



# A Multi-task Shared Cascade Learning for Aspect Sentiment Triplet Extraction Using BERT-MRC

Wang Zou<sup>1,2</sup> · Wubo Zhang<sup>1</sup> · Wenhuan Wu<sup>1,3</sup> · Zhuoyan Tian<sup>4</sup>

Received: 18 July 2023 / Accepted: 6 January 2024 / Published online: 29 January 2024  
© Springer Science+Business Media, LLC, part of Springer Nature 2024

## Abstract

The aspect sentiment triplet extraction (Triplet) aims at extracting aspect terms (AE), extracting aspect-oriented opinion terms (AOE), and discriminating aspect-level sentiment polarity (ASC) from the comments. To address the current study, the end-to-end framework-based approach suffers from the problem of contribution distribution among multiple components, while the pipeline framework-based approach is susceptible to error propagation. Moreover, the complexity of the model limits the detection of long-distance aspect terms and opinion terms. In this paper, we propose a framework based on multi-task shared cascade learning and machine reading comprehension (MRC), which is called Triple-MRC. The multi-task shared cascade learning can effectively avoid the problem of contribution distribution among components. The MRC approach leverages the prior knowledge from the question to reduce the error propagation between tasks and mitigate the limitation associated with model complexity. We conduct experiments on publicly available two benchmark datasets for the Triplet task. The experimental results demonstrate the superior performance of the Triple-MRC framework compared to the baseline model, which better achieves the Triplet task. Through the analysis of the comparison study, model training process, error analysis, ablation study, attention visualization, and case study, we have confirmed the effectiveness of introducing the multi-task shared cascade learning method and MRC method into the model.

**Keywords** Asepct sentiment triplet extraction · Multi-task shared cascade learning · Machine reading comprehension · Prior knowledge

## Introduction

With the rapid development of e-commerce, users share comments on the application platforms, providing a wealth of valuable information. For manufacturers, the information

within customer comments encapsulates valuable insights into consumer opinions and emotional attitudes towards their products. Mining the value information from comments can serve as feedback to aid in product improvement. For consumers, the details provided in product comments enable them to assess the product's quality and make informed choices. Hence, it is crucial to employ efficient techniques for extracting valuable information from reviews.

Utilizing the aspect-based sentiment analysis (ABSA) [1] technique allows for a comprehensive extraction of sentiment polarity and opinion attitudes related to specific aspect terms in the comments. Consequently, it finds extensive applications in extracting user feedback in various domains, including sales and services such as restaurants, electronics, clothing, and automobiles. Triplet [2] technology is a research direction within the ABSA task, with the goal of extracting aspect terms and opinion terms, as well as distinguishing the sentiment polarity of aspect terms. As illustrated in the example shown in Fig. 1, consider the sentence “The wine was delicious, but the service was not so great.” In this

---

✉ Wang Zou  
zouwang@huat.edu.cn

Wubo Zhang  
zhangwubo1234@126.com

Wenhuan Wu  
wuwenhuan5@163.com

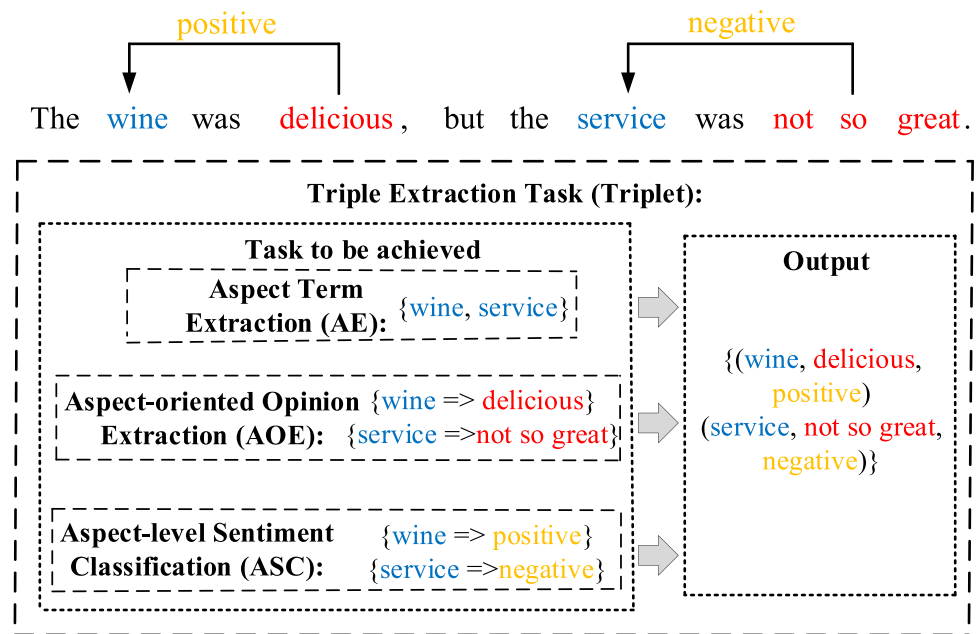
<sup>1</sup> School of Electrical and Information Engineering, Hubei University of Automotive Technology, Shiyan 442002, China

<sup>2</sup> Intelligent Manufacturing Institute, Dongfeng Design Institute Co Ltd, Wuhan 430000, China

<sup>3</sup> School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China

<sup>4</sup> School of Literature and Journalism Communication, South-Central Minzu University, Wuhan 430000, China

**Fig. 1** An illustrative example of the Triplet task, where we decompose it into three subtasks: AE, AOE, and ASC



example, the aspect terms are “wine” and “service,” the opinion terms are “delicious” and “not so great,” and their sentiment polarities are “positive” and “negative,” respectively. The Triplet extraction can acquire the triples (wine, delicious, positive) (service, not so great, negative) from the given sentence. According to the objective, we divide the Triplet task into three subtasks for execution: aspect term extraction (AE), aspect-oriented opinion extraction (AOE), and aspect-level sentiment classification (ASC) tasks. The AE task involves extracting aspect terms from the sentences, the AOE task focuses on extracting corresponding opinion terms based on the aspect terms, and the ASC task aims to identify the sentiment polarity to which the aspect terms belong.

In response to the current ABSA research landscape, most researchers concentrate on the ASC task, overlooking the role of opinion terms in a sentence. We contend that there exists a connection among aspect terms, opinion terms, and sentiment polarity. The end-to-end method [3] encounters challenges related to the distribution of contributions among multiple components, whereas the pipeline-based method [4] has the problem of error propagation. In this paper, we propose a multi-task shared cascade learning method based on machine reading comprehension (MRC) [5] for the Triplet task. The method utilizes the BERT-MRC model as its foundation and leverages the benefits of both multi-task shared parameter learning and multi-task cascade learning. Specifically, the MRC model accomplishes the AE, AOE, and ASC tasks individually through a question-answer (QA) [6] training methodology. The multi-task shared parameter learning [7] effectively avoids the problem of contribution distribution. Concurrently, multi-task cascade learning [8] maximizes the utilization of prior knowledge in the question,

thereby alleviating the problem of error propagation among tasks. The MRC model approach circumvents limitations in detecting aspect terms and opinion terms over long distances.

The main contributions of this paper are listed below:

1. We incorporate the multi-task learning (MTL) [9] and the MRC approach into the Triplet task. Additionally, we innovatively combine the respective advantages of multi-task shared parameter learning and multi-task cascade learning.
2. We introduce a multi-task shared cascade learning model based on MRC, called Triple-MRC.<sup>1</sup> The approach subdivides the Triplet task into AE, AOE, and ASC tasks, employing three BERT-MRC models with shared parameters to execute all subtasks. Additionally, we incorporate dependency syntactic features [10] into the model to help it better understand sentence structures.
3. This paper conducts a comparison study, error analysis, ablation experiment, attention visualization study, and case study to demonstrate the performance of our proposed model. The experimental results indicate that Triple-MRC outperforms the current state-of-the-art methods.

## Related Work

In this section, we first present the current state of research on traditional ABSA task. Next, we introduce ABSA based on

<sup>1</sup> The code is available at: <https://github.com/ZouWang-spider/Triple-MRC>

multi-task learning. Finally, we focus on the state of research in the recently proposed Triplet task.

To establish a connection between AE and ASC, researchers proposed employing the pipeline approach to address the ABSA task. Dai and Song [11] proposed a pipeline approach that leverages rule mining and a weakly supervised neural network model to perform AE and ASC. To tackle the issue of error propagation in practical applications with the pipeline method, Wu et al. [12] introduced a novel grid tagging scheme for the ASC model. The model also designs an effective grid tagging scheme inference strategy, leveraging mutual indications between different opinion factors to achieve more accurate extraction. Some scholars have suggested employing a tagging scheme to execute the ABSA task. Ji et al. [13] annotated sentiment polarity on aspect term labels, combining boundary information with sentiment polarity. They employed a conditional random field (CRF) [14] to minimize empirical risk and assigned each token in the given label set based on the learned transition matrix. Kumar et al. [15] proposed a BERT-based ensemble adversarial training interactive learning method to address the complete ABSA problem through a unified tagging scheme. Zhang et al. [16] employed a span-based annotation scheme, extracting multiple entities under the supervision of span boundary detection and accurately predicting sentiment polarity based on span distances. In response to the issue of overfitting and underfitting caused by imbalanced labels, He et al. [17] proposed a meta-based self-training and reweighting ABSA method. They trained a teacher model to generate domain-specific knowledge, while the student model was used for supervised learning. Subsequently, the meta-weights of the model were jointly trained with the student model, providing task-specific weights for each instance to coordinate their convergence speed and balance label categories.

In recent years, researchers have proposed introducing the MTL approach to mitigate the problem of error propagation between tasks. Wang et al. [18] proposed a novel multi-task neural learning framework to simultaneously address AE and ASC subtasks. Huan et al. [19] proposed a multi-task dual-encoder framework. Firstly, they constructed dual encoders to encode and fuse sentence information and semantic information separately. Then, they utilized implicit symbols and constraints between word pairs to accomplish multi-task inference and triplet decoding. Ma et al. [20] designed the hierarchical stacked bidirectional gated recurrent unit (HSBi-GRU) joint model to capture abstract features of both tasks, allowing the target label to influence the sentiment label.

The Triplet task was initially introduced by Peng et al. [21] in 2020, and it has garnered increasing attention recently. To address the issue of researchers not fully leveraging opinion term information, they proposed a two-stage framework. The first stage contains a unified AE-oriented sequence tagging model and an opinion term extraction (OE) oriented

graph convolutional neural network (GCN) [22]. In the second stage, all potential aspect-opinion pairs are enumerated, and a binary classifier is employed to determine whether aspect terms and opinion terms match. Finally, the sentiment polarity is computed by combining the aspect terms with the corresponding opinion term information. However, the method is computationally inefficient and necessitates the separate training of models in the two phases. Xu et al. [23] introduced the first location-aware tagging scheme with a novel end-to-end joint model to concurrently extract triplets. Additionally, the model employs CRF and semi-Markov CRF-based models [24] to predict sequence labels. Yan et al. [25] proposed using the pre-trained bidirectional and autoregressive transformers model (BART) [26] to address all ABSA subtasks within a unified generative framework. They redefined the goals of all subtasks in ABSA as a sequence composed of pointer indexes and sentiment class indexes. This transformation unifies all subtasks into a generative formula and employs the BART model for learning. However, the time complexity of the model hinders the detection of the aspect terms with long-distance opinion terms. Mao et al. [27] employed a pipeline framework to devise a two-step MRC model. In this model, the MRC model on the left facilitates aspect term extraction, and the MRC model on the right handles both aspect sentiment recognition and opinion term extraction. Additionally, two unified questions are also customized in the model for both left- and right-side tasks. Due to the length of sentences and the complexity of the model, the current model struggles to effectively detect opinion terms that are distant from aspect terms in the sentence. Li et al. [28] proposed the bidirectional encoder representations from the transformers-based MRC (BERT-MRC) model in the named-entity recognition (NER) [29] task. This model adopted the MRC approach instead of the traditional CRF model and demonstrated superior performance. The results indicate that the MRC-based approach can accurately identify aspect terms and opinion terms in sentences.

To summarize the current state of ABSA research, contemporary scholars have employed MTL methods to address the issue of error propagation inherent the pipeline framework. The design of a unified tagging scheme in multiple task models has gained popularity. Additionally, the MRC approach is being utilized to overcome the limitations of the models in effectively detecting long-distance aspect terms and opinion terms in sentences.

## Proposed Framework

In this section, we first provide a detailed description of the overall structure of the Triplet-MRC model. We then present the computational process of multi-task shared cascade learning for the Triplet task and Triple-MRC model, respec-

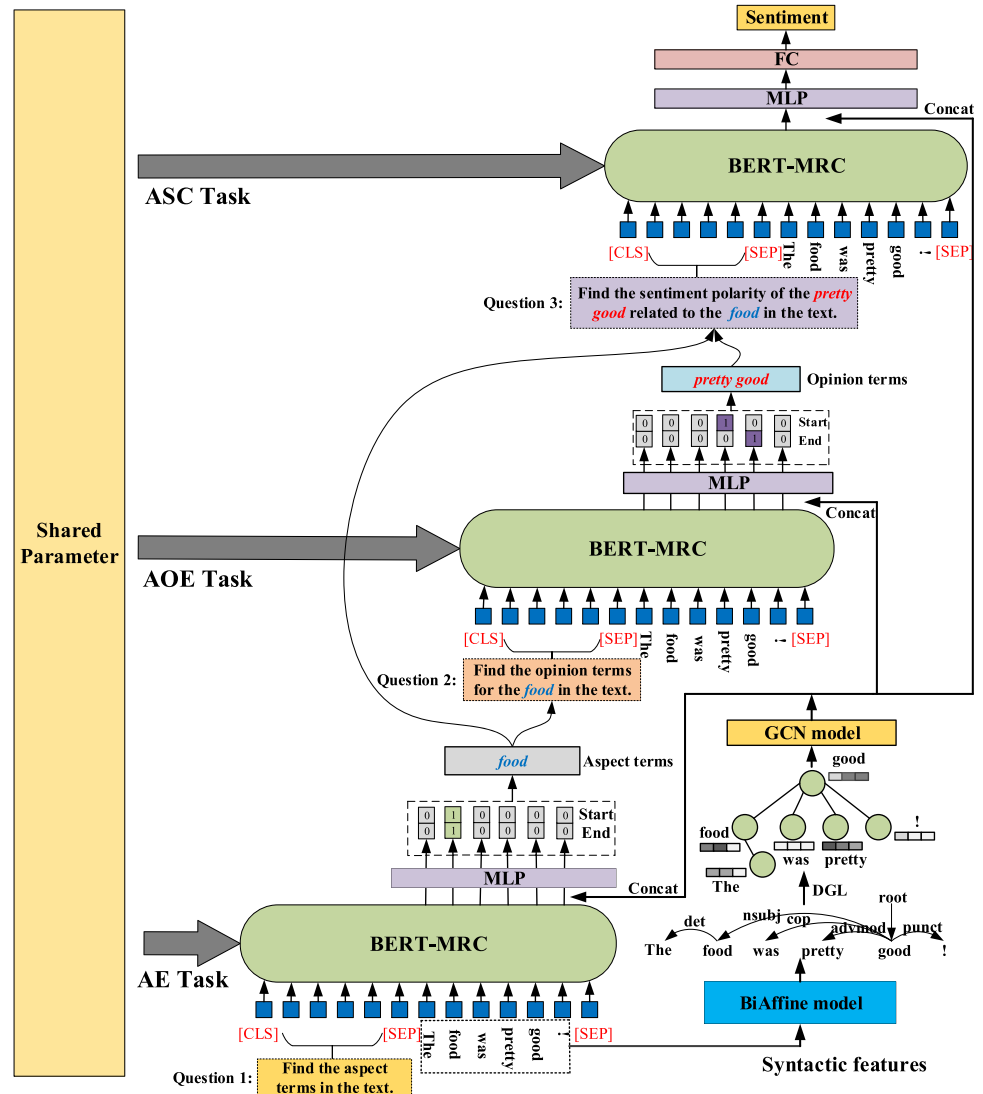
tively. Finally, we introduce the training inference process to illustrate how the model is trained and implements the Triplet task.

The overall structure of the Triple-MRC model is depicted in Fig. 2, comprising three primary BERT-MRC models: bottom BERT-MRC for the AE task, middle BERT-MRC for the AOE task, and top BERT-MRC for the ASC task. Additionally, syntactic features are incorporated into the model to enhance its ability to learn the dependency syntax of sentences. Table 1 provides a comprehensive overview of the specific inputs and outputs for the three BERT-MRC models. Given the input sentence, “The food was pretty good !”, the input for the bottom BERT-MRC model is [CLS] Question 1 [SEP] Sentence [SEP], with the design of Question 1 being “Find the aspect terms in the text.” Subsequently, the output hidden vectors are concatenated with the syntactic feature vectors and fed into the multi-layer perceptron layer (MLP). The resulting output signifies the starting and

ending positions of the aspect term *food*. The middle BERT-MRC model receives an input of [CLS] Question 2 [SEP] Sentence [SEP], with Question 2 designed as “Find the opinion terms for the {*food*} in the text.” Subsequently, an MLP is utilized to learn the concatenated output hidden vectors with the syntactic feature vectors. The resulting output signifies the starting and ending positions of the opinion term *pretty good*. The top BERT-MRC model is fed with an input of [CLS] Question 3 [SEP] Sentence [SEP], with Question 3 formulated as “Find the sentiment polarity of the {*pretty good*} related to the {*food*} in the text.” Subsequently, the output vectors are concatenated with syntactic feature vectors and fed into an MLP and fully connected layers. The softmax function is employed to compute the probability distribution of sentiment polarity.

Besides, the comprehensive model structure also includes the design of multi-task shared cascade learning, aiming to establishing connections among the three tasks through the

Fig. 2 The overall structure of Triple-MRC model



**Table 1** Methodology summary

Task	Component	Prior knowledge (question)	Input	Sentence	Output
AE	Bottom BERT-MRC	Find the aspect terms in the text.		The food was pretty good !	<i>food</i>
AOE	Middle BERT-MRC	Find the opinion term for the { <i>food</i> } in the text.		The food was pretty good !	<i>pretty good</i>
ASC	Top BERT-MRC	Find the sentiment polarity of the { <i>pretty good</i> } related to the { <i>food</i> } in the text.		The food was pretty good !	<i>positive</i>

utilization of shared parameters across the three BERT-MRC models. Cascade learning is implemented by employing the output of the previous task as the input for the subsequent task. The design of both Question 2 and Question 3 is based on the output of the preceding model, enabling the model to leverage prior knowledge in the questions, thereby mitigating the error propagation between tasks.

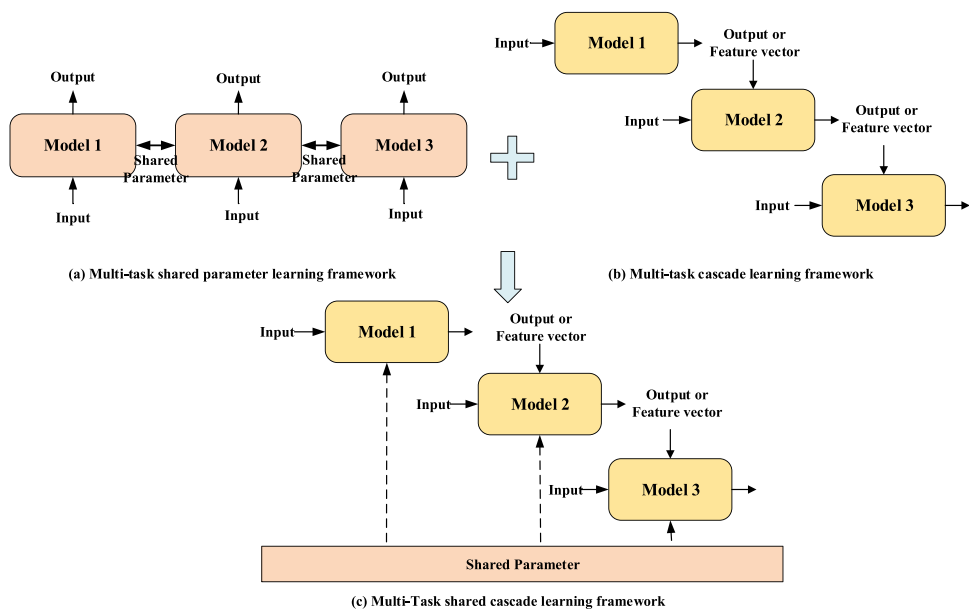
**Multi-task Shared Cascade Learning for Triplet Task**

The MTL framework is employed to learn the relationships between multiple tasks, which is widely used in multi-modal domains [30]. In devising the multi-task shared cascade learning, we amalgamate two frameworks: shared parameter learning and cascade learning within the MTL paradigm. As depicted in Fig. 3, the multi-task shared parameter learning framework in panel a comprises multiple independent models, each dedicated to executing distinct tasks. These models are interconnected by shared parameter to establish relationships among multiple tasks.

The multi-task cascade learning framework in panel b also comprises multiple independent models, but with a key distinction-here, the framework utilizes the output or feature vector of the current model as the input for the subsequent model. This approach is employed to establish connections between independent models and enhance the learning of relationships among multiple tasks. The multi-task shared cascade learning framework in panel c integrates the strengths of both frameworks. While conducting cascade learning, it incorporates shared parameter between models, facilitating a more robust learning of correlations among multiple tasks.

The implementation of multi-task shared cascade learning for the Triplet task is computed as follows. Given a sentence  $x_j$  with a maximum length of  $n$  as the input, the output of the given triplet with annotations is  $T_j = \{(a, o, s)\}$ , where  $(a, o, s)$  respectively denote (aspect term, opinion term, sentiment polarity) and  $s \in \{Positive, Neutral, Negative\}$ . For the training dataset  $D = \{(x_j, T_j)\}$ , the objective function of the model is as follows.

**Fig. 3** Multi-task shared cascade learning framework



$$L(D) = \prod_{j=1}^{|D|} \prod_{(a,o,s) \in T_j} P((a, o, s)|x_j) \tag{1}$$

Calculate the log-likelihood of  $x_j$ , where given sentence  $x_j$  and aspect term  $a$ , opinion term  $o$ , and sentiment polarity  $s$  are conditionally independent.

$$\begin{aligned} \ell(x_j) &= \sum_{(a,o,s) \in T_j} \log P((a, o, s)|x_j) \\ &= \sum_{a \in T_j} \sum_{(o,s) \in T_j|a} \log P(a|x_j) + \log P((o, s)|a, x_j) \\ &= \sum_{a \in T_j} \sum_{(o,s) \in T_j|a} \log P(a|x_j) \\ &+ \sum_{a \in T_j} \sum_{o \in T_j|a} \log P(o|a, x_j) \\ &+ \sum_{a \in T_j} \sum_{o \in T_j|a} \sum_{s \in T_j|a,o} \log P(s|(a, o, x_j)) \end{aligned} \tag{2}$$

We sum the above equation  $x_j \in D$  and normalize both sides, and then we obtain a log-likelihood of the following form:

$$\begin{aligned} \ell(D) &= \alpha \cdot \sum_{j=1}^{|D|} \sum_{(a,o,s) \in T_j|a} \log P(a|x_j) \\ &+ \beta \cdot \sum_{j=1}^{|D|} \sum_{a \in T_j} \sum_{o \in T_j|a} \log P(o|a, x_j) \\ &+ \gamma \cdot \sum_{j=1}^{|D|} \sum_{a \in T_j} \sum_{o \in T_j|a} \sum_{s \in T_j|a,o} \log P(s|(a, o, x_j)) \end{aligned} \tag{3}$$

where  $\alpha, \beta, \gamma \in [0, 1]$ , which are obtained by training the Triple-MRC model.

### Triple-MRC Model

The use of syntactic features in the Triple-MRC model contributes to a better understanding of sentence structure. Firstly, the BiAffine<sup>2</sup> model [31] is employed to compute the dependency relationships between words in the sentence. The BiAffine model is a dependency syntactic parsing tool proposed by NLP researchers at Stanford University. Subsequently, the Deep Graph Library (DGL)<sup>3</sup> tool is used to

construct a graph with words as node features and dependency relationships as edges. DLG is a graph neural network framework launched by New York University and Amazon. Finally, the graph is fed into the GCN model to learn the syntactic features and output hidden vectors. When inputting a sentence  $X = \{x_i, i \in [1, n]\}$ , the syntactic features are computed as follows.

$$(x_i, d_i, x_j) = BiAffine(X) \tag{4}$$

$$G = \sum_{i,j}^n DGL(x_i, x_j) \tag{5}$$

$$V = GCN(G) \tag{6}$$

where  $x_i, x_j$  denotes the word in the sentence, and  $i, j \in [1, n]$  and  $i \neq j, n$  is the number of words.  $d_i$  is the dependency relationship between  $x_i$  and  $x_j$ .  $G = (N, E)$  denotes graph, where  $N \in [x_1, x_n]$  is the set of nodes and  $E$  is the set of edges. Each edge  $e_i \in E$  connects two nodes  $x_i$  and  $x_j$ , usually denoted by  $e_i = (x_i, x_j)$ .  $V = \{v_1, \dots, v_n\}$  and  $V \in \mathbb{R}^{n \times h}$  is the syntactic feature hidden vector, and  $h$  represents the dimension of the vector.

The following will describe in detail how to fine-tune and design the bidirectional encoder representations from transformers (BERT) [32] model for application to MRC task. The central component of the Triple-MRC model is the BERT-MRC model, which employs BERT as the backbone model to encode the contextual information and MRC as the learning training method. When given a sentence  $X$  and the Question  $1 Q1 = \{q_i, i \in [1, m]\}$ , the output aspect term start position and end position are denoted as  $g_a^s$  and  $g_a^e$ , respectively. The model performs AE task computation as shown below.

$$H_a = BERT - MRC_{aspect}(Q1, X) \tag{7}$$

$$h_a^s, h_a^e = MLP(Concat(H_a, V)) \tag{8}$$

$$g_a^s = softmax(h_a^s) \tag{9}$$

$$g_a^e = softmax(h_a^e) \tag{10}$$

where  $BERT - MRC_{aspect}$  indicates that the model performs the AE task.  $Concat$  denotes the concatenated operation. The  $h_a^s \in \mathbb{R}^{n \times d}$  and  $h_a^e \in \mathbb{R}^{n \times d}$  denote the model calculates to obtain the hidden vectors of the sequence of aspect term start position and end position. The  $g_a^s \in \mathbb{R}^n$  and  $g_a^e \in \mathbb{R}^n$  denote the probability distributions of the sequence of aspect term start position and end position, respectively.

The prediction of the start and end position is typically achieved through a span-based approach, involving a binary

<sup>2</sup> The BiAffine model is available at <https://github.com/chantera/biaffineparser>

<sup>3</sup> The DGL Framework is available at <https://www.dgl.ai/pages/start.html>

classification task for each position in the sequence. The cross-entropy loss function is employed to compute the loss at the starting position and the loss at the ending position separately. Subsequently, these two losses are summed up to form the loss function for the AE task.

$$L_{aspect} = - \sum_i y_i^s \log(g_{ai}^s) - \sum_i y_i^e \log(g_{ai}^e) \quad (11)$$

where  $y^s \in \{1, 0\}$  denotes the starting position of the annotated aspect term and  $y^e$  is the ending position. The  $i \in [1, n]$  and  $n$  is the length of the sentence.

The AOE task is analogous to the AE task, as both tasks leverage the BERT-MRC model for sequence prediction. The process of calculating the AOE task is shown below.

$$H_o = BERT - MRC_{opinion}(Q2, X) \quad (12)$$

$$h_o^s, h_o^e = MLP(Concat(H_o, V)) \quad (13)$$

$$g_o^s = softmax(h_o^s) \quad (14)$$

$$g_o^e = softmax(h_o^e) \quad (15)$$

where  $X$  denotes the input sentence and  $Q2$  is the question designed for extracting opinion term.  $BERT - MRC_{opinion}$  denotes the model performs the AOE task. The  $h_o^s \in \mathbb{R}^{n \times d}$ ,  $h_o^e \in \mathbb{R}^{n \times d}$  denote the hidden vectors of the model output start position and end position sequences. The  $g_o^s \in \mathbb{R}^n$ ,  $g_o^e \in \mathbb{R}^n$  denote the probability distributions of the calculated start position and end position sequences.

The loss function for the opinion term extraction task is calculated as follows.

$$L_{opinion} = - \sum_i k_i^s \log(g_{oi}^s) - \sum_i k_i^e \log(g_{oi}^e) \quad (16)$$

where  $k^s \in \{1, 0\}$  denotes the starting position of the annotated opinion term and  $k^e$  is the ending position.

Different from the other two types of tasks, the ASC task is essentially a classification task using the BERT-MRC model. Therefore, we fine-tune the supervised training of the model, where the input is a sentence and Question 3  $Q3 = \{q_i, i \in [1, k]\}$ , and the output is the sentiment polarity label  $E \in \{Positive, Negative, Neutral\}$ . The process of calculating the ASC task is shown below.

$$H_s = BERT - MRC_{sentiment}(Q3, X) \quad (17)$$

$$h_s = MLP(Concat(H_s, V)) \quad (18)$$

$$f_s = FC(h_s) \quad (19)$$

$$p_s = softmax(f_s) \quad (20)$$

where  $h_s \in \mathbb{R}^{n \times h}$  denotes the hidden layer vector of the model output, and  $h$  is the dimension of the vector. The  $FC$  is the computation of the fully connected layer, and the  $p_s$  is the sentiment probability distribution.

The ASC task is essentially a multi-classification task, which is calculated using the cross-entropy loss function and L2 regularization as follows.

$$L_{sentiment} = CE(p_s, E) + \lambda \sum_{\theta \in \Theta} \theta^2 \quad (21)$$

where  $CE$  is the cross-entropy loss function and  $E$  is the annotated sentiment polarity. The  $\lambda$  is the weighting coefficient, and  $\Theta$  is the parameter-set of the model. Then, the final multi-task training loss function is as follows.

$$\zeta = \alpha \cdot L_{aspect} + \beta \cdot L_{opinion} + \gamma \cdot L_{sentiment} \quad (22)$$

where  $\alpha, \beta, \gamma \in [0, 1]$  are hyper-parameters to control the contributions of objectives.

### Training Process Inference

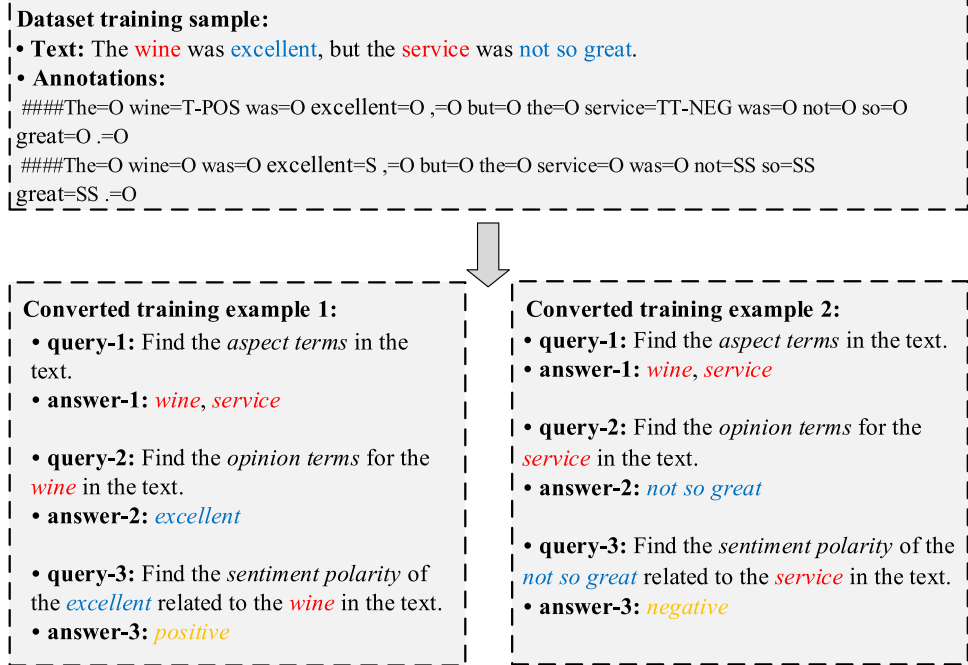
Question 1 = “Find the aspect terms in the text.” (23)

Question 2 = “Find the opinion term for the {*aspect term*} in the text.” (24)

Question 3 = “Find the sentiment polarity of the {*opinion term*} related to the {*aspect term*} in the text.” (25)

Typically, a sentence in the dataset will contain multiple aspect terms and opinion terms. To tackle the training problem that sentences contain multiple aspect terms and opinion terms, we have designed questions in the Triple-MRC framework to address each aspect term individually. First, the model leverages the prior knowledge from Question 1 to extract all the aspect terms in the sentence. Subsequently, it utilizes each aspect term to craft a specific Question 2. The generic sentence pattern for Question 2 is illustrated in Eq. (24), wherein each aspect term replaces the “{*aspect term*}” in Question 2. Finally, the output aspect terms and their corresponding opinion terms are employed as prior knowl-

**Fig. 4** The dataset training process. The sentences contain two aspect terms: “wine” and “service.” We have designed common Questions 2 and 3 to facilitate the specific training of each aspect term



edge to formulate Question 3. The generic sentence pattern for Question 3 is depicted in Eq. (25), using the aspect term and opinion term instead of “{*opinion term*” and “{*aspect term*” in the Question 3. Algorithm 1 provides a details description of the training inference process within the Triple-MRC framework (Fig. 4).

**Algorithm 1** The inference process for Triplet task of the Triple-MRC framework

**Require:** Sentence  $x$ , Question 1  $Q_1$ , Question 2  $Q_2$ , Question 3  $Q_3$   
**Ensure:**  $T = \{(a, o, s)\}$  triples

- 1: Initialize  $T = \{ \}$ ;
- 2:  $D = BiAffine(x)$ ;
- 3:  $G = DGL(D)$ ;
- 4:  $h_1 = BERT - MRC(x, Q_1)$ ;
- 5:  $A = MLP(Concat(h_1, G))$ ;
- 6: **if**  $A = \{ \}$  **then**
- 7:   **Return**  $T$ ;
- 8: **end if**
- 9: **for**  $a_i \in A$  **do**
- 10:    $h_o = BERT - MRC(x, Q_2)$ , where  $Q_2$  contain  $a_i$ ;
- 11:    $o_i = MLP(Concat(h_o, G))$ ;
- 12:    $h_m = BERT - MRC(x, Q_3)$ , where  $Q_3$  contain  $a_i$  and  $o_i$ ;
- 13:    $h_s = MLP(Concat(h_m, G))$ ;
- 14:    $s_i = FC(h_s)$ ;
- 15:    $T \leftarrow T \cup \{(a_i, o_i, s_i)\}$ ;
- 16: **end for**
- 17: **Return**  $T$ .

## Experiments

### Datasets

The original datasets are sourced from Pontiki et al. [33–35] in SemEval challenges, comprising 14restaurant, 14laptop,

15restaurant, and 16restaurant. The datasets include annotations for aspect terms along with their corresponding sentiment polarities. The datasets pertain to the comments from both the restaurant and laptop domains. However, for the Triplet task, annotations of the opinion terms in the datasets are required. To fulfill this requirement, we utilize the publicly available datasets provided by Peng et al. [21]<sup>4</sup>. Those datasets are annotated with (aspect term, opinion term, sentiment) triplet, building upon the foundation of the original datasets. Additionally, we utilized the publicly available datasets published by Xu et al. [23]<sup>5</sup>, a corrected version of Peng et al.’s datasets. Within these datasets, we identified overlapping opinion items and missing triples, and we have thoroughly rectified these issues.<sup>6</sup> The statistics for both datasets are presented in Table 2.

### Baselines

To evaluate the performance of our proposed Triple-MRC model, we use the aspect sentiment triplet extraction model proposed in recent years as the baseline model. All of these models are capable of implementing the Triplet task, and their details are described below.

<sup>4</sup> Datasets available at <https://github.com/xuuluuu/SemEval-Triplet-data>

<sup>5</sup> Datasets available at [https://github.com/xuuluuu/Position-Aware-Tagging-for-ASTE/tree/master/data/triplet\\_data](https://github.com/xuuluuu/Position-Aware-Tagging-for-ASTE/tree/master/data/triplet_data)

<sup>6</sup> Corrected dataset available: <https://github.com/ZouWang-spider/Triple-MRCdatasets>



**Table 2** The statistics of datasets

Datasets		14res	14lap	15res	16res
Peng et al.	train	1300	920	593	842
	dev	323	323	148	210
	test	496	339	318	320
Xu et al.	train	1266	906	605	857
	dev	310	219	148	210
	test	492	328	322	326

- **RINATE** (Bai et al. 2020 [10]) is a weakly supervised co-extraction method for AE and AOE tasks that make use of the dependency relations among words in a sentence.
- **CMLA** (Wang et al. 2017 [36]) is a multi-layer attention network designed for AE and AOE tasks. Each layer in the network is composed of pairs of attentions with tensor operators.
- **Li-unified-R** (Peng et al. 2020 [21]) is a modified version of Li-unified, initially designed for AESC through a unified tagging scheme. Li-unified-R specifically adjusts the original module for the AOE task.
- **Peng-two-stage** (Peng et al. 2020 [21]) is a two-stage framework employing distinct models for various sub-tasks in ABSA, standing out as the state-of-the-art method for the Triplet task.
- **JET-BERT** (Xu et al. 2020 [23]) is based on an end-to-end framework model, featuring a novel location-aware tagging scheme that facilitates the joint extraction of triplet.
- **GTS-BERT** (Wu et al. 2020 [12]) designs an effective grid tagging scheme (GTS) inference strategy, leveraging the mutual indication between different opinion factors to achieve more accurate opinion term extraction.
- **BART-ABSA** (Yan et al. 2021 [25]) transforms all ABSA subtasks into a unified generative formulation and addresses them within an end-to-end framework, employing the pre-trained sequence-to-sequence BART model.
- **MEC-GCN** (Chen et al. 2022 [37]) introduces an efficient refining strategy for word-pair representation refinement based on dependent syntactic features. This approach takes into consideration the implicit results of AE and AOE when determining whether word pairs match.
- **Dual-MRC** (Mao et al. 2021 [27]) is a dual BERT-MRC model based on the pipeline framework. In this model, the left model is employed for AE task, while the right component is dedicated to ASC and AOE tasks.
- **BMRC** (Chen and Qian 2020 [38]) transforms the Triplet task into a multi-round machine reading comprehension task and introduces a bidirectional machine reading comprehension (BMRC) framework to address this challenge.

## Hyper-parameter Setting

We employed the shared parameter learning approach in the Triple-MRC model. As a result, identical parameters are utilized for each segment of the BERT-MRC model, while other fine-tuning components are configured based on the specific tasks. We conducted an analysis of the experimental results to optimize the hyper-parameter settings. The hyper-parameter settings are shown in Table 3. The hyper-parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  of computing multi-task learning are not sensitive to the results, so we fix them at 1/3 in our experiments.

## Evaluation Metrics

For all tasks in our experiments, we employ the F1 score as an evaluation metric. The F1 score involves the calculation of the precision and recall for each category separately. The F1 score represents the harmonic mean when the precision is equal to the recall and is computed as follows.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (26)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (27)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (28)$$

where  $TP$  is the number of samples that are positive and the predicted outcome is positive.  $FP$  is the number of samples that are negative, and the predicted outcome is positive.  $FN$  is the number of samples that are positive, and the predicted outcome is negative.

**Table 3** The hyper-parameter settings

Hyper-parameter	Setting
BERT dim	768
Multi-Head	12
GCN layer	2
epochs	50
batch_size	32
Fully connected layer	128
L2 regularization	0.01
learning_rate	2e−5
dropout	0.1
optimizer	Adam

**Table 4** Comparison of the F1 score (%) for the Triplet task on the Peng et al. (2020) annotated datasets. The baseline results with “†” are retrieved from Mao et al. (2021), and results with “‡” are from Yan et al. (2021)

Frameworks	Models	14res	14lap	15res	16res
Pipeline	RINATE†	34.03	20.00	28.00	23.30
	CMLA†	43.12	32.90	35.90	41.60
	Li-unified-R†	51.68	42.47	46.69	44.51
	Peng-two-stage†	51.89	43.50	46.79	53.62
End-to-end	JET-BERT‡	63.92	50.00	54.67	62.98
	GTS-BERT‡	70.33	54.38	55.74	66.73
	BART-ABSA‡	72.46	57.59	60.11	69.98
	MEC-GCN‡	77.64	62.67	65.79	72.19
MRC	Dual-MRC†	70.32	58.58	57.21	67.40
	BMRC‡	75.20	61.68	63.88	69.61
Our	Triple-MRC	<b>78.52</b>	<b>63.25</b>	<b>66.42</b>	<b>72.85</b>

## Comparison Study

In the comparison experiment section, we conduct Triplet task on the two datasets using the baseline model and the proposed Triple-MRC model in this paper. The F1 scores obtained are presented in Tables 4 and 5. Analyzing the model’s performance from the comparative results in Tables 5 and 6, we can obtain several observational conclusions. Firstly, the Triple-MRC model surpasses the majority of baseline models on the datasets res14, lap14, res15, and res16. We attribute this superior performance to our model effectively leveraging the benefits of multi-task shared cascade learning and BERT-MRC models. Secondly, when compared to the Dual-MRC and BMRC models utilizing the MRC method, our Triple-MRC model demonstrates superior performance. This indicates that multi-task shared cascade learning contributes to establishing connections between the training of each task. Thirdly, in comparison to the JET-BERT, GTR-BERT, and BART-ABAS models based on

end-to-end methods, the Triple-MRC model exhibits superior performance. This shows that multi-task shared cascade learning can effectively mitigate the issue of contribution allocation. Finally, the Triple-MRC model outperforms the CMLA, Li-unified-R, and Peng-two-stage models based on the pipeline approach. This suggests that multi-task cascade learning can make full use of the prior knowledge in the questions, reducing the error propagation problem among tasks. Additionally, the assistance of the MRC method and dependency syntactic features benefits the model in the long-distance learning of aspect terms and opinion terms.

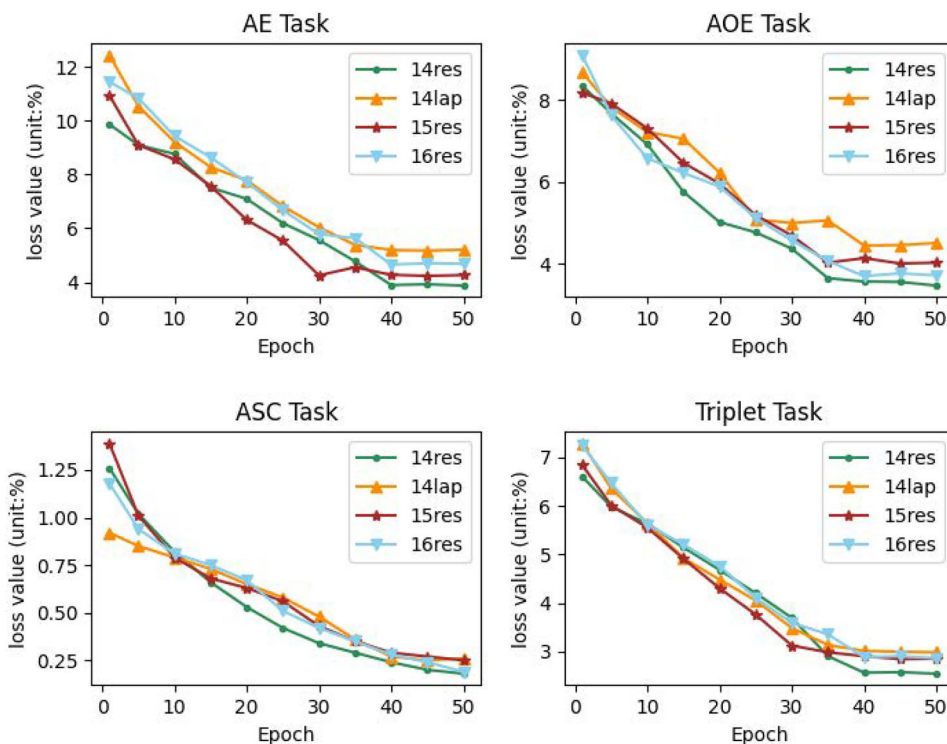
## Model Training Process

In the model training process section, we present the variation curves of the loss function for the Triple-MRC model trained on the 14res, 14lap, 15res, and 16res datasets (Peng et al.). To validate the effectiveness of multi-task shared cascade learning, the experiments showcase the loss function variation

**Table 5** Comparison of the F1 score (%) for the Triplet task on the Xu et al. (2020) annotated datasets. The baseline results with “†” are retrieved from Yan et al. (2021), and results with “‡” are from Chen et al. (2022)

Frameworks	Models	14res	14lap	15res	16res
Pipeline	RINATE†	34.95	20.07	29.97	23.87
	CMLA†	42.79	33.16	37.01	41.72
	Li-unified-R†	51.10	42.34	47.82	44.31
	Peng-two-stage†	51.46	42.87	52.32	54.21
End-to-end	JET-BERT†	62.40	51.04	57.53	63.83
	GTS-BERT‡	68.81	55.42	58.60	67.58
	BART-ABSA†	65.25	58.69	59.26	67.62
	MEC-GCN‡	71.78	58.81	61.93	68.33
MRC	Dual-MRC†	67.03	57.63	60.15	67.42
	BMRC‡	67.99	57.82	60.02	65.75
Our	Triple-MRC	<b>72.45</b>	<b>60.72</b>	<b>62.86</b>	<b>68.65</b>

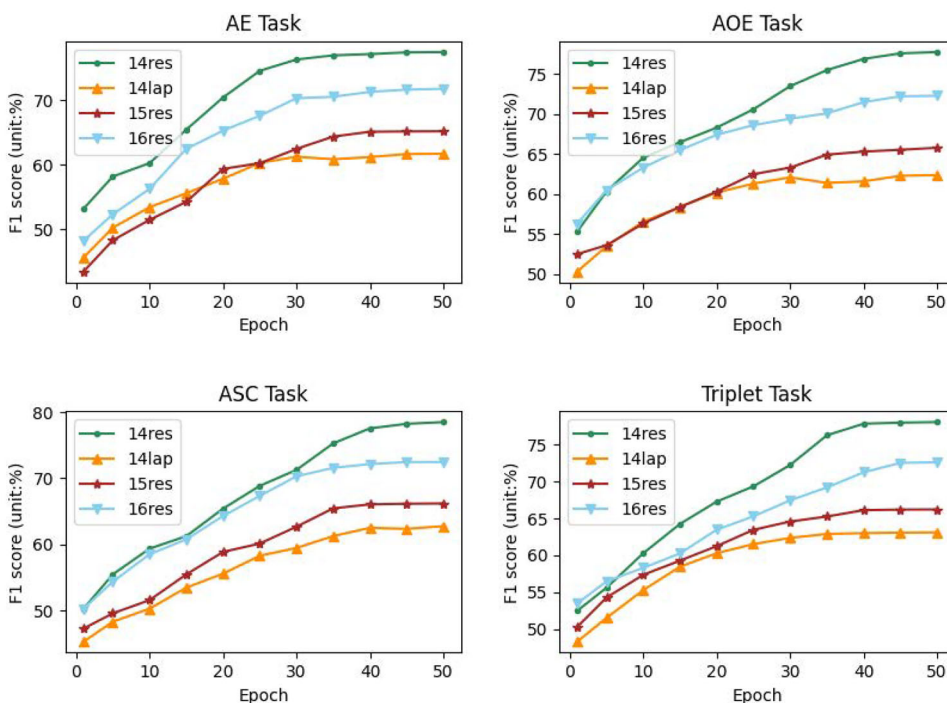
**Fig. 5** Loss value variation curves of the Triple-MRC model training process on 14res, 14lap, 15res, and 16res (Peng et al.) datasets



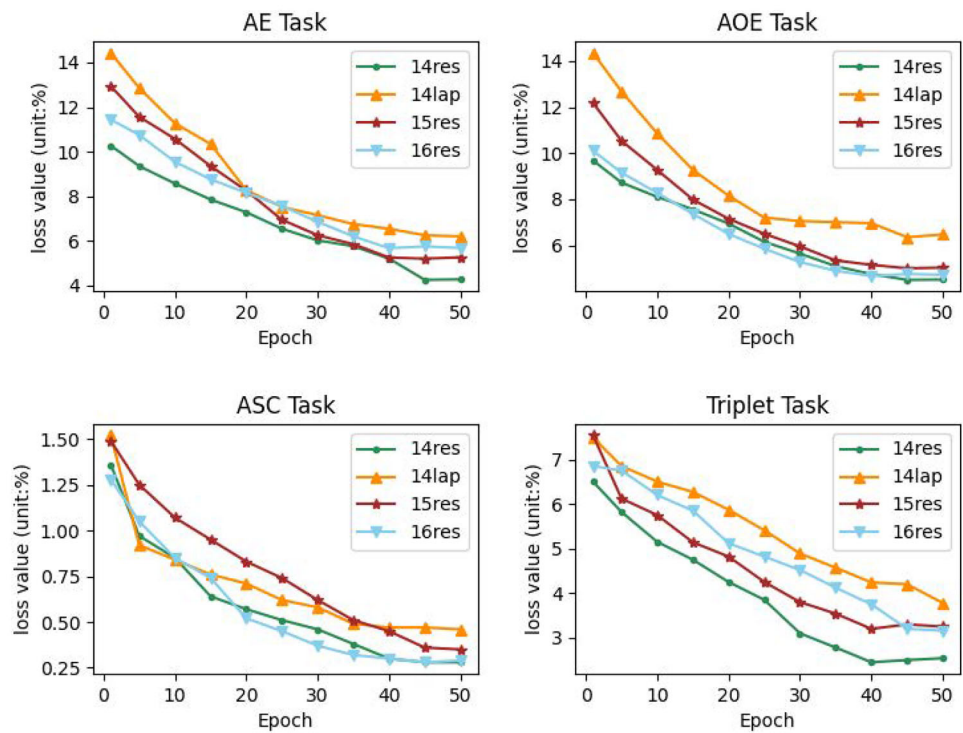
curves of the Triple-MRC model for the AE task, AOE task, ASC task, and Triplet Task throughout the training process. Upon analyzing the experimental results depicted in Figs. 5, 6, 7, and 8, it is observed that the model exhibits significantly larger loss values for AE and AOE compared to ASC during training. To analyze the reason, the AE and AOE tasks

belong to the sequence annotation task, essentially involving binary classification for each sequence position to determine whether it is the start or end position. The cross-entropy function is then employed to calculate the loss value for the probability of all sequence positions concerning the true value. Consequently, the loss value of all predicted sequence

**Fig. 6** F1 score variation curves of the Triple-MRC model training process on 14res, 14lap, 15res, and 16res (Peng et al.) datasets



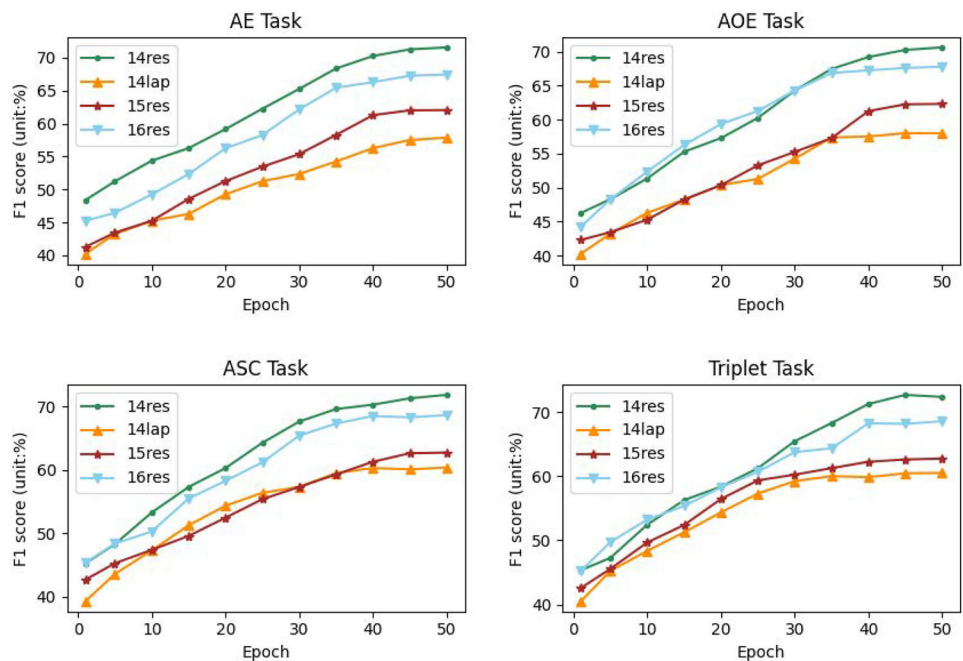
**Fig. 7** Loss value variation curves of the Triple-MRC model training process on 14res, 14lap, 15res, and 16res (Xu et al.) datasets



positions is summed, resulting in larger loss values. In contrast, the ASC task is the sentiment triple classification task using the cross-entropy loss function, which calculates the loss value for the sentiment probability distribution between 0 and 1. Additionally, from the change in the loss value graph, it is evident that the multi-task shared cascade model effectively reduces the overall loss value of the Triple-MRC model across the four datasets.

From the data in Tables 6 and 7, it is evident that the F1 score achieved on the AOE task is higher than on the AE task. Analyzing the reason, the result of extracting aspect terms from the AE task was utilized as prior knowledge when the model was trained for the AOE task. Therefore, the AOE task attains a higher F1 score compared to the AE task. Similarly, the ASC task demonstrates superior performance as aspect terms and opinion terms are serve as prior knowl-

**Fig. 8** F1 score variation curves of the Triple-MRC model training process on 14res, 14lap, 15res, and 16res (Xu et al.) datasets



**Table 6** The Triple-MRC model of each task’s achieved F1 score (%) on the Peng et al. datasets

Datasets	AE task	AOE task	ASC task	Triplet task
14res	77.40	77.71	78.50	<b>78.52</b>
14lap	61.70	62.33	62.75	<b>63.25</b>
15res	65.18	65.75	66.19	<b>66.42</b>
16res	71.73	72.25	72.47	<b>72.85</b>

**Table 7** The Triple-MRC model of each task’s achieved F1 score (%) on the Xu et al. datasets

Datasets	AE Task	AOE Task	ASC Task	Triplet Task
14res	71.55	70.62	71.79	<b>72.45</b>
14lap	57.85	58.00	60.35	<b>60.72</b>
15res	62.04	62.30	62.70	<b>62.86</b>
16res	67.43	67.79	68.60	<b>68.65</b>

edge during model training. The Triple-MRC model attains an enhanced F1 score on the Triplet task compared to the ASC task, indicating that the extraction opinion terms assist the model in discerning the aspect terms sentiment polarity. This also validates the effectiveness of using a multi-task learning approach for the Triplet task.

**Error Analysis**

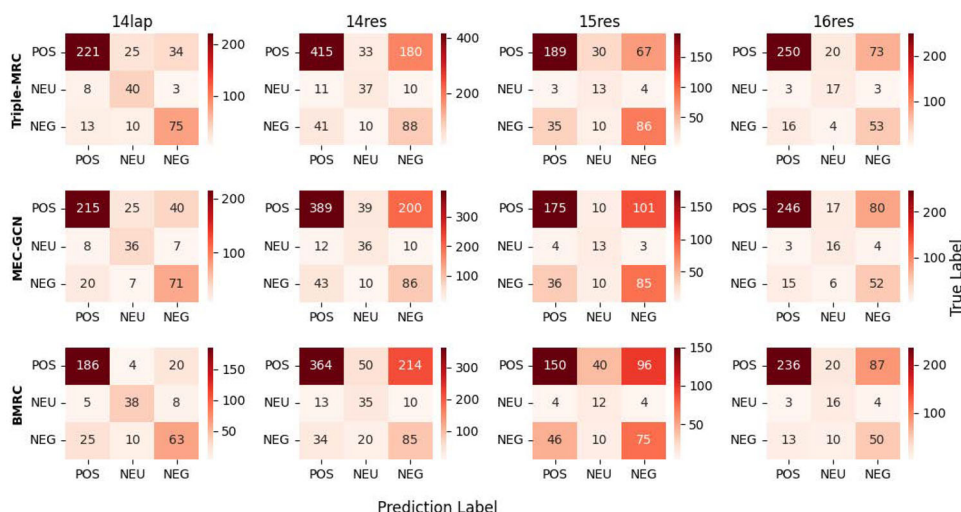
To evaluate the model’s performance in computing sentiment polarity, we conducted an error analysis experiment. The experimental results, depicted in Figs. 9 and 10, illustrate the confusion matrices of the Triple-MRC, MEC-GCN, and BMRC models on the test datasets, where the “POS” represents positive sentiment, “NEU” represents neutral sen-

timent, and “NEG” represents negative sentiment. In Fig. 9, the predictions of the Triple-MRC model on the Peng et al. test datasets exhibit a significant advantage over those of the MEC-GCN and BMRC models. The performance ranking of the models is Triple-MRC >MEC-GCN >BMRC. Turning to the comparison in Fig. 10, it is evident that on the Xu et al. test datasets, the predictions of the Triple-MRC model are consistently optimal. The predictions of the MEC-GCN and BMRC models are similar in performance. The visualization of the confusion matrices on the test datasets indicates the overall superior performance of our proposed Triple-MRC model.

**Ablation Study**

The ablation study aims to assess the impact of syntactic features and multi-task shared cascade learning on the overall model performance. The results of the ablation experiment are presented in Tables 8 and 9, along with Figs. 11 and 12, where “w/o” denotes components removed from the model. In the “w/o multi-task shared cascade learning” design, we employ the strategy of removing prior knowledge from the queries. Query 2 is “Find the opinion terms for the aspect terms in the text,” and Query 3 is “Find the sentiment polarity of the opinion terms related to the aspect terms in the text.” Ablation experiment results on Peng et al. datasets reveal that removing the syntactic features component leads to an average performance decrease of 1.13%, while removing the multi-task shared cascade learning component results in an average decrease of 3.96%. In the ablation experiment results on Xu et al. datasets, removing the two components led to an average performance decrease of 1.08% and 3.84%, respectively. The findings suggest that both components can effectively enhance the model’s performance, with multi-task shared cascade learning exerting a greater impact on the model.

**Fig. 9** The experimental results of the confusion matrix on Peng et al. datasets



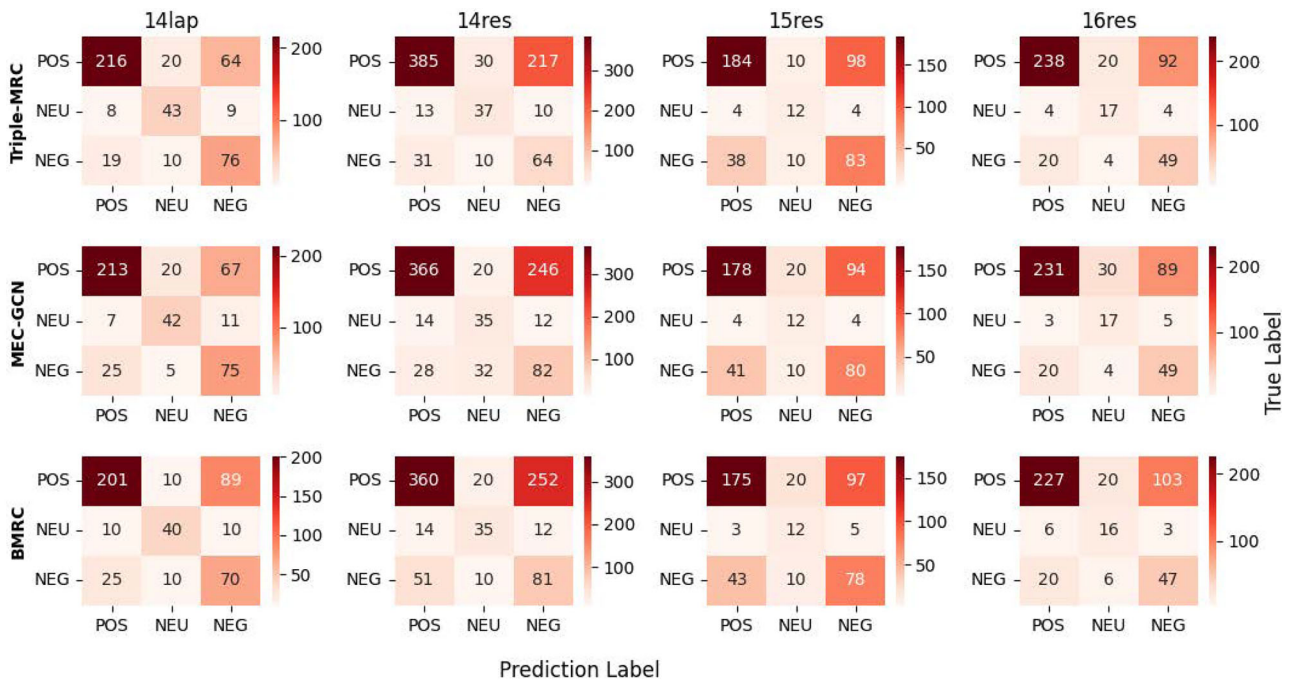


Fig. 10 The experimental results of the confusion matrix on Xu et al. datasets

Table 8 The F1 score (%) achieved by ablation experiment on the Peng et al. datasets

Model	14res	14lap	15res	16res
Triple-MRC	78.52	63.25	66.42	72.85
<i>w/o</i> syntactic features	(77.27)↓ <b>1.25</b>	(62.10)↓ <b>1.15</b>	(65.52)↓ <b>0.90</b>	(71.62)↓ <b>1.23</b>
<i>w/o</i> multi-task shared cascade learning	(74.02)↓ <b>4.50</b>	(59.63)↓ <b>3.62</b>	(62.97)↓ <b>3.45</b>	(68.58)↓ <b>4.27</b>

### Attention Visualization

To verify the impact of designing three types of questions in the BERT-MRC model on training improvement, the experiments demonstrate the visualization of attention [39] distribution for questions and sentences. The BERT-MRC model comprises 12 attention heads, and the experiments present results depicting representative attention distributions among them. The attention visualization for Question 1 and the sentence during the training of the AE task is depicted in Fig. 13. The observation reveals higher attention scores for “aspect terms” and “food,” indicating the effectiveness of the Ques-

tion 1 design. The attention visualization for Question 2 and the sentence during the training of the AOE task is illustrated in Fig. 14. Notably, the scores of “opinion terms” and “pretty good” are higher. The term “food” in Question 2 also receives a relatively high attention score along with “pretty good” in the sentence. This verifies that incorporating “food” as prior knowledge in Question 2 contributes to the model in learning the relationship between aspect terms and opinion terms. The attention visualization of Question 3 and the sentence when training the ASC task is depicted in Fig. 15. The attention scores of “sentiment polarity” in Question 3 and “pretty good” in the sentence are higher. Additionally, the scores

Table 9 The F1 score (%) achieved by ablation experiment on the Xu et al. datasets

Model	14res	14lap	15res	16res
Triple-MRC	72.45	60.72	62.86	68.65
<i>w/o</i> syntactic features	(71.35)↓ <b>1.10</b>	(59.52)↓ <b>1.20</b>	(61.91)↓ <b>0.95</b>	(67.57)↓ <b>1.08</b>
<i>w/o</i> multi-task shared cascade learning	(67.28)↓ <b>5.17</b>	(56.90)↓ <b>3.82</b>	(60.11)↓ <b>2.75</b>	(65.03)↓ <b>3.62</b>

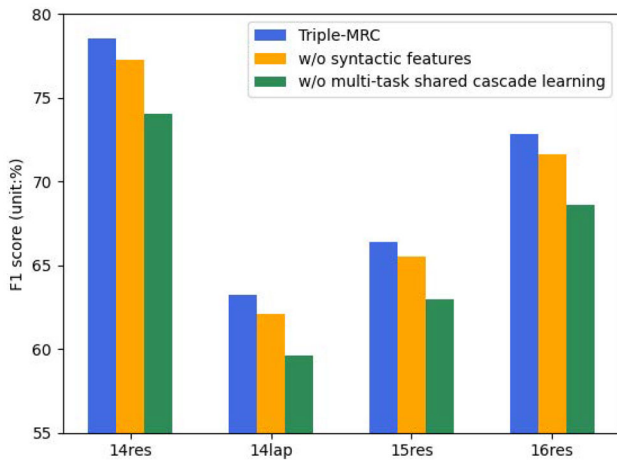


Fig. 11 The results of the ablation experiment on the Peng et al. datasets

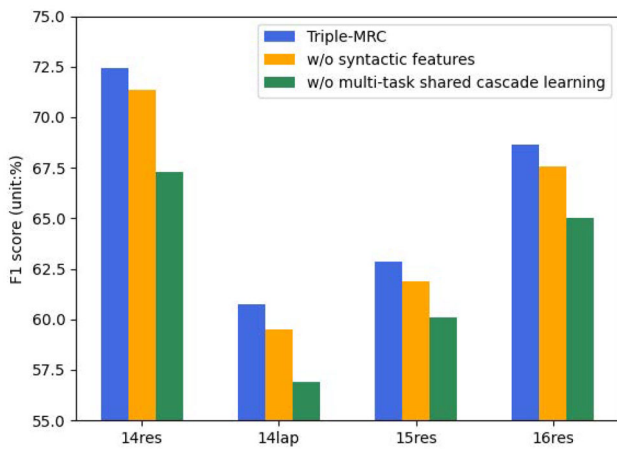


Fig. 12 The results of the ablation experiment on the Xu et al. datasets

Fig. 14 Attention visualization for AOE task in the Triple-MRC model

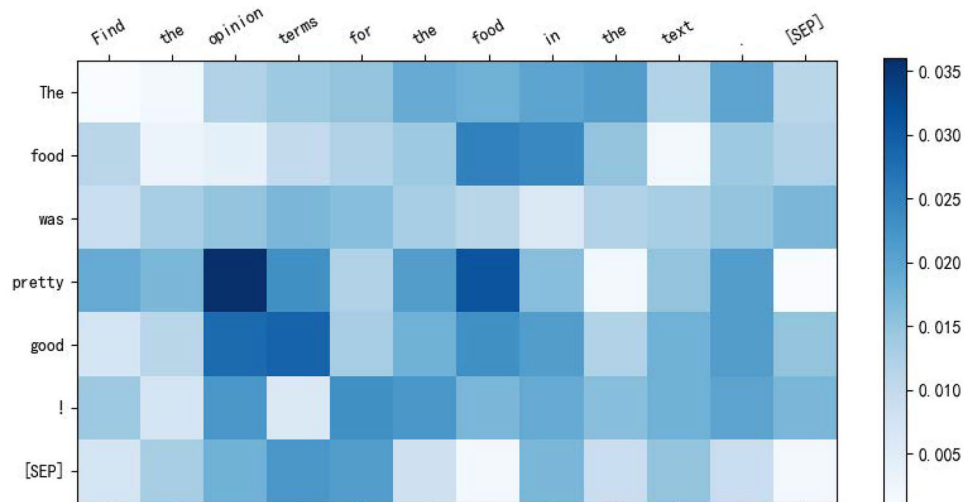


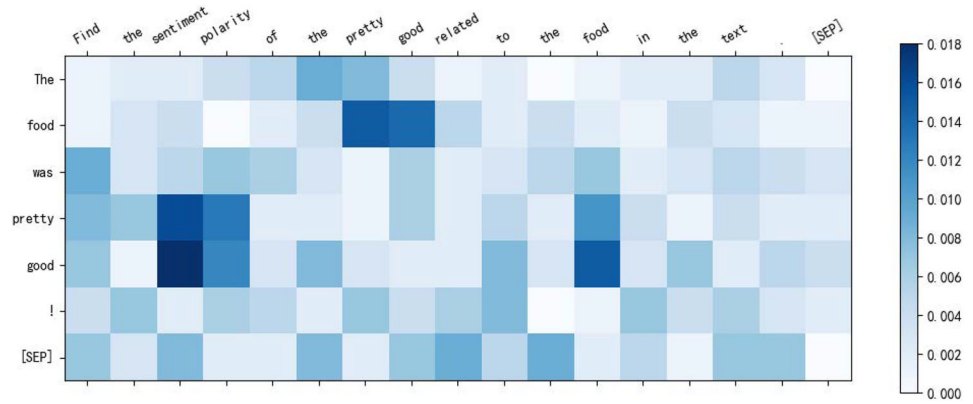
Fig. 13 Attention visualization for AE task in the Triple-MRC model

between aspect terms and opinion terms are also elevated. This confirms that the prior knowledge in Question 3 significantly aids the model in determining sentiment polarity.

### Case Study

The case study selected the MEC-GCN model and BMRC model as the comparison models. The case sentences were respectively chosen from the computer domain and the restaurant domain, and were extracted from online reviews on the website. The experimental results for the case study are presented in Table 10, where the MEC-GCN model failed to accurately determine the sentiment polarity of the “dinner special” in the fifth sentence. The BMRC model failed to accurately extract the multi-word aspect terms such as “blond wood decor” in the second sentence and “17 inch screen” in the sixth sentence. Furthermore, BMRC struggled to accu-

**Fig. 15** Attention visualization for ASC task in the Triple-MRC model



**Table 10** Case study of Triplet task, where the wrong predictions are marked with **X**. The blue highlighted words are aspect terms, and the red highlighted words are opinion terms

id	Sentence
1	The <b>speed</b> is incredible and I am more than <b>satisfied</b> . MEC-GCN: (speed, satisfied, POS) BMRC: (speed, satisfied, POS) Triple-MRC: (speed, satisfied, POS)
2	The <b>blond wood decor</b> is very <b>soothing</b> , the <b>premium sake</b> is <b>excellent</b> and the <b>service</b> is <b>great</b> . MEC-GCN: (blond wood decor, soothing, POS), (premium sake, excellent, POS), (service, great, POS) BMRC: (blond wood, soothing, NEU) <b>X</b> , (premium sake, excellent, POS), (service, great, POS) Triple-MRC: (blond wood decor, soothing, POS), (premium sake, excellent, POS), (service, great, POS)
3	The <b>service</b> is <b>excellent</b> , the <b>decor</b> is <b>great</b> , and the textcolorbluefood is <b>delicious</b> . MEC-GCN: (service, excellent, POS), (decor, great, POS), (food, delicious, POS) BMRC: (service, excellent, POS), (decor, great, POS), (food, delicious, POS) Triple-MRC: (service, excellent, POS), (decor, great, POS), (food, delicious, POS)
4	Sometimes the <b>screen</b> even <b>goes black</b> on this computer. MEC-GCN: (screen, goes black, NEG) BMRC: (screen, goes black, NEG) Triple-MRC: (screen, goes black, NEG)
5	Similar to other Indian restaurants, they use the <b>dinner special</b> to <b>attract customers</b> at the door. MEC-GCN: (dinner special, attract customers, POS) <b>X</b> BMRC: (dinner special, attract customers, NEU) Triple-MRC: (dinner special, attract customers, NEU)
6	The <b>17 inch screen</b> is <b>very large</b> , but the <b>computer</b> is <b>very light</b> . MEC-GCN: (17 inch screen, very large, NEU),(computer, very light, POS) BMRC: (screen, very large, POS) <b>X</b> , (computer, very light, POS) Triple-MRC: (17 inch screen, very large, NEU) (computer, very light, POS)

rately determine the sentiment polarities of these two aspect terms. In contrast, our Triple-MRC model excelled in accurately extracting both aspect terms and opinion terms while determining sentiment polarity. Analyzing the reasons, both the MEC-GCN model and our Triple-MRC utilize syntactic features. This is advantageous for the models to better learn the dependency relationships between words, thereby accurately extracting aspect terms and opinion terms. Moreover, our model employs a multi-task shared cascade learning structure that integrates prior knowledge, allowing for accurate sentiment polarity determination. The case study results once again validate the superior overall performance of the Triple-MRC model.

### Conclusion

In this paper, we propose a Triple-MRC framework for aspect sentiment triplet extraction. The model subdivides the Triplet task into AE, AOE, and ASC tasks, employing a multi-task shared cascade learning framework and the MRC method to implement all sub-tasks. The dependency syntactic features contribute to the model’s better understanding of the syntactic information in sentences. Additionally, we have also designed three questions as prior knowledge in the BERT-MRC model to train the model. The comparison study, model training process, error analysis, ablation study, attention visualization, and case study are conducted to demonstrate the performance of the Triple-MRC model. The experimental results demonstrate the superiority of our proposed framework over all compared baselines. In future work, we consider introducing part-of-speech features [40] of sentences. This enhancement aims to assist the BERT-MRC model in achieving a more accurate extraction of multi-word aspect terms and multi-word opinion terms.

**Funding** This study was funded by the Hubei Province Key Research Project (grant number No. TA02002), the Hubei Provincial Central Leading Local Science and Technology Development Special Project (grant number No. 2018ZYDD007), the Hubei Province Education



Department Science and Technology Research Project (grant number No. Q20201801), and the Hubei University of Automotive Technology PhD Research Start-up Fund Project (grant number No. BK202004).

**Data Availability** The data is available at (1) Peng et al. annotated dataset <https://github.com/xuuluuu/SemEval-Triplet-data>, (2) Xu et al. annotated dataset <https://github.com/xuuluuu/Position-Aware-Tagging-for-ASTE/tree/master/data/triplet-data>, and (3) our corrected annotated dataset <https://github.com/ZouWang-spider/Triple-MRCdatasets>.

**Code Availability** The code is available at <https://github.com/ZouWang-spider/Triple-MRC>.

## Declarations

**Ethical Approval** This article does not contain any studies with human participants or animals performed by any of the authors.

**Informed Consent** Informed consent was obtained from all individual participants included in the study.

**Competing Interest** The authors declare no competing interests.

## References

- Jiang Q, Chen L, Xu R, Xiang A, Min Y. A challenge dataset and effective models for aspect-based sentiment analysis. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). ACL; 2019. p. 6280–5.
- Hu Z, Wang Z, Wang Y, Tan AH. Aspect sentiment triplet extraction incorporating syntactic constituency parsing tree and common-sense knowledge graph. *Cogn Comput*. 2023;15(1):337–47.
- Yang L, Na JC, Yu J. Cross-modal multitask transformer for end-to-end multimodal aspect-based sentiment analysis. *Inf Process Manage*. 2022;59(5):103038.
- De Clercq O, Lefever E, Jacobs G, Carpels T, Hoste V. Towards an integrated pipeline for aspect-based sentiment analysis in various domains. In: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA). ACL; 2017. p. 136–42.
- Zhang W, Li X, Deng Y, Bing L, Lam W. A survey on aspect-based sentiment analysis: tasks, methods, and challenges. *IEEE Trans Knowl Data Eng*. 2022;35(11):11019–38.
- Du Y, Jin X, Yan R, Yan J. Sentiment-enhanced answer generation and information fusing for product-related question answering. *Inf Sci*. 2023;627:205–19.
- Zhou J, Huang JX, Hu QV, He L. Is position important? Deep multi-task learning for aspect-based sentiment analysis. *Appl Intell*. 2020;50:3367–78.
- Zhao G, Luo Y, Chen Q, Qian X. Aspect-based sentiment analysis via multitask learning for online reviews. *Knowl-Based Syst*. 2023;264:110326.
- Wang X, Xu G, Zhang Z, Li J, Sun X. End-to-end aspect-based sentiment analysis with hierarchical multi-task learning. *Neurocomputing*. 2021;455:178–88.
- Bai X, Liu P, Zhang Y. Investigating typed syntactic dependencies for targeted sentiment classification using graph attention neural network. *IEEE/ACM Trans Audio Speech Language Process*. 2020;29:503–14.
- Dai H, Song Y. Neural aspect and opinion term extraction with mined rules as weak supervision. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL). ACL; 2019. p. 5268–77.
- Wu Z, Ying C, Zhao F, Fan Z, Dai X, Xia R. Grid tagging scheme for aspect-oriented fine-grained opinion extraction. In: Findings of the Association for Computational Linguistics (ACL). ACL; 2020. p. 2576–85.
- Ji Q, Lin X, Ma Y, Liu G, Wang S. A unified labeling model for open-domain aspect-based sentiment analysis. In: 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC). IEEE; 2020. p. 186–9.
- Chen T, Xu R, He Y, Wang X. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Syst Appl*. 2017;72:221–30.
- Kumar A, Balan R, Gupta P, Neti LBM, Malapati A. BILEAT: a highly generalized and robust approach for unified aspect-based sentiment analysis: BILEAT. *Appl Intell*. 2022;52(12):14025–40.
- Zhang Y, Ding Q, Zhu Z, Liu P, Xie F. Enhancing aspect and opinion terms semantic relation for aspect sentiment triplet extraction. *J Intell Inf Syst*. 2022;59(2):523–42.
- He K, Mao R, Gong T, Li C, Cambria E. Meta-based self-training and re-weighting for aspect-based sentiment analysis. *IEEE Trans Affect Comput*. 2022;14(3):1731–42.
- Wang F, Lan M, Wang W. Towards a one-stop solution to both aspect extraction and sentiment analysis tasks with neural multi-task learning. In: 2018 International joint conference on neural networks (IJCNN). IEEE; 2018. p. 1–8.
- Huan H, He Z, Xie Y, Guo Z. A multi-task dual-encoder framework for aspect sentiment triplet extraction. *IEEE Access*. 2022;10:103187–99.
- Ma D, Li S, Wang H. Joint learning for targeted sentiment analysis. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP). ACL; 2018. p. 4737–42.
- Peng H, Xu L, Bing L, Huang F, Lu W, Si L. Knowing what, how and why: a near complete solution for aspect-based sentiment analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). AAAI. 2020;34(05):8600–7.
- Phan HT, Nguyen NT, Hwang D. Aspect-level sentiment analysis: a survey of graph convolutional network methods. *Inf Fusion*. 2023;91:149–72.
- Xu L, Li H, Lu W, Bing L. Position-aware tagging for aspect sentiment triplet extraction. [arXiv:2010.02609](https://arxiv.org/abs/2010.02609) [Preprint]. 2020. Available from: <http://arxiv.org/abs/2010.02609>.
- Sarawagi S, Cohen WW. SemiMarkov conditional random fields for information extraction. In: Advances in Neural Information Processing Systems 17 Neural Information Processing Systems (NIPS). NIPS; 2004. p. 1185–92.
- Yan H, Dai J, Ji T, Qiu X, Zhang Z. 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 (ACL-IJCNLP). ACL; 2021. p. 2416–29.
- Xiong H, Yan Z, Wu C, Lu G, Pang S, Xue Y, Cai Q. BART-based contrastive and retrospective network for aspect-category-opinion-sentiment quadruple extraction. *Int J Mach Learn Cybern*. 2023; 1–13.
- Mao Y, Shen Y, Yu C, Cai L. A joint training dual-MRC framework for aspect based sentiment analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). AAAI. 2021;35(15):13543–51.
- Li X, Feng J, Meng Y, Han Q, Wu F, Li J. A unified MRC framework for named entity recognition. [arXiv:1910.11476](https://arxiv.org/abs/1910.11476) [Preprint]. 2019. Available from: <http://arxiv.org/abs/1910.11476>.
- Batra S, Rao D. Entity based sentiment analysis on Twitter. *Science*. 2010;9(4):1–12.

30. Dozat T, Manning CD. Deep biaffine attention for neural dependency parsing. [arXiv: 1611.01734](https://arxiv.org/abs/1611.01734) [Preprint]. 2016. Available from: <http://arxiv.org/abs/1611.01734>.
31. Soleymani M, Garcia D, Jou B, Schuller B, Chang S, Pantic M. A survey of multimodal sentiment analysis. *Image Vis Comput.* 2017;65:3–14.
32. Devlin J, Chang MW, Lee K, Toutanova K. 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 (NAACL-HLT). ACL; 2019. p. 4171–86.
33. Pontiki M, Galanis D, Pavlopoulos J, Papageorgiou H, Androutsopoulos I, Manandhar S. SemEval-2014 task 4: aspect based sentiment analysis. In: SemEval@COLING. SemEval. 2014;2014: 27–35.
34. Pontiki M, Galanis D, Papageorgiou H, Manandhar S, Androutsopoulos I. SemEval-2015 task 12: aspect based sentiment analysis. In: SemEval@NAACLHLT. SemEval; 2015. p. 486–95.
35. 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, Eryigit G. SemEval-2016 task 5: aspect based sentiment analysis. In: SemEval@NAACLHLT. SemEval; 2016. p. 19–30.
36. Wang W, Pan SJ, Dahlmeier D, Xiao X. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). AAAI. 2017;31(1).
37. Chen H, Zhai Z, Feng F, Li R, Wang X. Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL). ACL; 2022. p. 2974–85.
38. Chen Z, Qian T. Enhancing aspect term extraction with soft prototypes. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). ACL; 2020. p. 2107–17.
39. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. *Adv Neural Inf Process Syst.* 2017;30.
40. Zou W, Zhang W, Tian Z, Wu H. A hybrid model for text classification using part-of-speech features. *J Intell Fuzzy Syst.* 2023;45(1):1235–49.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.