Knowledge-Based Systems 264 (2023) 110326

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Aspect-based sentiment analysis via multitask learning for online reviews

Guoshuai Zhao [a,*], Yiling Luo [a,1], Qiang Chen [a,1], Xueming Qian [b,c]

[a] School of Software Engineering, Xi'an Jiaotong University, Xi'an, 710049, China
[b] School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, 710049, China
[c] Key Laboratory for Intelligent Networks and Network Security, Ministry of Education, Xi'an, 710049, China

A R T I C L E I N F O

Article history:
Received 18 September 2022
Received in revised form 8 January 2023
Accepted 18 January 2023
Available online 21 January 2023

Keywords:
Aspect-based sentiment analysis
Aspect polarity classification
Aspect term extraction
BERT
Relational graph attention network
Multi-head attention

A B S T R A C T

Aspect based sentiment analysis(ABSA) aims to identify aspect terms in online reviews and predict their corresponding sentiment polarity. Sentiment analysis poses a challenging fine-grained task. Two typical subtasks are involved: Aspect Term Extraction (ATE) and Aspect Polarity Classification (APC). These two subtasks are usually trained discretely, which ignores the connection between ATE and APC. Concretely, we can relate ATE to APC through aspects and train them concurrently. We mainly use the ATE task as an auxiliary task, allowing the APC to focus more on relevant aspects to facilitate aspect polarity classification. In addition, previous studies have shown that utilizing dependency syntax information with a graph neural network (GNN) also contributes to the performance of the APC task. However, most studies directly input sentence dependency relations into graph neural networks without considering the influence of aspects, which do not emphasize the important dependency relationships. To address these issues, we propose a multitask learning model combining APC and ATE tasks that can extract aspect terms as well as classify aspect polarity simultaneously. Moreover, we exploit multihead attention(MHA) to associate dependency sequences with aspect extraction, which not only combines both ATE and APC tasks but also stresses the significant dependency relations, enabling the model to focus more on words closely related to aspects. According to our experiments on three benchmark datasets, we demonstrate that the connection between ATE and APC can be better established by our model, which enhances aspect polarity classification performance significantly. The source code has been released on GitHub https://github.com/winder-source/MTABSA.

1. Introduction

Aspect based sentiment analysis (ABSA) aims to mine sentiment information toward a given sentence, but is fine-grained. Specifically, its goal is to identify aspect terms in a comment and predict their corresponding sentiment polarity. In the example, "I like the service in the restaurant, but the environment is not very good", the aspect terms are "service" and "environment". The output emotional polarity of the two aspects is *positive* and *negative*. The sentiments corresponding to these two aspects are quite opposite, so it is not appropriate to conduct a sentiment analysis of the whole sentence but to conduct a more fine-grained analysis. The main research line of ABSA focuses on two subtasks, namely, ATE and APC.

The APC task is usually considered a classification task, or sentiment classification of a given aspect in a sentence. The approach to solving APC tasks has evolved from feature engineering to deep learning-based methods. The most common deep neural network architectures used in APC tasks are convolutional neural networks (CNNs) and recurrent neural networks (RNNs) [1,2]. Moreover, the application of attention mechanisms in neural networks is becoming increasingly extensive. The attention mechanism [3,4] is also suitable for ABSA tasks. In recent years, SOTA results have been obtained in numerous NLP tasks with proposed pretraining models. Therefore, many studies are based on the pretraining model, such as the AEN model [5] and BERT-PT [6]. In addition, a sentence contains not only semantic information but also syntactic structure information, such as dependency tree structure. Intuitively, it is helpful to integrate syntactic structure information into the APC task because the syntactic structure can better capture sentiment words related to aspect. Recently, many methods have regarded the dependency tree as an adjacency matrix and utilized GNNs to encode the entire adjacency matrix, such as graph attention networks [7] (GAT) and graph convolutional networks [8,9] (GCN).

In most of these studies, the ATE task was studied independently. The ATE task is regarded as a NER task aimed at extracting

* Corresponding author.
E-mail address: guoshuai.zhao@xjtu.edu.cn (G. Zhao).

1 They have equal contributions to this work.

https://doi.org/10.1016/j.knosys.2023.110326
0950-7051/

aspects of the sentences as a sequence labeling task [10–12]. The advancement of deep learning has shown its usefulness in tasks. Recent methods use deep neural networks to assist aspect extraction [13–15]. Furthermore, there are many models based on BERT to perform sequence labeling tasks due to the success of the model.

Other approaches for sentiment analysis [16,17] have emerged in recent years, with meta-based self-training sentiment analysis [18] and prompt-based sentiment analysis [19,20] being proposed to perform sentiment analysis more efficiently, and new trends in neurosymbolic AI for explainable sentiment analysis [21–24] have also emerged.

In addition, there are also some works that mainly consider multitask learning in ABSA [25–27] to achieve better performance with the help of interactions between tasks. Based on the inspiration of multitask learning, we also propose a multitask learning model that combines APC and ATE tasks. In our model, the APC task's performance is further boosted by using the feature of the ATE task. Inspired by the relational graph attention network (RGAT) presented by Wang et al. [28], we also use a series of RGAT processes to encode the reshaped and pruned dependency tree. Although previous studies have shown that graph neural networks contribute to the performance of APC tasks, most previous studies have fed sentence dependencies directly into graph neural networks. Such dependencies do not consider the influence of aspects. Therefore, to address the challenge of dependency relations, we apply MHA to associate dependency sequences with aspect extraction, enabling our model to focus on the dependency sequences that are more closely related to the aspects. To validate the effectiveness of the proposed model, extensive experiments are conducted on three public datasets, and according to the experimental results, our model has obvious improvements and achieves superior performance.

Our main contributions are as follows:

- We propose a multitask learning model that integrates BERT and RGAT models for APC and ATE tasks. The two tasks are conducted simultaneously in a joint training manner.
- We propose to associate dependency sequences with aspect extraction via MHA, which can enhance the connection between aspects and their associated dependency sequences.
- Three public datasets were used to confirm the validity of the model. As seen from the experimental results, the proposed model outperforms recent state-of-the-art models. We further conduct domain-adaptation experiments, achieving appealing results.

2. Related work

2.1. Aspect term extraction

As a subtask of ABSA, aspect term extraction works to identify different aspects mentioned in a given sentence. Aspect terms refer to specific characteristics or attributes of products or services discussed in the review. Aspect term extraction can be regarded as a textual entity. Aspect term extraction methods have undergone a development phase from traditional methods to deep learning methods.

2.1.1. Traditional methods

Traditional methods often require manual annotation of aspects, typically rely on the use of handcrafted features and rules, are time-consuming, and are often very complicated. Poria et al. [29] presented a novel rule-based approach. They use sentence dependency trees and common sense to extract aspects and detect explicit and implicit aspects. A novel approach was implemented by Liu et al. [30] to select rules automatically, which is unsupervised and domain-independent and can select effective rule sets.

2.1.2. Machine learning-based methods

Compared with traditional methods, methodologies based on machine learning or deep learning for ATE can learn aspect features and extract aspect terms directly from a given text automatically, which saves time and manpower. Machine learning includes both supervised and unsupervised methods. Supervised methods are used to train manually annotated data to extract aspects from comments. The commonly used supervised techniques are decision trees, support vector machines, and K-nearest neighbor [31]. However, unsupervised methods do not need annotated data, instead using syntactic and contextual patterns to extract aspect features automatically.

The NER problem widely uses the conditional random field, which is a supervised-based method. Because the ATE task is similar to the NER task, CRF is also widely used in aspect extraction tasks. Lample et al. [10] combined LSTM and CRF and proposed a training mode that used a large unlabeled corpus combined with a small amount of supervised data, using the biLSTM+CRF model for training.

2.1.3. Deep learning-based methods

The above methods have shortcomings. The CRF is a linear model that requires many features to work well, and the language patterns need to be crafted manually. In recent years, deep learning has been widely used, has become mainstream and has achieved good performance in aspect extraction tasks. Da'u et al. [13] utilized multichannel cnn to extract aspects and encode contextual information using the word embedding channel and POS Tag embedding channel to enhance the sequential tagging of words, respectively. Zhang et al. [15] presented a topic-aware dynamic cnn using dynamically generated filters to learn more significant features concerning aspects.

2.2. Aspect polarity classification

APC aims to analyze the sentiment polarity associated with each aspect. In recent years, the approaches for the APC task mainly include conventional machine learning methods and deep learning methods. Deep learning-based methods are gradually becoming mainstream in APC tasks, which are introduced in this section.

2.2.1. Convolutional neural network-based methods

CNNs are deep neural network architectures that are commonly applied in APC tasks. Li et al. [32] proposed transformation networks to transform embeddings and used a CNN layer for encoding. Xue and Li [1] extract aspect-specific features by combining convolutional neural networks with gating mechanisms.

2.2.2. Attention mechanism-based methods

There are also some attempts leveraging the attention mechanism [33,34] to focus on more significant information. An interactive attention network was used by Ma et al. [35] to learn attention in aspects and contexts and to model aspect-context relationships. Huang et al. [36] proposed an attention-over-attention neural network (AOA) to jointly learn aspects and sentence representations. A self-supervised attention learning method that can mine useful information about attentional supervision and refines attentional mechanisms automatically was proposed by Tang et al. [37]. A self-supervised attention learning method was presented by Su et al. [38] for continuously learning useful attention supervision information.

2.2.3. Pretrained language model-based methods

In recent years, the pretrained language model BERT [39] has produced promising results on many tasks in the field of NLP, including APC. Song et al. [5] employed attention to model the interaction between targets and contexts, using BERT to encode target mentions and surrounding contexts. Zhao et al. [40] developed a knowledge-enabled language representation model BERT to leverage a sentiment knowledge graph.

2.2.4. Syntax-based methods

Syntactic information is also available for the task. The early work relied on the manual definition of some syntactic rules and was labor intensive. Neural network models were used later. The adaptive recurrent neural network (AdaRNN) was developed by Dong et al. [41]. AdaRNN can encode dependency trees and propagate word sentiments to the target aspect adaptively. Considering both the complementarity and semantic relevance of syntactic structures, Li et al. [42] designed a module with rich syntactic knowledge to reduce dependency parsing errors.

2.2.5. Graph neural network-based methods

Recently, graph neural networks have been widely used to encode dependency trees, showing promising results. Zhang et al. [43] combined GCN with dependency trees of sentences to learn the representation of dependencies. Wang et al. [28] proposed a relational graph attention network to encode a new dependency parse tree that was reshaped and pruned. Zeng et al. [44] presented a new relation-building multitask learning network using graph convolution networks to encode aspect representations. Contextual information, semantic relations, and syntactic structures were used simultaneously to build graph convolutional networks by Phan et al. [45]. To enhance the current graph convolutional network for modeling sentiment dependencies, Zhao et al. [46] proposed an aggregated graph convolutional network (AGCN). Feng et al. [47] designed a novel model AG-VSR using variable sentence representation and the attention-assisted graph-based representation generated by GCN to implement aspect sentiment classification.

2.3. Differences with existing methods

In addition to the single task mentioned above, there are some works considering various subtasks of ABSA and using multitask learning to implement ABSA. Focusing on both ATE and APC tasks and considering them as sequence labeling tasks, Luo et al. [48] proposed a new RNN framework capable of generating aspect-polarity pairs. He et al. [49] proposed an interactive multitask learning network (IMN) that can jointly learn multiple related tasks at the token level and document level simultaneously. Chen and Qian. [25] focused on three subtasks of ABSA and explored their interactions to build a relation-aware collaborative learning framework (RACL). Luo et al. [50] extended the gradient coordination mechanism to ABSA and designed a cascaded annotation module to address the imbalance between ATE and APC as sequence labeling tasks. Liang et al. [51] migrate document-level knowledge of specific domains and related sentiments to aspect-level subtasks and explore the interactive associations between subtasks to improve performance. A new end-to-end ABSA-dependent syntactic knowledge-enhanced interaction architecture based on multitask learning was implemented by Liang et al. [52] by exploiting message passing mechanisms and graph convolutional networks.

Instead of a single-task ATE or APC, we also propose a multitask learning model to associate APC with ATE. However, unlike the other multitask learning models mentioned above, we use graphical representations, dependency relations, and attention mechanisms, mainly focusing on the performance enhancement of APC tasks by treating ATE tasks as auxiliary tasks. Intuitively, the feature extracted by the ATE task is conducive to the APC task. Therefore, we exploit MHA to further obtain the relations between aspects and their corresponding dependency relations and focus on the important dependencies. We also integrate BERT and RGAT models to implement our model. In contrast to the above work, we apply multitask learning to aspect-level sentiment analysis and use dependencies and attention mechanisms to focus more on relevant aspects, using the ATE task to assist the APC task to achieve better performance improvement of the APC task.

3. Our model

We introduce our multitask learning model in detail in this section. It consists of four main parts: BERT-APC, RGAT, BERT-ATE, and MHA. Fig. 1 shows the overall architecture of our model. The input of our model includes three parts: input sentences and aspects into the BERT-APC module simultaneously; input dependency relations into the RGAT module; and input the sentence into the BERT-ATE module. The model has two outputs: the extracted aspect and the aspect polarity. BERT-APC is used to extract the global features of sentences and aspects, so the input of BERT-APC contains the whole sentence and the target aspects in the sentence, which are separated by the separator token [SEP]. RGAT is applied to encode the aspect-oriented dependency tree [28] of sentences. BERT-ATE can be regarded as an aspect extractor, but it can not only extract the target aspects from sentences but also provide the relevant features to assist aspect polarity classification, allowing the model to leverage feature interactions and information to obtain better results. The multihead attention module combines the two tasks of ATE and APC, integrating dependency labels into the features extracted by BERT-ATE to emphasize the important dependency relations.

3.1. Problem formulation

3.1.1. Aspect term extraction

We treat the ATE task as a sequence labeling task. We preprocess the text with IOB tags, identifying each word in the sentence as a different token. We use B_{asp} to identify the first word of the aspect, I_{asp} to identify the internal words of the aspect, and O to identify the rest of the words. For example, consider the sentence "best spicy tuna roll, great Asian salad". This sentence will be processed as $S = (w_1, w_2, \dots, w_n)$, where w represents a token and $n = 9$ represents the total number of words. It will be identified as $Y = (O, B_{asp}, I_{asp}, I_{asp}, O, O, B_{asp}, I_{asp}, O)$.

3.1.2. Aspect polarity classification

The APC task is a fine-grained sentiment analysis task that aims to analyze the sentiment of different aspects of the comments. Suppose there is a sentence: "The food is surprisingly good, but the decor is not very good". We use $S = (w_1, w_2, \dots, w_n)$ to represent the sentence, where w represents each word, and the $\langle T, P\rangle$ tuple represents the aspect and the corresponding sentiment polarity. $T = (w_i, w_{i+1}, \dots, w_j)$ represents an aspect, where i and j represent the starting and ending positions of the aspect in the sentence, respectively. $P = p,\ p \in \{Positive, Negative, Neutral\}$. In this sentence, there are two tuples $\langle (food), Positive\rangle$ and $\langle (decor), Negative\rangle$.

3.2. BERT-APC

BERT is a pretraining model, and on many tasks, it has obtained SOTA results. In our model, we use BERT to extract the

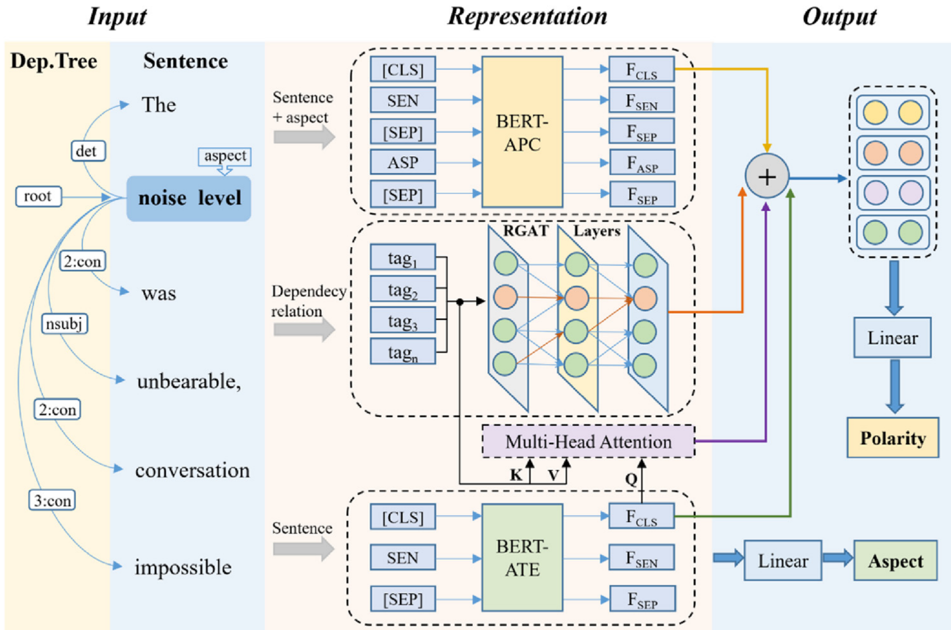

Fig. 1. An overview of the proposed multi-task learning model. It consists of four components: BERT-APC, RGAT, BERT-ATE, and MHA.

features of sentences and aspects. To obtain the global features from sentences and aspects, we use "$[CLS] + sentence + [SEP] + aspect + [SEP]$" as input. Suppose there is a sentence $S = (w_1, w_2, \ldots, w_n)$ and aspect $T = (w_i, w_{i+1}, \ldots, wj)$, w denotes the word in the sentence, i and j indicate where aspects start and end in the sentence, respectively. Then the input sequence is $I = (w_{cls}, w_1, w_2, \ldots, w_n, w_{sep}, w_i, w_{i+1}, \ldots, w_j, w_{sep})$. After feature extraction by BERT, we obtain a feature sequence h, and each feature in the sequence h corresponds to an input word.

$$h = (h_{cls}, h_1, h_2, \ldots, h_n, h_{sep}, h_i, h_{i+1}, \ldots, h_j, h_{sep}) \tag{1}$$

h_{cls} is the pooling vector of the whole sentence, which is used to represent the features of the whole sentence. We use h_{apc} to represent this feature.

3.3. RGAT

GAT is able to aggregate representations of neighboring nodes along dependency paths but the process does not consider the dependency labels. Inspired by Wang et al. [28], we use RGAT to encode dependency labels. Specifically, we choose R-GAT+BERT [28], where BERT replaces BiLSTM, and the attentional heads of RGAT are also replaced by those of BERT. For a sentence S, we obtain its dependency tree by exploiting a dependency parser first and build a new aspect-oriented dependency tree according to the tree reconstruction algorithm of Wang et al. [28], retaining the significant dependency relations. The final dependency relation features are calculated as follows:

$$\begin{aligned} g_{ij} &= \sigma\left(relu(r_{ij}W_{m1} + b_{m1})W_{m2} + b_{m2}\right) \\ \alpha_{ij} &= \frac{exp(g_{ij})}{\sum_{j=1}^{N_i} exp(g_{ij})} \\ h_{rgat} &= \|_{m=1}^{M} \sum_{j \in N_i} \alpha_{ij} W_m h_j \end{aligned} \tag{2}$$

where h_j represents the corresponding feature of the word j, which is calculated by BERT-APC. r_{ij} represents the relation embedding between nodes i and j. M is the relational head numbers in RGAT. h_{rgat} is the extracted feature by RGAT.

3.4. BERT-ATE

We regard the ATE task as a sequence annotation task. Suppose there is a sentence sequence $S = (w_1, w_2, \ldots, w_n)$, and the corresponding tag sequence is $Y = (t_1, t_2, \ldots, t_n)$, $t_i \in \{O, B_{asp}, I_{asp}, [CLS], [SEP]\}$. We input $[CLS] + W + [SEP]$ into BERT to obtain a feature sequence $F = (F_{cls}, F_1, F_2, \ldots, F_n, F_{sep})$. Based on the feature of each word, we use a linear layer to classify it and determine which tag it belongs to.

$$Y_{term} = \frac{exp(F_i)}{\sum_{k=1}^{N} exp(F_i)} \tag{3}$$

where N is the number of categories. In our model, $N = 5$, they are $O, B_{asp}, I_{asp}, [CLS], [SEP]$. F_{cls} is the feature of the whole input sentence, which we denote by h_{ate} for the final sentiment classification.

3.5. Multihead attention

To better correlate the ATE and APC tasks and highlight the important dependency relations, we apply the multihead attention mechanism to our model. We know that the attention mechanism can pay attention to some specific words. In the APC task, if some sentiment words can be associated with aspect, intuitively, it should be conducive to sentiment classification. Similar to Vaswani et al. [33], we use multiple scale dot attention (SDA) to capture the association between two tasks. The scale-dot attention is calculated as follows:

$$\begin{aligned} Q &= Q_{ate} \cdot W^q \\ K &= K_{dep} \cdot W^k \\ V &= V_{dep} \cdot W^v \end{aligned} \tag{4}$$

$$SDA(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \cdot V$$

where Q_{ate} is the sentence feature extracted by BERT-ATE, K_{dep} and V_{dep} are the dependent tag sequences, and $K_{dep} = V_{dep}$. W^q, W^k and W^v are weight matrices, $W^q \in \mathbb{R}^{d_{model} \times d_q}$, $W^k \in \mathbb{R}^{d_{model} \times d_k}$, $W^v \in \mathbb{R}^{d_{model} \times d_v}$, where d_{model} is the dimension of the model. MHA

Table 1
The distribution of the datasets.

Dataset	Positive		Negative		Neutral		Total	
	Train	Test	Train	Test	Train	Test	Train	Test
Laptop14	994	341	870	128	464	169	2328	638
Restaurant14	2164	728	807	196	637	196	3608	1120
Twitter	1561	173	1560	173	3127	346	6248	692

calculates multiple SDAs and concatenates multiple calculation results.

$$\begin{aligned} MHA(Q,K,V) &= Concat(head_1, head_2, \ldots, head_n)W^O \\ head_i &= SDA(Q,K,V) \end{aligned} \tag{5}$$

where $W^O \in \mathbb{R}^{hd_v \times d_{model}}$ is the parameter matrices for projection. h is the number of attention heads and we set it to 4. h_{mha} represents the feature extracted by the multihead attention mechanism.

3.6. Output layer

There are two outputs in this model: aspect term and aspect sentiment polarity. The aspect extraction module obtains the feature sequence corresponding to each word, and then uses a linear layer to obtain the probability that each word belongs to various labels.

$$p(l) = \frac{\exp\left(W_l F_i + b_l\right)_l}{\sum_{l' \in L} \exp\left(W_l F_i + b_l\right)_{l'}} \tag{6}$$

where L is the set of tags, $L = \{O, B_{asp}, I_{asp}, [CLS], [SEP]\}$. The above four modules all extract their features. We concatenate the four features as the final feature and exploit a linear layer for classification to obtain the probability that it belongs to each emotion category C.

$$\begin{aligned} h_f &= [h_{apc}; h_{ate}; h_{mha}; h_{rgat}] \\ p(c) &= \frac{\exp\left(W_f h_f + b_f\right)_c}{\sum_{c' \in C} \exp\left(W_f h_f + b_f\right)_{c'}} \end{aligned} \tag{7}$$

where h_f is the final feature.C is the set of sentiment types, $C = \{Positive, Negative, Neutral\}$.

3.7. Loss function

Our model is a multitask learning model, so we optimize the loss of two tasks simultaneously. The cross entropy loss is suitable for both APC and ATE tasks. The loss of the APC task is calculated as follows:

$$\mathcal{L}_{apc} = \sum_{1}^{C} \widehat{y}_i \log y_i \tag{8}$$

where C is the number of polarity categories. The loss function of the ATE task is:

$$\mathcal{L}_{ate} = \sum_{1}^{N} \sum_{1}^{k} \widehat{t}_i \log t_i \tag{9}$$

where N represents the number of categories and k represents the number of tokens of the input sequence. We calculate the final loss function as follows:

$$\mathcal{L} = \alpha \times \mathcal{L}_{apc} + (1-\alpha) \times \mathcal{L}_{ate} \tag{10}$$

where α is a hyperparameter, which indicates the proportional relationship between the loss of two tasks.

Algorithm 1 shows the algorithm of our proposed model.

Algorithm 1 The proposed algorithm

Input: sentence s, aspect a, dependency relations r
for all epoch in epochs **do**
  obtain feature h_{apc} by Eq. (1)
  obtain feature h_{rgat} by Eq. (2)
  obtain feature h_{ate} and the extracted aspect by BERT-ATE(s)
  obtain feature h_{mha} by Eq. (5)
  obtain the final feature and polarity probabilities by Eq. (7)
  calculate the loss of APC task by Eq. (8)
  calculate the loss of ATE task by Eq. (9)
  calculate the final loss by Eq. (10)
  calculate gradients and update parameters
end for

4. Experiments

We first present the three datasets used in this section. Then, we introduce the evaluation metrics and parameter settings and the baseline approaches used for comparison. Finally, the experimental results are given and analyzed.

4.1. Datasets

We conduct experiments on three public sentiment analysis datasets, including Laptop14, Restaurant14 and Twitter. Laptop14 and Restaurant14 are from SemEval 2014 task 4 [53], which includes sentiment comments from the restaurant and laptop areas. The Twitter dataset is mainly composed of tweets processed by Dong et al. [41]. The affective polarities in these datasets include three categories: positive, neutral and negative. Table 1 shows the number distribution of sentences of each sentiment polarity in the three datasets.

4.2. Metrics

Aspect sentiment classification is a multiclassification task. In this paper, we treat the APC task as a triple classification task and focus on the performance of the APC task. We use the accuracy and macroaverage F1 score as evaluation metrics of the proposed model. The higher the accuracy and macroaverage F1 values are, the better the performance of the model. Accuracy is the most intuitive performance measure, which is the proportion of the total number of correct results predicted. Accuracy measures the global sample predictions. The accuracy is calculated as follows:

$$Accuracy = \frac{C_{pre}}{T_{pre}} \tag{11}$$

where T_{pre} represents the total number of predictions and C_{pre} represents the number of correct predictions.

Macroaveraging is used to first count the index values for each class and then to find the arithmetic mean for all classes. Therefore, to calculate the macroaverage F1, we first calculate the F1 score for each class and then average it. The F1 score is the weighted average of precision and recall. For multiclassification, precision, recall, and F1 values are all specific to a class. For class A, class A is the positive class, and all other classes are negative. We have three classes: positive, neutral and negative. We first calculate the precision, recall and F1 score for each class as follows:

$$\begin{aligned} Precision_A &= \frac{C_A}{T_{A1}} \\ Recall_A &= \frac{C_A}{T_{A2}} \\ F1_A &= \frac{2 \times Precision_A \times Recall_A}{Precision_A + Recall_A} \end{aligned} \tag{12}$$

where $A \in \{Positive, Negative, Neutral\}$, C_A represents the number of correctly predicted Class A samples, T_{A1} represents the total number of samples predicted as Class A, and T_{A2} represents the total number of samples that are actually Class A.

The macroaverage F1 score is calculated as follows:

$$Macro - F1 = \frac{1}{n}\sum_{i=1}^{n} F1_i \tag{13}$$

where n represents the number of classes for sentiment classification and $F1_i$ represents the F1 score of each class.

4.3. Setup

We use Biaffine Parse to parse the sentence to obtain the dependency relations. RGAT encodes the dependency label. We set the dimension of dependency label embeddings to 300. We use BERT-BASE as BERT-APC and BERT-ATE. To extract global features, BERT-APC adopts the BERT-SPC [5] type input format. BERT-ATE extracts the aspects and uses the general BERT-type input format. We use the hidden states of the last layer of BERT as the feature of the words and the feature corresponding to the [CLS] token as the feature of the whole sentence. Dropout also has some influence on the experimental results. We set it to 0.2. Our model was trained with a maximum of 20 epochs and a mini-batch size of 16. The number of attention heads in MHA is set as 4, and the number of Heads in RGAT is set as 6. We set the hyperparameter $\alpha = 0.2$. We use the Adam optimizer with a learning rate of 5e-5 to minimize the loss. For baseline RGAT-BERT, we run the experiment according to the source code[2] provided by the authors. We have released the source code of this work on GitHub.

4.4. Comparison models

We selected some mainstream models for ABSA as our baseline, including the following:

- **With BERT:** BERT-BASE [39], BERT-PT [6], LCF-BERT [54], AEN-BERT [5], RGAT-BERT [28], BERT-ADA [55] and UGF-ABSA [56].
  - BERT-BASE [39] is the basic BERT model proposed by Google that achieves the best results on multiple tasks. We fine-tune it in experimental datasets.
  - BERT-PT [6] applies reading comprehension techniques to web review data. The authors refer to this problem as the Review Reading Comprehension task, and it can be applied to ABSA.
  - LCF-BERT [54] uses the local context attention mechanism to segment different aspects of a sentence and their corresponding emotional words to reduce the mutual influence between aspects.
  - AEN-BERT [5] adopts an attention coding network based on pretrained BERT model designed to address aspect-oriented sentiment classification problems.
  - RGAT-BERT [28] proposes a novel relationship graph attention network to encode sentence dependencies and make connections between aspects and opinions. Moreover, it proposes an aspect-oriented dependency tree structure. This experiment was rerun, using the parameters provided by the authors.
  - BERT-ADA [55] develops a domain-based adaptive BERT model for the APC task. The model uses a task-related corpus to continue training BERT-base and obtains good results on the Laptop dataset.
  - UGF-ABSA [56] builds a text generation model based on BART [57], which unifies the seven subtasks in ABSA.
- **Without BERT:** ASGCN [43], MGAN [58], AOA [36], TNet [32], RAM [59], Cabasc [60], IAN [35], IMN [49], RACL [25], DREGCN [52].
  - ASGCN [43] uses a graph convolutional network to encode sentence dependency trees, which solves the multiword dependence problem in ABSA.
  - MGAN [58] captures contextual information using a bidirectional LSTM network and captures the relationship between aspects and contexts using a multigranularity attention mechanism.
  - AOA [36] proposes an attention-over-attention neural network that captures the interaction between aspects and contexts.
  - TNet [32] converts BiLSTM embeddings into target-specific embeddings and encodes them using CNNs.
  - RAM [59] extracts sentence features using LSTM and captures distant sentiment word features using a multilayer attention mechanism.
  - Cabasc [60] proposes a novel sentiment classification model based on content attention and aspect, with two attention enhancing mechanisms: sentence-level content attention and context attention.
  - IAN [35] represents an interactive attention network that implements interactive learning of the target and context, generating attentional representations of the target and context, respectively.
  - IMN [49] proposes an interactive multitask learning network and designs a message passing architecture.
  - RACL [25] uses multitask learning to represent a relational-aware collaborative learning framework.
  - DREGCN [52] implemented a new end-to-end ABSA-dependent syntactic knowledge-enhanced interaction architecture based on multitask learning using graph convolutional networks and message passing mechanisms.

4.5. Results and analysis

4.5.1. Overall performance

Table 2 shows the overall performance of different models. Overall, the best performance is achieved by our model. Taking the restaurant dataset as an example, the results of the methods without BERT are not very good in summary. Their accuracy is mostly approximately 80, and the F1 value is approximately 71. Among them, the MGAN model has the highest accuracy of 81.25, which is 5.63 points lower than our model. The ASGCN model has the highest F1 value, which is 9.14 points lower than that of our model. From the experimental results, we find that the overall performance of the model with the pretrained BERT model is better, which also shows the powerful feature representation ability of BERT. With the assistance of BERT, our model obtains better results than other comparable models, with accuracies of 86.88, 80.56, and 76.59 on the Restaurant, Laptop, and Twitter datasets, respectively. In terms of accuracy and F1, our model outperforms the state-of-the-art RGAT-BERT model by 1.17 and 1.73 points on the Restaurant dataset, respectively. Compared with the best baseline RGAT-BERT, our results are significantly improved, especially on the laptop dataset, and our model has improved by 2.35 points. Our model combines ATE and APC tasks, which shows that multitasking can further improve the performance of sentiment classification. Additionally, using MHA

[2] https://github.com/shenwzh3/RGAT-ABSA

Table 2
Performance comparison of different models on the three datasets. The best results are bolded and we have underlined the second best results.

	Model	Restaurant		Laptop		Twitter	
		ACC	Macro-F1	ACC	Macro-F1	ACC	Macro-F1
w/o BERT	ASGCN	80.77	72.02	75.55	71.05	72.15	70.4
	MGAN	81.25	71.95	75.39	72.47	72.54	70.81
	AOA	81.20	–	74.50	–	–	–
	TNet	80.69	71.27	76.54	71.75	74.97	73.60
	RAM	80.23	70.80	74.49	71.35	69.36	67.30
	Cabasc	77.14	63.99	70.85	64.56	67.34	64.71
	IAN	78.60	–	72.10	–	–	–
	IMN	83.89	75.66	75.36	72.02	–	–
	RACL	81.42	74.46	73.53	71.09	–	–
	DREGCN	81.88	73.32	77.86	73.46	–	–
w BERT	BERT-BASE	85.09	78.74	78.06	72.51	72.83	70.74
	RGAT-BERT	<u>85.71</u>	<u>79.43</u>	78.21	74.07	74.71	73.42
	UGF-ABSA	–	75.56	–	<u>76.76</u>	–	–
	BERT-PT	84.95	76.96	78.07	75.08	–	–
	LCF-BERT	84.29	76.79	<u>78.53</u>	75.15	<u>75.43</u>	<u>74.46</u>
	AEN-BERT	83.12	73.76	78.06	73.97	72.98	69.83
	Ours	**86.88**	**81.16**	**80.56**	**77.00**	**76.59**	**74.67**

Table 3
Tukey's HSD Test Results between the Four Methods in three datasets. The numbers are mean-diff between methods X and Y.

Method pair		Mean-diff(X-Y)
X	Y	
AOA	LCF-BERT	−0.0812*
AOA	RGAT-BERT	−0.1073*
RGAT-BERT	LCF-BERT	0.0261
Ours	AOA	0.1392*
Ours	LCF-BERT	0.058*
Ours	RGAT-BERT	0.0318*

*means that the difference is statistically significant at α = 0.05.

between the two tasks can help the model pay attention to the sentiment words related to aspects, which is also helpful in sentiment classification. These results have demonstrated the effectiveness of our model.

We supplement Tukey's HSD test as shown in Table 3. The numbers are the mean-diff between Models X and Y. The measurement here is the accuracy of the sentiment classification. "*" means that the difference is statistically significant at α = 0.05. As seen from the table, our model surpasses other models, which shows that the gains of our model over AOA, LCF-BERT and RGAT-BERT are statistically significant.

4.5.2. Ablation study

We further perform an ablation study to evaluate the influence of each module. Our model mainly includes four components: BERT-APC, RGAT, BERT-ATE and MHA. We conduct ablation experiments on these modules to obtain a better understanding of their relative importance, and we retained the BERT-APC module in all ablation experiments because it extracts global features of sentences and aspects, which is very important and is the core part of the model. Table 4 shows the experimental results. We find that the more components we use, the higher the performance is. Overall, using RGAT to encode dependency labels improves the performance significantly, especially on the Restaurant dataset, where the accuracy reaches 86.25 by adding RGAT. After adding BERT-ATE and multihead attention, the performance on the three datasets is improved more obviously, especially on the Twitter dataset, in which the accuracy is improved by 1.73 percentage points, and the F1 is improved by 1.16 percentage points. This also demonstrates the effectiveness of the proposed model. Multitask learning and multihead attention mechanisms enable the model to better pay attention to words related to aspects, thus improving the performance for aspect sentiment classification tasks.

4.5.3. Domain-adaptation experiment

The BERT model is pretrained on a large-scale general corpus and then applied to various downstream tasks through fine-tuning. Meanwhile, the ABSA datasets are small-scale corpora. We know that we can further train BERT in specific domains to improve performance. Therefore, we continue to train the BERT-BASE model to obtain the domain-adapted pretrained BERT model on the Yelp Dataset Challenge reviews and the Amazon Laptops review dataset. Table 5 shows the results of the domain-adaptation experiment. We find that after domain adaptation training, the performance of all models is enhanced. Our model also obtains the best results, which also shows the effectiveness of domain adaptation.

4.5.4. Sample analysis

To visualize the capability of the model, we analyze 100 randomly selected test data from the restaurant test set and classify these sentences into five categories: neutral, double negative, suggestion, comprehensive understanding, and transition words. Table 6 shows some examples, and the bold part of the sentence represents aspects. The first category is neutral comments. The sentiment of such comments is not particularly obvious, but the model may be affected by some words in the sentence, resulting in misjudgments of neutral emotion. For example, in the sentence, "it was good, but none of the flavors wow", "but" is a transition word, which may lead to the model mistaking this as a negative attitude. The second category is the double negative sentence, for example, in the sentence, "you will not be disappointed by any of the choices in the menu", "disappointed" is negative. However, there is another negative word (not) in front of it, so it is positive on the whole. This kind of sentence is difficult to understand for the current model, which is easily affected by a single negative word. The third category is suggested comments, which only recommend or do not recommend users to try. This kind of sentence sometimes shows relatively positive sentiment, but there is no obvious emotional word in the sentence, leading to the model misjudgment of neutral sentiment. The fourth category is sentences that require comprehensive understanding. This kind of sentence is difficult in natural language processing and requires in-depth language understanding techniques. For example, the sentence, "since I cook for a living, I'm very fussy about the food I eat in restaurants", is simply an opinion without any emotion but

Table 4
Ablation study results on Restaurant, Laptop and Twitter datasets.

Components				Restaurant		Laptop		Twitter	
BERT-APC	RGAT	BERT-ATE	MHA	ACC	Macro-F1	ACC	Macro-F1	ACC	Macro-F1
✓	✓			86.25	80.27	79.15	75.59	74.86	73.51
✓		✓		85.89	79.43	78.06	73.94	74.42	73.16
✓			✓	82.32	73.48	77.12	72.11	70.95	68.95
✓	✓		✓	85.18	78.13	78.68	75.09	75.43	73.75
✓		✓	✓	85.54	79.19	76.18	69.66	74.42	72.21
✓	✓	✓		86.07	79.57	78.84	74.41	75.58	74.35
✓	✓	✓	✓	**86.88**	**81.16**	**80.56**	**77.00**	**76.59**	**74.67**

Table 5
Domain-adaption experiment results on Restaurant and Laptop datasets.

Model	Restaurant		Laptop	
	ACC	Macro-F1	ACC	Macro-F1
BERT-BASE	85.97	77.68	79.72	75.18
AEN-BERT	83.04	71.88	78.68	74.43
LCF-BERT	85.45	79.67	80.09	76.64
BERT-ADA	87.14	80.09	80.23	75.77
RGAT-BERT	86.51	80.75	79.31	75.3
Ours	**87.59**	**81.81**	**81.19**	**77.13**

can easily be misjudged as negative because of the word fussy. The fifth category is sentences with transition words, which may cause the model to predict completely opposite sentiment polarity and require a deeper semantic understanding of the sentence. Through the example analysis, we find that the predicted polarity of our model for the five categories of sentences matches the true polarity more than the other two comparison models, and we can better handle sentences with suggestions, sentences that require comprehensive understanding, and sentences with transitions.

5. Discussion

To further understand the effects of some important parameters and modules on the experimental results, this section discusses them in detail, including the impact of multitask learning and MHA, the effects of the number of heads in MHA and RGAT, the effects of hyperparameter α, the effects of different syntactic parsers, and the effects of multitask learning on ATE. In the discussion of the effects of multitask learning and MHA, we go further by changing the calculation of MHA to discuss how the ATE task in multitask learning contributes to the performance of the APC task. Then, we discuss the number of attention heads in the attention mechanism in MHA and RGAT. To investigate the effect of the ratio between the two losses, we also discuss the hyperparameter α in the final loss function. We also focus on and discuss the effect of different syntactic parsers on the model performance. While our previous focus was on improving the performance of the APC task, to understand the influence of multitask learning on the ATE task, we discuss and compare the performance of a single BERT-ATE model and the whole proposed model on the ATE task.

5.1. Effect of multitask learning and MHA

The ablation experiment illustrates the effectiveness of multitask learning and MHA. To further prove that the ATE task in multitask learning contributes to the performance of the APC task, we change the calculation method of MHA.

In our model, we use BERT to extract aspects. Meanwhile, we calculate the MHA based on BERT's global feature and dependency label sequence. Here, we can think that the BERT-ATE module has extracted a soft-aspect feature, so when we use the MHA mechanism, the model will focus more on this soft aspect to improve the performance of APC. To illustrate the effectiveness of the features extracted by the BERT-ATE module, we try to calculate MHA based on the features extracted by the BERT-APC and dependency label sequences. The results are shown in Table 7. It can be seen that the performance of *"MHA+BERT-ATE"* is significantly better than that of *"MHA+BERT-APC"*, which also shows that multitasking and MHA can improve the performance.

5.2. Effect of the number of heads in the MHA and RGAT

We exploit MHA to relate ATE and APC tasks. As we know, different attention heads attend to different places, so we conduct

Table 6
Some examples of Restaurant test set. Bold words in sentences indicate aspects.

Category	Sentence	Actual	Ours	RGAT-BERT	LCF-BERT	AOA
Neutral	it was good,but none of the **flavors** wow.	Neutral	Negative ✗	Negative ✗	Negative ✗	Positive ✗
	the **decor** is designed in a contemporary Japanese style restaurant.	Neutral	Neutral ✓	Neutral ✓	Neutral ✓	Positive ✗
Double negation	you will not be disappointed by any of the choices in the **menu**.	Positive	Positive ✓	Positive ✓	Positive ✓	Negative ✗
	the **prices** are not terrible.	Positive	Positive ✓	Positive ✓	Positive ✓	Negative ✗
Suggest	give it a try,**menu** is typical french but varied.	Neutral	Neutral ✓	Neutral ✓	Positive✗	Positive ✗
	i recommend the fried pork dumplings, the orange chicken/beef, and the **fried rice**.	Positive	Positive ✓	Neutral ✗	Positive ✓	Positive ✓
Comprehensive	the **food** did take a few extra minutes to come, but the cute waiter's jokes and friendliness made up for it.	Neutral	Neutral ✓	Neutral ✓	Negative ✗	Positive ✗
	since i cook for a living, i'm very fussy about the **food** i eat in restaurants.	Neutral	Neutral ✓	Negative ✗	Neutral ✓	Positive ✗
Transition word	the food was great - sushi was good, but the **cooked food** amazed us.	Positive	Positive ✓	Positive ✓	Positive ✓	Positive ✓
	good **cake** but: it was not the best cake i've ever had, and definitely not worth standing outside on the sidewalk being herded like cattle by indifferent and overworked employees.	Negative	Negative ✓	Positive ✗	Negative ✓	Negative ✓

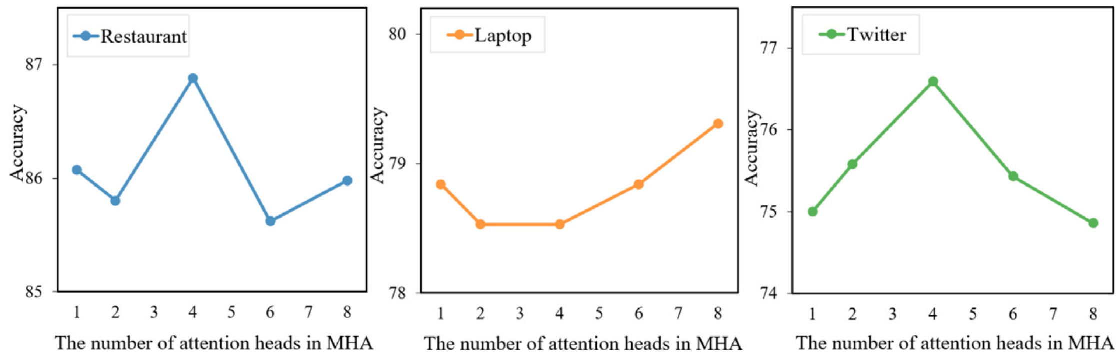

Fig. 2. The effect of the number of attention heads in MHA.

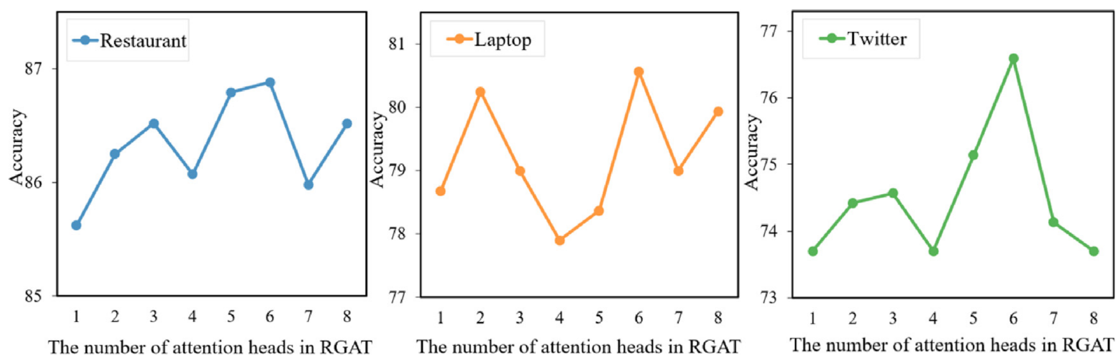

Fig. 3. The effect of the number of attention heads in RGAT.

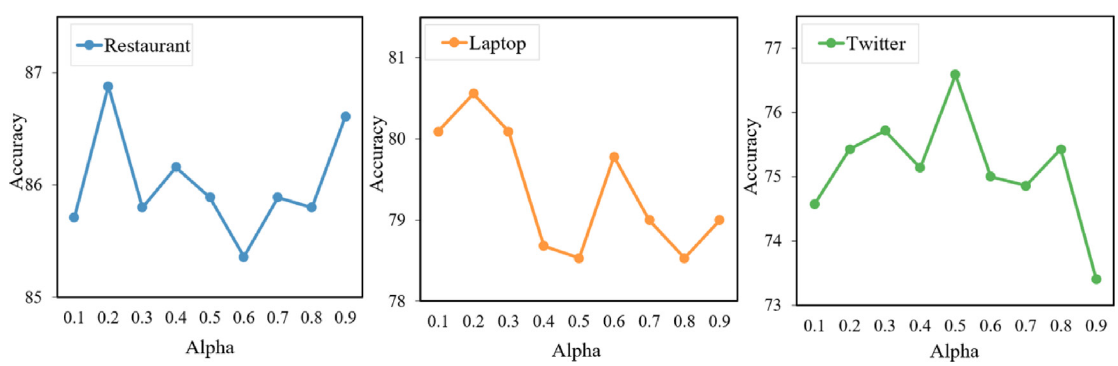

Fig. 4. The effect of the hyperparameter α.

Table 7
Results of different MHA calculation methods.

Model	Restaurant		Laptop	
	ACC	Macro-F1	ACC	Macro-F1
MHA+BERT-APC	86.34	80.29	78.68	74.96
MHA+BERT-ATE	**86.88**	**81.16**	**80.56**	**77.00**

experiments on the number of attention heads, and the results are shown in Fig. 2. On the Restaurant and Twitter datasets, when the number of attention heads is 4, we obtain the best results, and the results are also good on Laptop. Therefore, this parameter is set as 4 in our model for the convenience of subsequent experiments.

RGAT is also an attention mechanism that can also use multiple heads of attention to focus on different parts. The RGAT module uses multihead attention, so we discuss the effect of the number of attention heads in the RGAT on model performance. Fig. 3 presents the experimental results. It can be observed from the figure that the model gives the best results on the three datasets when the number of heads is 6. However, if there is only one attention head, the worst effect is achieved, which also means that multiple attention heads can improve the performance of the model, and different attention heads can concentrate on different information in different subspaces.

5.3. Effect of the hyperparameter α

To investigate the effect of the ratio between the two losses, we perform an experiment on the hyperparameter α, and we show the experimental results in Fig. 4. Our model reaches the best results on both the restaurant and laptop datasets when $\alpha = 0.2$. When $\alpha = 0.5$, the best result is obtained on the Twitter dataset. This shows that different parameters have a significant effect on the results. To facilitate the subsequent experiments, we set $\alpha = 0.2$.

Table 8
The effect of syntactic parsers.

Parser	Model	Restaurant		Laptop		Twitter	
		ACC	Macro-F1	ACC	Macro-F1	ACC	Macro-F1
Biaffine	RGAT-BERT	85.71	79.43	78.99	74.27	74.71	73.42
	Our Model	85.44	81.16	80.56	77.00	76.59	74.67
FT-RoBERTa	RGAT-BERT	85.44	78.41	78.53	74.66	74.71	72.82
	Our Model	86.07	79.76	79.31	74.64	76.01	75.04

Table 9
The effect of multi-task learning on ATE.

Model	Restaurant	Laptop	Twitter
	F1	F1	F1
BERT-ATE	86.42	80.23	97.49
Our Model	87.45	81.55	97.61

5.4. Effect of different parsers

We exploit RGAT to encode the dependency tree of a sentence in our model. However, different parsers parse different dependency tree structures, so we focus on the impact of different syntactic parsers on model performance. Two syntactic parsers are mainly compared: Biaffine and FT-RoBERTa [61], and the comparison results are shown in Table 8. It can be seen from Table 8 that the Biaffine parser has better performance, so the Biaffine parser is chosen to parse the dependency tree of sentences in our model.

5.5. Effect of multitask learning on ATE

All the previous experiments were conducted on the APC task. To understand the influence of multitask learning on the ATE task, we compared the performance of the single BERT-ATE model and the whole proposed model on the ATE task, and Table 9 shows the results. For the evaluation metric, we chose the F1-score. As seen from Table 9, our model significantly outperforms the single BERT-ATE model on the ATE task, which is a component of the proposed model, indicating that multitask learning can also improve the performance of the ATE task, further demonstrating the effectiveness of multitask learning.

6. Conclusion

We propose a new multitask learning model for ABSA by combining aspect term extraction and aspect polarity classification. Our model consists of four main modules: BERT-APC, BERT-ATE, RGAT, and MHA. The model can not only extract aspects but also classify aspects, however we mainly focus on the APC task, with the ATE task as an auxiliary means to improve the performance of the APC task. To correlate the two tasks and highlight important dependencies, we leverage a multihead attention mechanism to correlate the dependency sequences and aspect extraction that allows the model to pay more attention to the aspect-related words, thus improving the performance of APC. On all three widely used benchmark datasets, our model obtains state-of-the-art performance, demonstrating its effectiveness. Further analysis and discussion indicate the usefulness of incorporating multitask learning into the proposed model and applying MHA to combine them and obtain important dependency relations. In recent years, some new tendencies on neurosymbolic AI for explainable sentiment analysis have emerged, which are worth investigating. In the future, we will focus on explainable sentiment analysis.

CRediT authorship contribution statement

Guoshuai Zhao: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Supervision. **Yiling Luo:** Methodology, Software, Data curation, Writing – original draft, Writing – review & editing, Formal analysis, Validation. **Qiang Chen:** Methodology, Software, Data curation, Writing – original draft, Formal analysis. **Xueming Qian:** Resources, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 61902309, in part by China Postdoctoral Science Foundation under Grant 2020M683 496 and BX20190273, in part by the Fundamental Research Funds for the Central Universities, China (xxj022019003, xzd012022006) ; in part by the Humanities and Social Sciences Foundation of Ministry of Education, China under Grant 16XJAZH003, and in part by the Science and Technology Program of Xi'an, China under Grant 21RGZN0017.

References

[1] W. Xue, T. Li, 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), 2018, pp. 2514–2523.
[2] E.F. Ayetiran, Attention-based aspect sentiment classification using enhanced learning through cnn-Bilstm networks, Knowl.-Based Syst. 252 (2022) 109409.
[3] G. Zhao, X. Lei, X. Qian, T. Mei, Exploring users' internal influence from reviews for social recommendation, IEEE Trans. Multimed. 21 (3) (2018) 771–781.
[4] R.K. Yadav, L. Jiao, M. Goodwin, O.-C. Granmo, Positionless aspect based sentiment analysis using attention mechanism, Knowl.-Based Syst. 226 (2021) 107136.
[5] Y. Song, J. Wang, T. Jiang, Z. Liu, Y. Rao, Targeted sentiment classification with attentional encoder network, in: International Conference on Artificial Neural Networks, Springer, 2019, pp. 93–103.
[6] H. Xu, B. Liu, L. Shu, S.Y. Philip, 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), 2019, pp. 2324–2335.
[7] X. Bai, P. Liu, Y. Zhang, Investigating typed syntactic dependencies for targeted sentiment classification using graph attention neural network, IEEE/ACM Trans. Audio Speech Lang. Process. 29 (2020) 503–514.
[8] B. Liang, H. Su, L. Gui, E. Cambria, R. Xu, Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks, Knowl.-Based Syst. 235 (2022) 107643.

[9] H. Tang, G. Zhao, X. Bu, X. Qian, Dynamic evolution of multi-graph based collaborative filtering for recommendation systems, Knowl.-Based Syst. 228 (2021) 107251.

[10] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, C. Dyer, Neural architectures for named entity recognition, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016, pp. 260–270.

[11] C. Wan, Y. Peng, K. Xiao, X. Liu, T. Jiang, D. Liu, An association-constrained LDA model for joint extraction of product aspects and opinions, Inform. Sci. 519 (2020) 243–259.

[12] M. Venugopalan, D. Gupta, An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis, Knowl.-Based Syst. 246 (2022) 108668.

[13] A. Da'u, N. Salim, Aspect extraction on user textual reviews using multi-channel convolutional neural network, PeerJ Comput. Sci. 5 (2019) e191.

[14] G. Zhao, P. Lou, X. Qian, X. Hou, Personalized location recommendation by fusing sentimental and spatial context, Knowl.-Based Syst. 196 (2020) 105849.

[15] Z. Zhang, Y. Rao, H. Lai, J. Wang, J. Yin, TADC: A topic-aware dynamic convolutional neural network for aspect extraction, IEEE Trans. Neural Netw. Learn. Syst. (2021).

[16] E. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (02) (2016) 102–107.

[17] G. Zhao, Z. Liu, Y. Chao, X. Qian, CAPER: context-aware personalized emoji recommendation, IEEE Trans. Knowl. Data Eng. 33 (9) (2021) 3160–3172.

[18] K. He, R. Mao, T. Gong, C. Li, E. Cambria, Meta-based self-training and re-weighting for aspect-based sentiment analysis, IEEE Trans. Affect. Comput. (2022).

[19] R. Mao, Q. Liu, K. He, W. Li, E. Cambria, The biases of pre-trained language models: An empirical study on prompt-based sentiment analysis and emotion detection, IEEE Trans. Affect. Comput. (2022).

[20] T. Gao, J. Fang, H. Liu, Z. Liu, C. Liu, P. Liu, Y. Bao, W. Yan, LEGO-ABSA: A prompt-based task assemblable unified generative framework for multi-task aspect-based sentiment analysis, in: Proceedings of the 29th International Conference on Computational Linguistics, 2022, pp. 7002–7012.

[21] E. Cambria, Y. Li, F.Z. Xing, S. Poria, K. Kwok, SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 105–114.

[22] E. Cambria, Q. Liu, S. Decherchi, F. Xing, K. Kwok, SenticNet 7: a commonsense-based neurosymbolic AI framework for explainable sentiment analysis, in: Proceedings of LREC 2022, 2022.

[23] J. Kocoń, J. Baran, M. Gruza, A. Janz, M. Kajstura, P. Kazienko, W. Korczyński, P. Miłkowski, M. Piasecki, J. Szołomicka, Neuro-symbolic models for sentiment analysis, in: International Conference on Computational Science, Springer, 2022, pp. 667–681.

[24] J. Zhu, Y. He, G. Zhao, X. Bo, X. Qian, Joint reason generation and rating prediction for explainable recommendation, IEEE Trans. Knowl. Data Eng. (2022) 1, http://dx.doi.org/10.1109/TKDE.2022.3146178.

[25] Z. Chen, T. Qian, Relation-aware collaborative learning for unified aspect-based sentiment analysis, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 3685–3694.

[26] R. Mao, X. Li, Bridging towers of multi-task learning with a gating mechanism for aspect-based sentiment analysis and sequential metaphor identification, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, (15) 2021, pp. 13534–13542.

[27] C. Ke, Q. Xiong, C. Wu, Z. Liao, H. Yi, Prior-bert and multi-task learning for target-aspect-sentiment joint detection, in: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2022, pp. 7817–7821.

[28] K. Wang, W. Shen, Y. Yang, X. Quan, R. Wang, Relational graph attention network for aspect-based sentiment analysis, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 3229–3238.

[29] S. Poria, E. Cambria, L.-W. Ku, C. Gui, A. Gelbukh, A rule-based approach to aspect extraction from product reviews, in: Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP), 2014, pp. 28–37.

[30] Q. Liu, Z. Gao, B. Liu, Y. Zhang, Automated rule selection for aspect extraction in opinion mining, in: 24th International Joint Conference on Artificial Intelligence, IJCAI 2015, International Joint Conferences on Artificial Intelligence, 2015, pp. 1291–1297.

[31] K. Shah, H. Patel, D. Sanghvi, M. Shah, A comparative analysis of logistic regression, random forest and KNN models for the text classification, Augment. Hum. Res. 5 (1) (2020) 1–16.

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

[33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[34] X. Zheng, G. Zhao, L. Zhu, X. Qian, PERD: Personalized emoji recommendation with dynamic user preference, in: Proc. ACM SIGIR, 2022, pp. 1922–1926.

[35] D. Ma, S. Li, X. Zhang, H. Wang, Interactive attention networks for aspect-level sentiment classification, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017, pp. 4068–4074.

[36] B. Huang, Y. Ou, K.M. Carley, Aspect level sentiment classification with attention-over-attention neural networks, in: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, Springer, 2018, pp. 197–206.

[37] J. Tang, Z. Lu, J. Su, Y. Ge, L. Song, L. Sun, J. Luo, Progressive self-supervised attention learning for aspect-level sentiment analysis, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 557–566.

[38] J. Su, J. Tang, H. Jiang, Z. Lu, Y. Ge, L. Song, D. Xiong, L. Sun, J. Luo, Enhanced aspect-based sentiment analysis models with progressive self-supervised attention learning, Artificial Intelligence 296 (2021) 103477.

[39] J.D.M.-W.C. Kenton, L.K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of NAACL-HLT, 2019, pp. 4171–4186.

[40] A. Zhao, Y. Yu, Knowledge-enabled BERT for aspect-based sentiment analysis, Knowl.-Based Syst. 227 (2021) 107220.

[41] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, K. Xu, 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), 2014, pp. 49–54.

[42] R. Li, H. Chen, F. Feng, Z. Ma, X. Wang, E. Hovy, Dual graph convolutional networks for aspect-based sentiment analysis, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 6319–6329.

[43] C. Zhang, D. Song, Q. Li, Aspect-based sentiment classification with aspect-specific graph convolutional networks, in: EMNLP-IJCNLP 2019-2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference, 2020, pp. 4568–4578.

[44] J. Zeng, T. Liu, W. Jia, J. Zhou, Relation construction for aspect-level sentiment classification, Inform. Sci. 586 (2022) 209–223.

[45] H.T. Phan, N.T. Nguyen, D. Hwang, Convolutional attention neural network over graph structures for improving the performance of aspect-level sentiment analysis, Inform. Sci. 589 (2022) 416–439.

[46] M. Zhao, J. Yang, J. Zhang, S. Wang, Aggregated graph convolutional networks for aspect-based sentiment classification, Inform. Sci. 600 (2022) 73–93.

[47] S. Feng, B. Wang, Z. Yang, J. Ouyang, Aspect-based sentiment analysis with attention-assisted graph and variational sentence representation, Knowl.-Based Syst. (2022) 109975.

[48] H. Luo, T. Li, B. Liu, J. Zhang, DOER: Dual cross-shared RNN for aspect term-polarity co-extraction, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 591–601.

[49] R. He, W.S. Lee, H.T. Ng, D. Dahlmeier, An interactive multi-task learning network for end-to-end aspect-based sentiment analysis, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 504–515.

[50] H. Luo, L. Ji, T. Li, D. Jiang, N. Duan, GRACE: Gradient harmonized and cascaded labeling for aspect-based sentiment analysis, in: Findings of the Association for Computational Linguistics: EMNLP 2020, 2020, pp. 54–64.

[51] Y. Liang, F. Meng, J. Zhang, Y. Chen, J. Xu, J. Zhou, An iterative multi-knowledge transfer network for aspect-based sentiment analysis, in: Findings of the Association for Computational Linguistics: EMNLP 2021, 2021, pp. 1768–1780.

[52] Y. Liang, F. Meng, J. Zhang, Y. Chen, J. Xu, J. Zhou, A dependency syntactic knowledge augmented interactive architecture for end-to-end aspect-based sentiment analysis, Neurocomputing 454 (2021) 291–302.

[53] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, 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, 2014, pp. 27–35, http://dx.doi.org/10.3115/v1/S14-2004, URL https://aclanthology.org/S14-2004.

[54] B. Zeng, H. Yang, R. Xu, W. Zhou, X. Han, Lcf: A local context focus mechanism for aspect-based sentiment classification, Appl. Sci. 9 (16) (2019) 3389.

[55] A. Rietzler, S. Stabinger, P. Opitz, S. Engl, Adapt or get left behind: Domain adaptation through BERT language model finetuning for aspect-target sentiment classification, in: Proceedings of the 12th Language Resources and Evaluation Conference, 2020, pp. 4933–4941.
[56] H. Yan, J. Dai, T. Ji, X. Qiu, Z. Zhang, A unified generative framework for aspect-based sentiment analysis, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 2416–2429.
[57] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 7871–7880.
[58] F. Fan, Y. Feng, D. Zhao, Multi-grained attention network for aspect-level sentiment classification, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 3433–3442.
[59] P. Chen, Z. Sun, L. Bing, W. Yang, Recurrent attention network on memory for aspect sentiment analysis, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 452–461.
[60] Q. Liu, H. Zhang, Y. Zeng, Z. Huang, Z. Wu, Content attention model for aspect based sentiment analysis, in: Proceedings of the 2018 World Wide Web Conference, 2018, pp. 1023–1032.
[61] J. Dai, H. Yan, T. Sun, P. Liu, X. Qiu, Does syntax matter? A strong baseline for Aspect-based Sentiment Analysis with RoBERTa, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 1816–1829.