.

3.2. Model Architecture

Aiming at the problem of insufficient research on aspect term extraction task, a joint deep learning model is designed in this section. This model combines aspect polarity classification task and aspect term extraction task, and two independent Bert layers are adopted to model the global context and the local context respectively. For conducting multi-task training at the same time, the input sequences are tokenized into different tokens and the each token is assigned two kinds of label. The first label indicates whether the token belongs to an aspect; the second label marks the polarity of the tokens belongs to the aspect.

Fig 2 is the network architecture of LCF-ATEPC. Local context feature generator (LCFG) unit is on the left and a global context feature generator (GCFG) unit is on the right. Both context feature generator units contain an independent pre-trained BERT layer, $BERT^l$ and $BERT^g$ respectively. The LCFG unit extracts the features of the local context by a local context focus layer and a MHSA encoder. The GCFG unit deploys only one MHSA encoder to learn the global context feature. The feature interactive learning (FIL) layer combines the learning of the interaction between local context features and global context features and predicts the sentiment polarity of aspects. The extraction of aspects based on the features of the global context.

3.2.1. BERT-Shared Layer

The pre-trained BERT model is designed to improve performance for most NLP tasks, and The LCF-ATEPC model deploys two independent BERT-Shared layers that are aimed to extract local and global context features. For pre-trained BERT, the fine-tuning learning process is indispensable. Both BERT-Shared layers are regarded as embed-

ded layers, and the fine-tuning process is conducted independently according to the joint loss function of multi-task learning. X^l and X^g are used to represent the tokenized inputs of LCFG and GCFG respectively, and we can obtain the preliminary outputs of local and global context features.

$$O^l_{BERT} = BERT^l\left(X^l\right) \tag{1}$$

$$O^g_{BERT} = BERT^g\left(X^g\right) \tag{2}$$

O^l_{BERT} and O^g_{BERT} are the output features of the LCFG and the GCFG, respectively. $BERT^l$ and $BERT^g$ are the corresponding BERT-shared layer embedded in the LCFG and the GCFG respectively.

3.3. Multi-Head Self-Attention

Multi-head self-attention is based on multiple scale-dot attention (SDA), which can be utilized to extract deep semantic features in the context, and the features are represented in self-attention score. The MHSA can avoids the negative influence caused by the long distance dependence of the context when learning the features. Suppose X_{SDA} is the input features learned by the LCFG. The scale-dot attention is calculate as follows:

$$\text{SDA}(X_{SDA}) = Softmax\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V \tag{3}$$

$$Q, K, V = f_x(X_{SDA}) \tag{4}$$

$$f_x(X_{SDA}) = \begin{cases} Q = X_{SDA} \cdot W^q \\ K = X_{SDA} \cdot W^k \\ V = X_{SDA} \cdot W^v \end{cases} \tag{5}$$

Q, K and V are the abstract matrices packed from the input features of SDA by three weight matrices $W_q \in \mathbb{R}^{d_h \times d_q}$, $W_k \in \mathbb{R}^{d_h \times d_k}$, $W_v \in \mathbb{R}^{d_h \times d_v}$. The MHSA performs multiple scaled-dot attention in parallel and concatenate the output features, then transform the features by multiplying a vector W^{MH}. h represents the number of the attention heads and equal to 12.

$$MHSA(X) = \tanh\left(\left\{H_1; \dots; H_h\right\} \cdot W^{MH}\right) \tag{6}$$

The ";" means feature concatenation of each head. $W^{MH} \in \mathbb{R}^{hd_v \times d_h}$ is the parameter matrices for projection . Additionally, we apply a tanh activation function for the MHSA learning process, which significantly enhanced feature-capture capability.

3.4. Local Context Focus
3.4.1. Semantic-Relative Distance

The determination of local context depends on semantic-relative distance (SRD), which is proposed to determine whether the context word belongs to the local context of a targeted aspect to help the model capture the local context. Local context is a new concept that can be adapted to most fine-grained NLP tasks. In the ABSA field, existing models generally segment input sequences into aspect sequences and context sequences, treat aspects and context as independent segments and model their characteristics separately. Instead of leaving the aspect alone as part of the input, this paper mines the aspect and its local context, because the empirical result shows the local context of the target aspect contains more important information.

SRD is a concept based on token-aspect pairs, describing how far a token is from the aspect. It counts the number of tokens between each specific token towards a targeted aspect as the SRD of all token-aspect pairs. The SRD is calculated as:

$$SRD_i = |i - P_a| - \lfloor \frac{m}{2} \rfloor \tag{7}$$
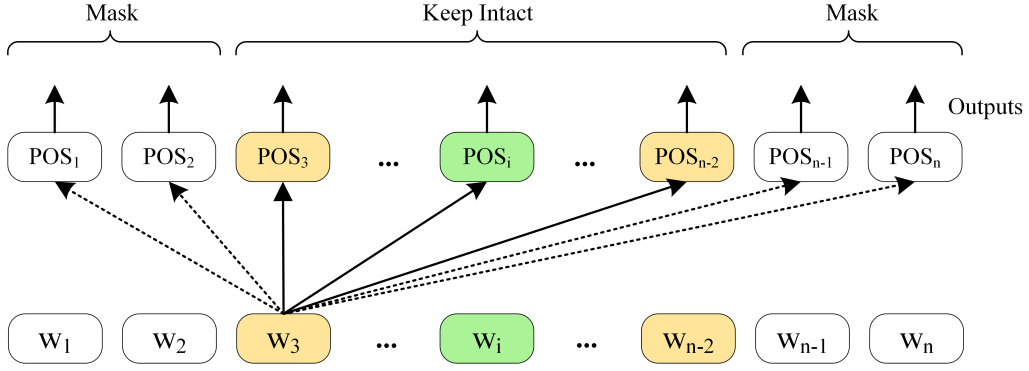
Figure 3: The simulation of the context-feature dynamic mask (CDM) mechanism. The arrows mean the contribution of the token in the computation of the self-attention score to arrowed positions (POS). And the features of the output position that the dotted arrow points to will be masked.

Figure 4: The simulation of the context-feature dynamic weighting (CDW) mechanism. The features of the output position (POS) that the dotted arrow points to will be weighted decay.
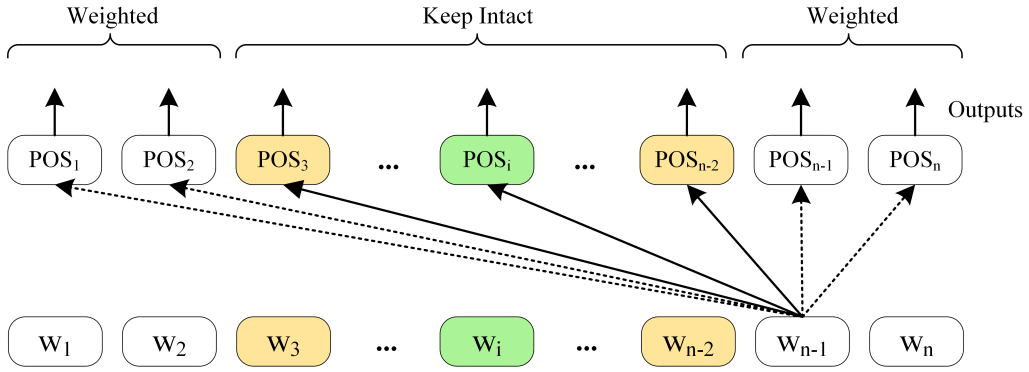
where i ($1 < i < n$) is the position of the specific token, P_a is the central position of aspect. m is the length of targeted aspect, and SRD_i represents for the SRD between the i-th token and the targeted aspect.

Figure 3 and Figure 4 are two implementations of the local context focus mechanism, the context-feature dynamic mask (CDM) layer and context-feature dynamic weighting (CDW) layer, respectively. The bottom and top of the figures represent the feature input and output positions (POS) corresponding to each token. The self-attention mechanism treats all tokens equally, so that each token can generate the self-attention score with other tokens through parallel matrix operation. According to the definition of MHSA, the features of the output position corresponding to each token are more closely related to itself. After calculating the output of all tokens by MHSA encoder, the output features of each output position will be masked or attenuated, except that the local context will be retained intact.

3.4.2. Context-features Dynamic Mask

Apart from to the features of the local context, the CDM layer will mask non-local context's features learned by the $BERT^l$ layer. Although it is easy to directly mask the non-local context words in the input sequence, it is inevitable to discard the features of non-local context words. As the CDM layer is deployed, only a relatively small amount of the semantic context itself will be masked at the corresponding output position. The relative representation of context words and aspects with relatively few semantics is preserved in the corresponding output position.

According to the CDM implementation, the features on all the positions of non-local context words will be set to zero vectors. In order to avoid the unbalanced distribution of features after the CDM operation, an MHSA encoder is

utilized to learn and rebalance the masked local context features. Suppose that the O_{BERT^l} is the preliminary output features of $BERT^l$, then we get the local context feature output as follows,

$$V_i = \begin{cases} E & SRD_i \leq \alpha \\ O & SRD_i > \alpha \end{cases} \tag{8}$$

$$M = \left[V_1^m, V_2^m, \ldots V_n^m \right] \tag{9}$$

$$O_{CDM}^l = O_{BERT^l} \cdot M \tag{10}$$

To mask the features of non-local context, we defines a feature masking matrix M, and V_i^m is the mask vectors for each token in the input sequence. α is the SRD threshold and n is the length of input sequence including aspect. Tokens whose SRD regarding to the targeted aspect is less than the threshold α are the local contexts. The $E \in \mathbb{R}^{d_h}$ represents the ones vector and $O \in \mathbb{R}^{d_h}$ is the zeros vectors. "." denotes the dot-product operation of the vectors.

$$O^l = MHSA \left(O_{CDM}^l \right) \tag{11}$$

Finally the local context features learned by the CDM layer are delivered as O^l.

3.4.3. Context-features Dynamic Weighting

Although empirical results show that the CDM has achieved excellent performance compared with existing models, we design the CDW to explore the potential of LCF mechanism. The CDW is another implementation of the LCF mechanism, takes a more modest strategy compared to the CDM layer, which simply drops the features of the non-local context completely. While the features of local context retained intact, the features of the non-local context words will be weighted decay according to their SRD concerning a targeted aspect.

$$V_i = \begin{cases} E & SRD_i \leq \alpha \\ \frac{n - (SRD_i - \alpha)}{n} \cdot E & SRD_i > \alpha \end{cases} \tag{12}$$

$$W = \left[V_1^w, V_2^w, \ldots V_n^w \right] \tag{13}$$

$$O_{CDW}^l = O_{BERT^l} \cdot W \tag{14}$$

where W is the constructed weight matrix and V_i^w is the weight vector for each non-local context words. Consistently with CDM, SRD_i is the SRD between the i-th context token and a targeted aspect. n is the length of the input sequence. α is the SRD threshold. "." denotes the vector dot-product operation.

$$O^l = MHSA \left(O_{CDW}^l \right) \tag{15}$$

O_{CDW}^l is the output of CDW layer. The CDM and CDW layers are independent, which mean they are alternative. Both the output features of CDM and CDW layers are denoted as O^l. Besides, we tried to concatenate the learned features of CDM and CDW layers and take linear transformation as the features of local context.

$$O_{fusion}^l = [O_{CDM}^l; O_{CDW}^l] \tag{16}$$

$$O_{fusion}^l = W^f \cdot O^f + b^f \tag{17}$$

$$O^l = MHSA \left(O_{fusion}^l \right) \tag{18}$$

W^f, O^f and b^f are weight matrix and bias vector, respectively. The model can choose one of the three approaches to learn the local context features.

3.5. Feature Interactive Learning

LCF-ATEPC does not only rely on local context features for sentiment polarity classification, but combines and learns the local context features and the global context features to conduct polarity classification.

$$O^{lg} = \left[O^l; O^g \right] \tag{19}$$

$$O^{lg}_{dense} = W^{lg} \cdot O^{lg} + b^{lg} \tag{20}$$

$$O^{lg}_{FIL} = MHSA \left(O^{lg}_{dense} \right) \tag{21}$$

O^l and O^g are the local context features and global context features, respectively. $W^{lg} \in \mathbb{R}^{d_h \times 2d_h}$ and $b^{lg} \in \mathbb{R}^{d_h}$ are the weights and bias vectors, respectively. To learn the features of the concatenated vectors, an MHSA encoding process is performed on the O^{lg}_{dense}.

3.6. Aspect Polarity Classifier

Aspect polarity classifier performs a head-pooling on the learned concatenated context features. Head-pooling is to extract the hidden states on the corresponding position of the first token in the input sequence. then a Softmax operation is applied to predict the sentiment polarity.

$$X^{lg}_{pool} = POOL \left(O^{lg}_{FIL} \right) \tag{22}$$

$$Y_{polarity} = \frac{\exp(X^{lg}_{pool})}{\sum_{k=1}^{C} \exp(X^{lg}_{pool})} \tag{23}$$

where C is the number of sentiment categories, and $Y_{polarity}$ represents the polarity predicted by aspect polarity classifier.

3.7. Aspect Term Extractor

Aspect term extractor first performs the token-level classification for each token, suppose T_i is the features on the corresponding position of token T,

$$Y_{term} = \frac{\exp(T_i)}{\sum_{k=1}^{N} \exp(T_i)} \tag{24}$$

where N is the number of token categories, and Y_{term} represents the token category inferred by aspect polarity classifier.

3.8. Training Details

The LCFG and the GCFG are based on the BERT-BASE and BERT-SPC models, respectively. And the BERT-SPC Song et al. (2019) significantly improved the performance of APC tasks. In LCF-ATEPC, BERT-SPC only refactored the input sequence form compared with BERT-BASE model. The input sequence of BERT-BASE is formed in *"[CLS]" + sequence + "[SEP]"*, while it is formed in *"[CLS]" + sequence + "[SEP]" + aspect + "[SEP]"* for BERT-SPC.

Since LCF-ATEPC is a multi-task learning model, we redesigned the form of data input and adopted dual labels of sentiment polarity and token category. The Figure 5 are the input samples of BERT-BASE and BERT-SPC model, respectively.

<center>

O　　O　*B-asp*　*I-asp*　O　　O　　　O　O　O

[CLS]　The　**food　price**　is　reasonable　here　.　[SEP]

Positive

(a) Sample 1

O　　O　　O　　　O　　O　O　*B-asp*　*I-asp*　O　O　*B-asp*　*I-asp*　　O

[CLS]　That　computer　contains　a　large　**hard　disk**　.　[SEP]　**hard　disk**　[SEP]

Positive　　　　　　*Positive*

(b) Sample 2

[CLS] 这部手机的屏幕像素很高，**电池**容量大 [SEP] 电池 [SEP]

Positive

(c) Sample 3

</center>

Figure 5: 5(a) is he input sequence formatted for the BERT-BASE model. The first line is composed of the aspect term labels; The second and third lines are the input sequence after tokenization and polarity label, respectively. 5(b) represent the sample input sequence formatted for BERT-SPC model and the aspects in two position are labeled simultaneously. 5(c) depicts a sample of Chinese sequence input format for BERT-SPC model.

The cross-entropy loss is adopted for APC and ATE subtask and the $\mathbf{L}_2$ regularization is applied in LCF-ATEPC, here is the loss function for APC task,

$$\mathcal{L}_{apc} = \sum_{1}^{C} \widehat{y}_i \log y_i + \lambda \sum_{\theta \in \Theta} \theta^2 \tag{25}$$

where C is the number of polarity categories, λ is the L_2 regularization parameter, and Θ is the parameter-set of the LCF-ATEPC. The loss function for ATE task is

$$\mathcal{L}_{ate} = \sum_{1}^{N} \sum_{1}^{k} \widehat{t}_i \log t_i + \lambda \sum_{\theta \in \Theta} \theta^2 \tag{26}$$

where N is the number of token classes and k is the sum of the tokens in each input sequence. Accordingly, the loss function of LCF-ATEPC is as follows:

$$\mathcal{L}_{atepc} = \mathcal{L}_{apc} + \mathcal{L}_{ate} \tag{27}$$

4. Experiments

4.1. Datasets and Hyperparameters Setting

To comprehensive evaluate the performance of the proposed model, the experiments were conducted in three most commonly used ABSA datasets, the Laptops and Restaurant datasets of SemEval-2014 Task4 subtask2 Pontiki et al. (2014) and an ACL Twitter social dataset Dong, Wei, Tan, Tang, Zhou and Xu (2014). To evaluate our model capability with processing the Chinese language, we also tested the performance of LCF-ATEPC on four Chinese comment datasets Che, Zhao, Guo, Su and Liu (2015); Zhao, Pan, Du and Zheng (2015b); Peng et al. (2018) (Car, Phone, Notebook, Camera). We preprocessed the seven datasets. We reformatted the origin dataset and annotated each

sample with the IOB labels for ATE task and polarity labels for APC tasks, respectively. The polarity of each aspect on the Laptops, Restaurants and datasets may be positive, neutral, and negative, and the conflicting labels of polarity are not considered. The reviews in the four Chinese datasets have been purged, with each aspect may be positive or negative binary polarity. To verify the effectiveness and performance of LCF-ATEPC models on multilingual datasets, we built a multilingual dataset by mixing the 7 datasets. We adopt this dataset to conduct multilingual-oriented ATE and APC experiments.

The table demonstrates the details of these datasets[4].

Table 1
The ABSA datasets for ATE and APC subtasks, including three English datasets and four Chinese datasets.

Datasets	Positive		Negative		Neural	
	Train	Test	Train	Test	Train	Test
Laptop	994	341	870	128	464	169
Restaurant	2164	728	807	196	637	196
Twitter	1561	173	1560	173	3127	346
Car	708	164	213	66	-	-
Phone	1319	341	668	156	-	-
Notebook	328	88	168	35	-	-
Camera	1197	322	546	113	-	-
Multilingual Mixed	8271	2157	4340	867	4228	711

The samples distribution of those datasets is not balanced. For example, most samples in the restaurant dataset are positive, while the neutral samples in the Twitter dataset account for the majority.

Table 2
Global hyperparameters settings for the LCF-ATEPC model, BERT-BASE and BERT-SPC models in the experiments.

Hyperparameters	Setting
learning rate	3×10^{-5}
batch size	16
training epochs	5
max sequence length	80
SRD	5
optimizer	AdamW

Apart from some hyperparameters setting referred to previous researches, we also conducted the controlled trials and analyzed the experimental results to optimize the hyperparameters setting. The superior hyperparameters are listed in Table 2. The default SRD setting for all experiments is 5, with additional instructions for experiments with different SRD.

4.2. Compared Methods

We compare the LCF-ATEPC model to current state-of-the-art methods. Experimental results show that the proposed model achieves state-of-the-art performance both in the ATE and APC tasks.

ATAE-LSTM Wang et al. (2016) is a classical LSTM-based network for the APC task, which applies the attention mechanism to focus on the important words in the context. Besides, ATAE-LSTM appends aspect embedding and the learned features to make full use of the aspect features. The ATAE-LSTM can be adapted to the Chinese review datasets.

ATSM-S Peng et al. (2018) is a baseline model of the ATSM variations for Chinese language-oriented ABSA task. This model learns the sentence and aspect terms at three perspectives of granularity.

GANN is novel neural network model for APC task aimed to solve the shortcomings of traditional RNNs and CNNs.

---

[4]The dataset processed for joint ATE and APC task are available at `https://github.com/yangheng95/LCF-ATEPC`

The GANN applied the Gate Truncation RNN (GTR) to learn informative aspect-dependent sentiment clue representations. GANN obtained the state-of-the-art APC performance on the Chinese review datasets.

AEN-BERT Song et al. (2019) is an attentional encoder network based on the pretrained BERT model, which aims to solve the aspect polarity classification.

BERT-PT Xu, Liu, Shu and Yu (2019) is a BERT-adapted model for Review Reading Comprehension (RRC) task, a task inspired by machine reading comprehension (MRC), it could be adapted to aspect-level sentiment classification task.

BERT-BASE Devlin et al. (2019) is the basic pretrained BERT model. We adapt it to ABSA multi-task learning, which equips the same ability to automatically extract aspect terms and classify aspects polarity as LCF-ATEPC model.

BERT-SPC Song et al. (2019) is a pretrained BERT model designed for the sentence-pair classification task. Consistent with the basic BERT model, we implemented this model for ABSA multitasking.

BERT-ADA Rietzler et al. (2019) is a domain-adapted BERT-based model proposed for the APC task, which fine-tuned the BERT-BASE model on task-related corpus. This model obtained state-of-the-art accuracy on the Laptops dataset.

LCF-ATEPC[5] is the multi-task learning model for the ATE and APC tasks, which is based on the the BERT-SPC model and local context focus mechanism.

LCF-ATE are the variations of the LCF-ATEPC model which only optimize for the ATE task.

LCF-APC are the variations of LCF-ATEPC and it only optimize for the APC task during training process.

4.3. Results Analysis

The experiments are conducted in several segments. First, the baseline performance of LCF-ATEPC on all Chinese and English data sets was tested, and then the effectiveness of multi-task learning was demonstrated. Finally, the assistance of domain-adapted BERT model in improving performance was evaluated and the sensitivity of different datasets to SRD was studied.

4.3.1. Performance on Chinese Review Datasets

Table 3 are the experimental results of LCF-ATEPC models on four Chinese review datasets.

Table 3
The experimental results (%) of LCF-ATEPC models on four Chinese datasets. "-" represents the result is not unreported. The optimal performance is in **bold**.

Model	Car			Phone			Notebook			Camera		
	$F1_{ate}$	Acc_{apc}	$F1_{apc}$	$F1_{ate}$	Acc_{apc}	$F1_{apc}$	$F1_{ate}$	Acc_{apc}	$F1_{apc}$	$F1_{ate}$	Acc_{apc}	$F1_{apc}$
ATAE-LSTM	-	81.90	76.88	-	85.77	83.87	-	83.47	82.14	-	85.54	84.09
ATSM-S	-	82.94	64.18	-	84.86	75.35	-	75.59	60.09	-	82.88	72.50
GANN	-	83.71	77.66	-	89.17	88.16	-	82.65	82.16	-	87.99	86.75
BERT-BASE	98.16	99.13	98.95	98.45	**99.6**	**99.53**	96.55	94.31	93.29	97.66	98.16	97.65
LCF-ATEPC-CDM	99.13	98.7	98.41	99.4	**99.6**	**99.53**	98.99	92.68	91.61	98.86	98.62	98.2
LCF-ATEPC-CDW	99.35	98.7	98.43	**99.5**	99.4	99.3	98.59	**95.12**	94.2	98.86	**98.85**	**98.5**
LCF-ATEPC-Fusion	**99.7**	**99.8**	**99.77**	98.45	**99.6**	**99.53**	99.19	**95.12**	**94.29**	**99.2**	98.16	97.61

4.3.2. Performance on SemEval-2014 task4

Table 4 lists the main experimental results of LCF-ATEPC models to compare the performance with other ABSA-oriented models.

The LCF-ATEPC models are multilingual-oriented. To demonstrate its ability to simultaneously input and analyze reviews in multiple languages, we constructed and experimented with a multilingual dataset fore-mentioned. And result on the multilingual mixed dataset illustrates the effectiveness of the LCF-ATEPC models.

---

[5]We implement our model based on pytorch-transformers: https://github.com/huggingface/pytorch-transformers

---

Table 4

Experimental results (%) of the LCF-ATEPC model. $F1_{ate}$, Acc_{apc} and $F1_{apc}$ are the macro-F1 score of ATE subtask, accuracy and macro-F1 score of the APC subtask. The unreported experimental results are denoted by "-", and the optimal performance is in **bold**.

Model	Laptop			Restaurant			Twitter			Multilingual Mixed		
	$F1_{ate}$	Acc_{apc}	$F1_{apc}$	$F1_{ate}$	Acc_{apc}	$F1_{apc}$	$F1_{ate}$	Acc_{apc}	$F1_{apc}$	$F1_{ate}$	Acc_{apc}	$F1_{apc}$
BERT-PT	-	78.07	75.08	-	84.95	76.96	-	-	-	-	-	-
AEN-BERT	-	79.93	76.31	-	83.12	73.76	-	74.71	73.13	-	-	-
SDGCN-BERT	-	81.35	78.34	-	83.57	76.47	-	-	-	-	-	-
BERT-BASE	80.39	79.78	75.24	80.11	82.77	77.91	98.02	75.43	73.41	92.05	81.98	78.62
BERT-SPC	98.44	80.41	76.82	99.18	85.54	79.19	99.64	76.16	74.48	97.64	82.46	80.07
LCF-ATEPC-CDM	**98.79**	81.35	78.05	**99.38**	86.25	79.63	99.93	**77.02**	**75.62**	**98.73**	**83.19**	**80.58**
LCF-ATEPC-CDW	98.25	80.41	**79.99**	99.29	**86.7**	79.78	**99.96**	76.73	75.44	98.7	82.01	79.16
LCF-ATEPC-Fusion	98.71	**81.5**	77.75	99.33	86.25	**80.26**	99.89	76.73	75.48	97.7	81.85	78.71

4.4. Overall Performance Analysis

Many models for ABSA tasks do not take into account the ATE subtask, but there are still some joint models Nguyen and Shirai (2018) based on the traditional neural network architecture to conduct the APC and ATE tasks simultaneously. Benefit from the joint training process, the two ABSA subtasks of APC and ATE can promote each other and improve the performance.

The CDM layer works better on twitter dataset because there are a lot of non-standard grammar usage and language abbreviations within it, and the local context focus techniques can promote to infer the polarity of terms. Surprisingly, for the Laptop and Restaurant datasets, guests occasionally have a unified "global" view in a specific review. That is, if the customer is not satisfied with one aspect, it is likely to criticize the other. Things will be the same if a customer prefers a restaurant he would be tolerant of some small disamenity, so the CDW mechanism performs better because it does not completely mask the local context of the other aspect. In the multi-task learning process, the convergence rate of APC and ATE tasks is different, so the model does not achieve the optimal effect at the same time.

We build a joint model for the multi-task of ATE and APC based on the BERT-BASE model. After optimizing the model parameters according to the empirical result, the joint model based on BERT-BASE achieved hopeful performance on all three datasets and even surpassed other proposed BERT based improved models on some datasets, such as BERT-PT, AEN-BERT, SDGCN-BERT, and so on. Meanwhile, we implement the joint-task model based on BERT-SPC. Compared with the BERT-BASE model, BERT-SPC significantly improves the accuracy and F1 score of aspect polarity classification. In addition, for the first time, BERT-SPC has increased the F1 score of ATE subtask on three datasets up to 99%.

ATEPC-Fusion is a supplementary scheme of LCF mechanism, and it adopts a moderate approach to generate local context features. The experimental results show that its performance is also better than the existing BERT-based models.

4.4.1. Effectiveness of Multi-task Learning

Keeping the main architecture of the LCF-ATEPC model unchanged, we tried to only optimize parameters for a single task in the multi-task model to explore the difference between the optimal performance of a single task and the multi-task learning model [6].

The Figure 5 depicts the performance of the LCF-ATEPC model when performing an single APC or ATE task. Experimental results show that on some datasets the LCF-ATEPC model performs better concerning APC or ATE single task than conducting ABSA multi-task on some datasets. In general, the proposed model LCF-ATEPC proposed in this paper is still superior to other ABSA-oriented multi-task models and even the single-task models aim to APC or ATE. When optimizing the model parameters for through back-propagation of multiple tasks, the multi-task learning model needs to take into account multiple loss functions of the different subtasks. So sometimes the multi-task

---

[6]The loss function of the LCF-ATEPC is set to $\mathcal{L}_{apc}$ and $\mathcal{L}_{ate}$ while optimizing for APC and ATE task, respectively.

Table 5

The empirical performance comparison between multi-task and single-task learning. The "-" indicates that the statistics are not important during single-task learning optimization and not listed in the table. The optimal performance is in **bold**. "*" indicates the real performance is almost up to 100%.

Model	Laptop			Restaurant			Twitter		
	$F1_{ate}$	Acc_{apc}	$F1_{apc}$	$F1_{ate}$	Acc_{apc}	$F1_{apc}$	$F1_{ate}$	Acc_{apc}	$F1_{apc}$
LCF-ATEPC-CDM	98.79	81.35	78.05	99.38	86.25	79.63	99.93	77.02	**75.62**
LCF-ATEPC-CDW	98.25	80.41	**79.99**	99.29	**86.7**	79.78	99.96	**76.73**	75.44
LCF-ATEPC-Fusion	98.71	81.5	77.75	99.33	86.25	80.26	99.89	**76.73**	75.48
LCF-ATE-CDM	99.14	-	-	**99.55**	-	-	**99.99***	-	-
LCF-APC-CDM	-	80.72	77.16	-	85.8	75.21	-	76.01	74.65
LCF-ATE-CDW	99.14	-	-	**99.55**	-	-	**99.99***	-	-
LCF-APC-CDW	-	**81.97**	78.37	-	86.25	**80.8**	-	76.3	74.29
LCF-ATE-Fusion	**99.22**	-	-	99.49	-	-	99.98	-	-
LCF-APC-Fusion	-	80.56	77.41	-	85.0	78.4	-	74.86	73.74

learning cannot achieve as the best effect as single-task learning does, which is also the compromise of the multi-task learning model when dealing with multiple tasks.

4.4.2. Domain-adaption for LCF-ATEPC

Table 6

Experimental results of LCF-ATEPC on Laptop and Restaurant data sets using domain-adapted pretrained BERT.

| Model	Laptop			Restaurant		
	$F1_{ate}$	Acc_{apc}	$F1_{apc}$	$F1_{ate}$	Acc_{apc}	$F1_{apc}$
BERT-BASE	78.86	79.94	76.26	78.43	85.45	77.37
BERT-SPC	95.46	80.09	75.96	95.97	89.38	84.6
BERT-ADA	-	80.23	75.77	-	87.14	80.09
LCF-ATEPC-CDM	97.81	**82.29**	78.57	**99.2**	**90.18**	**85.88**
LCF-ATEPC-CDW	**98.51**	81.66	77.68	99.04	89.38	84.57
LCF-ATEPC-Fusion	98.44	81.97	**78.66**	99.15	88.48	82.79

The BERT-BASE model is trained on a large-scale general corpus, so the fine-tuning during process during training process is significant and inevitable for BERT-based models. Meanwhile, the ABSA datasets commonly benchmarked are generally small with the domain-specific characteristic, the effect of BERT-BASE model on the most ABSA datasets can be further improved through domain-adaption. Domain adaption is a effective technique while integrating the pre-trained BERT-BASE model. By further training the BERT-BASE model in a domain-related corpus similar to or homologous to the target ABSA dataset, then domain-related pretrained BERT model can be obtained. We adopted the method proposed in Rietzler et al. (2019) to obtain the domain-adapted pre-trained BERT model based on the corpus of Yelp Dataset Challenge reviews[7] and the amazon Laptops review datasetHe and McAuley (2016). Table 6 shows that the performance of APC task significantly improved by domain-adapted BERT model. The accuracy benchmark in the classical Restaurant achieving more than 90%, which means that the LCF-ATEPC is the first ABSA-oriented model obtained up to 90% accuracy on the Restaurant dataset. In addition, experimental result on the Laptop dataset also prove the effectiveness of domain-adaption in multi-task learning. Besides, the experimental results on the laptop dataset also validate the effectiveness of domain-adapted BERT model for ABSA multi-task learning.

---

[7]This corpus are available at https://www.yelp.com/dataset/challenge

### 4.4.3. SRD Sensitivity on Different Datasets

We tested the sensitivity of SRD threshold on the typical Chinese and English ABSA datasets: the Phone dataset and The Restaurant dataset, respectively. Besides, for the evaluation of the restaurant dataset, we adopted the domain-adapted BERT model as the underlying architecture of the LCF-ATEPC model. The experimental result of Figure 6, 7 are evaluated in multi-task learning process.
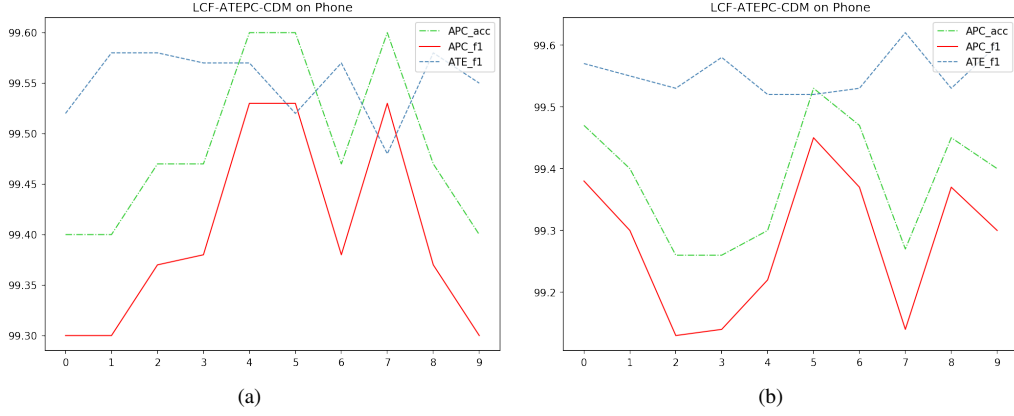
Figure 6: The 6(a), 6(b) are the performance visualization of LCF-ATEPC-CDM and LCF-ATEPC-CDW on the Chinese Phone dataset under different SRDs, respectively.

For the Chinese Phone dataset, the LCF-ATEPC-CDM model can achieve the best APC accuracy and F1 score when the SRD threshold is about 4-5, while the best ATE task performance reaches the highest when the SRD threshold is about 1-3. The LCF-ATEPC-CDW model obtains the best APC performance on the Phone dataset when the SRD threshold is 5, while the best ATE F1 score is approximately obtained when the SRD threshold is 7.

For the Restaurant dataset, the optimal APC accuracy and F1 score achieved by LCF-ATEPC-CDM while the SRD threshold is approximately between 4 and 6. While the SRD threshold for the LCF-ATEPC-CDW is set to 8, the model achieves the optimal aspect classification accuracy and F1 score. However, the F1 score of the ATE task is less sensitive to the SRD threshold, indicating that aspect polarity classification task has less assistance on it during the joint learning process.
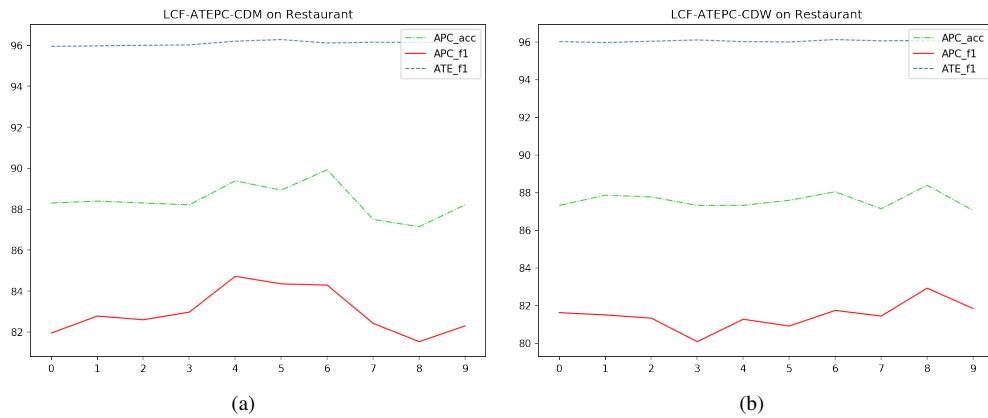
Figure 7: The 7(a), 7(b) are the performance visualization of LCF-ATEPC-CDM and LCF-ATEPC-CDW on the Restaurant dataset under different SRDs, respectively.

5. Conclusion

The ATE and APC subtasks were treated as independent tasks in previous studies. Moreover, the multi-task learning model for ATE and APC subtasks has not attracted enough attention from researchers. Besides, the researches concerning the Chinese language-oriented ABSA task are not sufficient and urgent to be proposed and developed. To address the above problems, this paper proposes a multi-task learning model LCF-ATEPC for aspect-based sentiment analysis based on the MHSA and the LCF mechanisms and applies the pre-trained BERT to the ATE sub-tasks for the first time. Not only for the Chinese language, but the models proposed in this paper are multilingual and applicable to the classic English review sentiment analysis task, such as the SemEval-2014 task4. The proposed model can automatically extract aspects from reviews and infer aspects' polarity. Empirical results on 3 commonly English datasets and four Chinese review datasets for ABSA tasks show that, compared with all models based on basic BERT, the LCF-ATEPC model achieves state-of-the-art performance on ATE and APC tasks.

Acknowledgments and Funding

References

Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 URL: https://arxiv.org/abs/1409.0473.

Che, W., Zhao, Y., Guo, H., Su, Z., Liu, T., 2015. Sentence compression for aspect-based sentiment analysis. IEEE/ACM Transactions on audio, speech, and language processing 23, 2111–2124. doi:10.1109/TASLP.2015.2443982.

Chen, F., Huang, Y., 2019. Knowledge-enhanced neural networks for sentiment analysis of chinese reviews. Neurocomputing 368, 51 – 58. URL: http://www.sciencedirect.com/science/article/pii/S0925231219311920, doi:https://doi.org/10.1016/j.neucom.2019.08.054.

Chen, P., Sun, Z., Bing, L., Yang, W., 2017a. Recurrent attention network on memory for aspect sentiment analysis, in: Proceedings of the 2017 conference on empirical methods in natural language processing, pp. 452–461. doi:10.18653/v1/D17-1047.

Chen, T., Xu, R., He, Y., Wang, X., 2017b. Improving sentiment analysis via sentence type classification using bilstm-crf and cnn. Expert Systems with Applications 72, 221–230. doi:10.1016/j.eswa.2016.10.065.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota. pp. 4171–4186. URL: https://www.aclweb.org/anthology/N19-1423, doi:10.18653/v1/N19-1423.

Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., Xu, K., 2014. Adaptive recursive neural network for target-dependent twitter sentiment classification, in: Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers), pp. 49–54. doi:10.3115/v1/p14-2009.

Fan, F., Feng, Y., Zhao, D., 2018. Multi-grained attention network for aspect-level sentiment classification, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3433–3442. doi:10.18653/v1/D18-1380.

He, R., Lee, W.S., Ng, H.T., Dahlmeier, D., 2017. An unsupervised neural attention model for aspect extraction, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 388–397. doi:10.18653/v1/p17-1036.

He, R., McAuley, J., 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering, in: Proceedings of the 25th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland. pp. 507–517. URL: https://doi.org/10.1145/2872427.2883037, doi:10.1145/2872427.2883037.

Huang, B., Ou, Y., Carley, K.M., 2018. Aspect level sentiment classification with attention-over-attention neural networks, in: Thomson, R., Dancy, C., Hyder, A., Bisgin, H. (Eds.), Social, Cultural, and Behavioral Modeling, Springer International Publishing, Cham. pp. 197–206. URL: https://link.springer.com/chapter/10.1007/978-3-319-93372-6_22.

Li, X., Bing, L., Lam, W., Shi, B., 2018. Transformation networks for target-oriented sentiment classification, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 946–956. doi:10.18653/v1/p18-1087.

Liu, N., Shen, B., 2019. Aspect-based sentiment analysis with gated alternate neural network. Knowledge-Based Systems , 105010doi:10.1016/j.knosys.2019.105010.

Liu, Q., Gao, Z., Liu, B., Zhang, Y., 2015. Automated rule selection for aspect extraction in opinion mining, in: Proceedings of the 24th International Conference on Artificial Intelligence, AAAI Press. pp. 1291–1297. URL: http://dl.acm.org/citation.cfm?id=2832415.2832429.

Ma, D., Li, S., Zhang, X., Wang, H., 2017. Interactive attention networks for aspect-level sentiment classification, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, pp. 4068–4074. URL: https://doi.org/10.24963/ijcai.2017/568, doi:10.24963/ijcai.2017/568.

Mukherjee, A., Liu, B., 2012. Aspect extraction through semi-supervised modeling, in: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1, Association for Computational Linguistics, Stroudsburg, PA, USA. pp. 339–348. URL: http://dl.acm.org/citation.cfm?id=2390524.2390572.

Nguyen, H., Shirai, K., 2018. A joint model of term extraction and polarity classification for aspect-based sentiment analysis, in: 2018 10th International Conference on Knowledge and Systems Engineering (KSE), IEEE. pp. 323–328. doi:10.1109/kse.2018.8573340.

Peng, H., Ma, Y., Li, Y., Cambria, E., 2018. Learning multi-grained aspect target sequence for chinese sentiment analysis. Knowledge-Based Systems 148, 167–176. doi:10.1016/j.knosys.2018.02.034.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S.M., Eryiğit, G., 2016. SemEval-2016 task 5: Aspect based sentiment analysis, in: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics, San Diego, California. pp. 19–30. URL: https://www.aclweb.org/anthology/S16-1002, doi:10.18653/v1/S16-1002.

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., Androutsopoulos, I., 2015. SemEval-2015 task 12: Aspect based sentiment analysis, in: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado. pp. 486–495. URL: https://www.aclweb.org/anthology/S15-2082, doi:10.18653/v1/S15-2082.

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., 2014. SemEval-2014 task 4: Aspect based sentiment analysis, in: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Association for Computational Linguistics, Dublin, Ireland. pp. 27–35. URL: https://www.aclweb.org/anthology/S14-2004, doi:10.3115/v1/S14-2004.

Poria, S., Cambria, E., Gelbukh, A., 2016. Aspect extraction for opinion mining with a deep convolutional neural network. Knowledge-Based Systems 108, 42–49. doi:10.1016/j.knosys.2016.06.00.

Poria, S., Cambria, E., Ku, L.W., Gui, C., Gelbukh, A., 2014. A rule-based approach to aspect extraction from product reviews, in: Proceedings of the second workshop on natural language processing for social media (SocialNLP), pp. 28–37. doi:10.3115/v1/w14-5905.

Rietzler, A., Stabinger, S., Opitz, P., Engl, S., 2019. Adapt or get left behind: Domain adaptation through bert language model finetuning for aspect-target sentiment classification. arXiv preprint arXiv:1908.11860 URL: https://arxiv.org/abs/1908.11860.

Sang, E.F.T.K., De Meulder, F., 2003. Introduction to the conll-2003 shared task: language-independent named entity recognition , 142–147doi:10.1016/j.eswa.2016.10.065.

Song, Y., Wang, J., Jiang, T., Liu, Z., Rao, Y., 2019. Attentional encoder network for targeted sentiment classification. arXiv preprint arXiv:1902.09314 URL: https://arxiv.org/abs/1902.09314.

Tang, D., Qin, B., Feng, X., Liu, T., 2016. Effective LSTMs for target-dependent sentiment classification, in: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, The COLING 2016 Organizing Committee, Osaka, Japan. pp. 3298–3307. URL: https://www.aclweb.org/anthology/C16-1311.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, Curran Associates Inc., USA. pp. 6000–6010. URL: http://dl.acm.org/citation.cfm?id=3295222.3295349.

Vo, D.T., Zhang, Y., 2015. Target-dependent twitter sentiment classification with rich automatic features, in: Proceedings of the 24th International Conference on Artificial Intelligence, AAAI Press. pp. 1347–1353. URL: http://dl.acm.org/citation.cfm?id=2832415.2832437.

Wagner, J., Arora, P., Cortes, S., Barman, U., Bogdanova, D., Foster, J., Tounsi, L., 2014. Dcu: Aspect-based polarity classification for semeval task 4 doi:10.3115/v1/s14-2036.

Wang, W., Pan, S.J., Dahlmeier, D., Xiao, X., 2017. Coupled multi-layer attentions for co-extraction of aspect and opinion terms, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI Press. pp. 3316–3322. URL: http://dl.acm.org/citation.cfm?id=3298023.3298050.

Wang, Y., Huang, M., Zhu, X., Zhao, L., 2016. Attention-based lstm for aspect-level sentiment classification, in: Proceedings of the 2016 conference on empirical methods in natural language processing, pp. 606–615. doi:10.18653/v1/d16-1058.

Xu, H., Liu, B., Shu, L., Yu, P., 2019. 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, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota. pp. 2324–2335. URL: https://www.aclweb.org/anthology/N19-1242, doi:10.18653/v1/N19-1242.

Xue, W., Li, T., 2018. Aspect based sentiment analysis with gated convolutional networks, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia. pp. 2514–2523. URL: https://www.aclweb.org/anthology/P18-1234, doi:10.18653/v1/P18-1234.

Zeng, B., Yang, H., Xu, R., Zhou, W., Han, X., 2019a. Lcf: A local context focus mechanism for aspect-based sentiment classification. Applied Sciences 9, 3389. doi:10.3390/app9163389.

Zeng, Z., Ma, J., Chen, M., Li, X., 2019b. Joint learning for aspect category detection and sentiment analysis in chinese reviews, in: China Conference on Information Retrieval, Springer. pp. 108–120. doi:10.1007/978-3-030-31624-2_9.

Zhang, Z., Zou, Y., Gan, C., 2018. Textual sentiment analysis via three different attention convolutional neural networks and cross-modality consistent regression. Neurocomputing 275, 1407 – 1415. URL: http://www.sciencedirect.com/science/article/pii/S0925231217316090, doi:https://doi.org/10.1016/j.neucom.2017.09.080.

Zhao, Y., Dong, S., Yang, J., 2015a. Effect research of aspects extraction for chinese hotel reviews based on machine learning method. International Journal of Smart Home 9, 23–34. doi:10.14257/ijsh.2015.9.3.03.

Zhao, Y., Pan, H., Du, C., Zheng, Y., 2015b. Principal curvature for infrared small target detection. Infrared Physics & Technology 69, 36 – 43. URL: http://www.sciencedirect.com/science/article/pii/S1350449514002825, doi:10.1016/j.infrared.2014.12.014.

Zhu, J., Wang, H., Zhu, M., Tsou, B.K., Ma, M., 2011. Aspect-based opinion polling from customer reviews. IEEE Transactions on Affective Computing 2, 37–49. doi:10.32657/10356/61830.