

# Enhanced Multi-Channel Graph Convolutional Network for Aspect Sentiment Triplet Extraction

Hao Chen\*, Zepeng Zhai\*, Fangxiang Feng, Ruifan Li†, Xiaojie Wang

School of Artificial Intelligence, Beijing University of Posts and Telecommunications, China  
{ccchenhao997, zepeng, fxfeng, rfl1, xjwang}@bupt.edu.cn

## Abstract

*Aspect Sentiment Triplet Extraction* (ASTE) is an emerging sentiment analysis task. Most of the existing studies focus on devising a new tagging scheme that enables the model to extract the sentiment triplets in an end-to-end fashion. However, these methods ignore the relations between words for ASTE task. In this paper, we propose an *Enhanced Multi-Channel Graph Convolutional Network* model (EMC-GCN) to fully utilize the relations between words. Specifically, we first define ten types of relations for ASTE task, and then adopt a **biaffine attention** module to embed these relations as an adjacent tensor between words in a sentence. After that, our EMC-GCN transforms the sentence into a multi-channel graph by treating words and the relation adjacent tensor as nodes and edges, respectively. Thus, relation-aware node representations can be learnt. Furthermore, we consider diverse linguistic features to enhance our EMC-GCN model. Finally, we design an effective refining strategy on EMC-GCN for word-pair representation refinement, which considers the implicit results of aspect and opinion extraction when determining whether word pairs match or not. Extensive experimental results on the benchmark datasets demonstrate that the effectiveness and robustness of our proposed model, which outperforms state-of-the-art methods significantly.<sup>1</sup>

## 1 Introduction

Aspect Sentiment Triplet Extraction (ASTE) is a new variant of Aspect-based Sentiment Analysis (ABSA). The ASTE task aims to extract aspect sentiment triplets from a sentence, and **each triplet contains three elements, namely aspect term, opinion term and their associated sentiment**. In Figure 1, an example illustrates the definition of ASTE.

\*These authors contributed equally to this work.

†Corresponding author.

<sup>1</sup>Code and datasets are available at <https://github.com/CCChenhao997/EMCGCN-ASTE>.

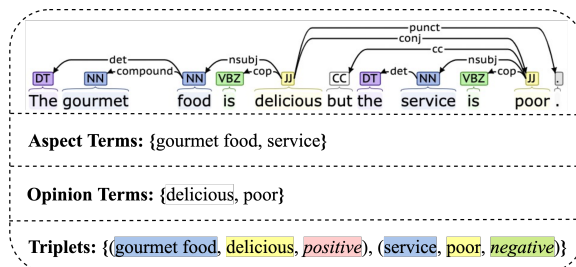


Figure 1: A sentence with its dependency tree is given to illustrate ASTE task. In the triplet set, aspect terms, opinion terms are highlighted in blue and yellow, respectively. The *positive* sentiment polarity is highlighted in red, while the *negative* in green.

To extract the triplets, previous studies have developed three types of approaches. Pipeline approaches (Peng et al., 2020) independently extract elements of the triplet. However, such techniques ignore the interaction between them, and potentially lead to error propagation and extra costs. To utilize the associations among the multiple subtasks, Mao et al. (2021) and Chen et al. (2021a) formulate the ASTE task as a multi-turn machine reading comprehension (MRC) problem and design a model based on BERT to jointly train multiple subtasks. Meanwhile, some efforts devote to extracting the triplets in an end-to-end framework (Xu et al., 2020; Wu et al., 2020a; Zhang et al., 2020; Chen et al., 2021b; Yan et al., 2021), which is constructed mainly by designing new tagging scheme. Although previous works have achieved significant fruits, there exists still several challenges.

Here, two questions arise naturally for ASTE task by our observations. **1) How to utilize various relations between words to help ASTE task?** Take Figure 1 as an example; for word pair ("gourmet", "food"), "gourmet" and "food" belong to the same aspect term "gourmet food". Likewise, for word pair ("food", "delicious"), "food" is an opinion target of "delicious" and is endowed with a positive sentiment polarity. Therefore, to

effectively extract the aspect term “*gourmet food*”, we expect that “*gourmet*” can obtain the information of “*food*” and vice versa. To judge the sentiment polarity of the aspect term, information of the opinion term “*delicious*” should be delivered to “*gourmet food*”. In short, we need to learn task-dependent word representations based on the relations between words. **2) How to utilize the linguistic features to help ASTE task?** First, we observe that aspect terms “*gourmet food*” and “*service*” are nouns, while opinion terms “*delicious*” and “*poor*” are adjectives. Thus, the word pair composed of a noun and an adjective tend to form aspect-opinion pair. Second, from the syntactic dependency tree in Figure 1, different dependency types exist in word pairs. For instance, “*gourmet*” and “*food*” comprise a compound noun because the dependency type between them is “*compound*”, while “*food*” is the nominal subject of “*delicious*” due to the type “*nsubj*”. Thus, these dependency types can help not only the extraction of aspect and opinion terms but also their matching<sup>2</sup>. In addition, we consider the tree-based and relative position distances which describe the relevance of two words.

In this paper, we propose a novel architecture, *Enhanced Multi-Channel Graph Convolutional Network* model (EMC-GCN), to answer the aforementioned questions. **Firstly**, we utilize a biaffine attention module to model the relation probability distribution between words in a sentence and use a vector to represent it. Each dimension in the vector corresponds to a certain relation type. To this end, we can derive a relation adjacency tensor from a sentence. Furthermore, our EMC-GCN transforms the sentence to a multi-channel graph by treating words and the relation adjacency tensor as nodes and edges, respectively. In order to learn precise relation between words, we impose relation constraint on the relation adjacency tensor. **Secondly**, to exploit linguistic features, including lexical and syntactic information, we obtain the part-of-speech combination, syntactic dependency type, tree-based distance and relative position distance of each word pair in the sentence. Similarly, we respectively transform these features into the edges for the multi-channel graphs to further enhance our model. Although part of linguistic fea-

tures has been applied in other tasks (Kouloumpis et al., 2011; Sun et al., 2019; Phan and Ogunbona, 2020; Li et al., 2021), to the best of our knowledge, they are rarely used in ASTE task. It is non-trivial to explore various linguistic features, adapt and apply them to ASTE in a novel way. **Thirdly**, inspired by the classifier chains method (Read et al., 2011) in multi-label classification task, we devise an effective refining strategy. Our strategy considers the implicit results of aspect and opinion extraction for word-pair representation refinement when judging whether word pairs match.

Our contributions are highlighted as follows:

- 1) We propose a novel EMC-GCN model for ASTE task. EMC-GCN exploits the multi-channel graph to encode relations between words. Convolution function over the multi-channel graph is applied to learn relation-aware node representations.
- 2) We propose a novel way to fully develop linguistic features to enhance our GCN-based model, including the part-of-speech combination, syntactic dependency type, tree-based distance and relative position distance of each word pair in a sentence.
- 3) We propose an effective refining strategy for refined word-pair representation. It considers the implicit results of aspect and opinion extraction when detecting if word pairs match.
- 4) We conduct extensive experiments on benchmark datasets. The experimental results show the effectiveness of our EMC-GCN model.

## 2 Related Work

Traditional sentiment analysis tasks are sentence-level (Yang and Cardie, 2014; Severyn and Moschitti, 2015) or document-level (Dou, 2017; Lyu et al., 2020) oriented. In contrast, *Aspect-based Sentiment Analysis* (ABSA) is an aspect or entity oriented fine-grained sentiment analysis task. The most three basic subtasks are *Aspect Term Extraction* (ATE) (Hu and Liu, 2004; Yin et al., 2016; Xu et al., 2018; Ma et al., 2019; Chen and Qian, 2020; Wei et al., 2020), *Aspect Sentiment Classification* (ASC) (Tang et al., 2016; Ma et al., 2017; Li et al., 2018; Zhang et al., 2019; Wang et al., 2020; Li et al., 2021) and *Opinion Term Extraction* (OTE) (Yang and Cardie, 2012, 2013; Fan et al., 2019; Wu et al., 2020b). The studies solve these tasks separately and ignore the dependency between these subtasks. Therefore, some efforts devoted to couple the two subtasks and proposed effective models to jointly extract aspect-based pairs.

<sup>2</sup>Matching of word-pair denotes that given  $w_i$  and  $w_j$  which respectively belong to an aspect term and an opinion term, if the aspect term and the opinion term form a triplet, then word-pair  $(w_i, w_j)$  matches.

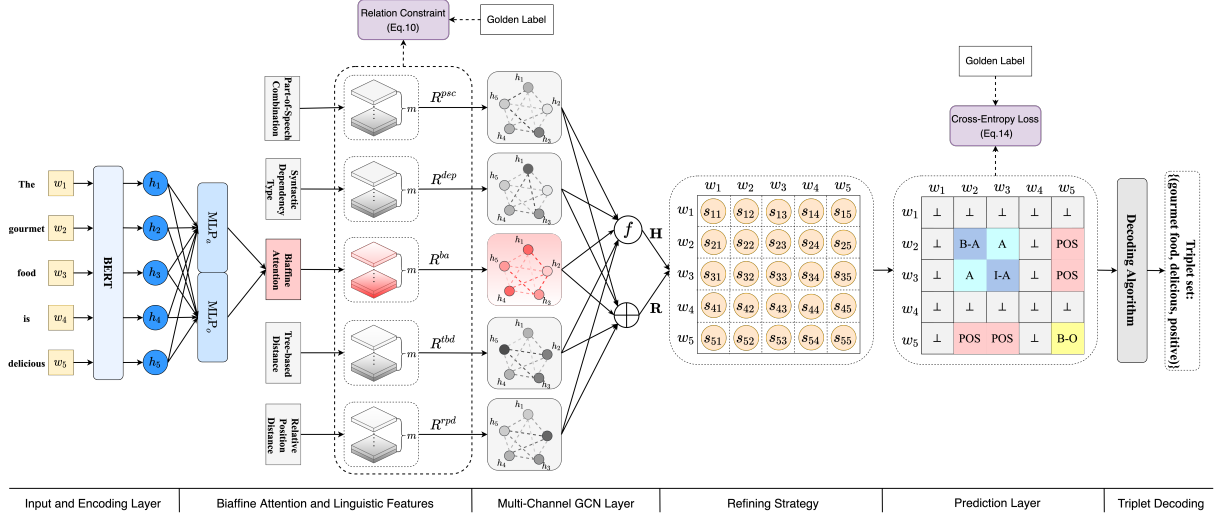


Figure 2: The overall architecture of our end-to-end model EMC-GCN.

This kind of work mainly has two tasks: *Aspect and Opinion Term Co-Extraction* (AOTE) (Wang et al., 2016, 2017; Dai and Song, 2019; Wang and Pan, 2019; Chen et al., 2020b; Wu et al., 2020a) and *Aspect-Sentiment Pair Extraction* (ASPE) (Ma et al., 2018; Li et al., 2019a,b; He et al., 2019).

Most recently, Peng et al. (2020) first proposed the ASTE task and developed a two-stage pipeline framework to couple together aspect extraction, aspect sentiment classification and opinion extraction. To further explore this task, (Mao et al., 2021; Chen et al., 2021a) transformed ASTE to a machine reading comprehension problem and utilized the shared BERT encoder to obtain the triplets after multiple stages decoding. Another line of research focuses on designing a new tagging scheme that makes the model can extract the triplets in an end-to-end fashion (Xu et al., 2020; Wu et al., 2020a; Zhang et al., 2020; Xu et al., 2021; Yan et al., 2021). For instance, Xu et al. (2020) proposed a position-aware tagging scheme, which solves the limitations related to existing works by enriching the expressiveness of labels. Wu et al. (2020a) proposed a grid tagging scheme, similar to table filling (Miwa and Sasaki, 2014; Gupta et al., 2016), to solve this task in an end-to-end manner. Yan et al. (2021) converted ASTE task into a generative formulation. However, these approaches generally ignore the relations between words and linguistic features which effectively promote the triplet extraction.

### 3 Proposed Framework

In this section, we elaborate on the details of EMC-GCN. The overview of the EMC-GCN framework

#	Relation	Meaning
1	B-A	beginning of aspect term.
2	I-A	inside of aspect term.
3	A	word pair $(w_i, w_j)$ belongs to the same aspect term.
4	B-O	beginning of opinion term.
5	I-O	inside of opinion term.
6	O	word pair $(w_i, w_j)$ belongs to the same opinion term.
7	POS	$w_i$ and $w_j$ of the word pair $(w_i, w_j)$ respectively belong to
8	NEU	an aspect term and an opinion term, and they form aspect-
9	NEG	opinion pair with positive/neutral/negative sentiment.
10	$\perp$	no above relations between word pair $(w_i, w_j)$ .

Table 1: The meanings of our defined ten relations. Note that these relations can also be seen as labels.

is shown in Figure 2.

#### 3.1 Problem Formulation

Given an input sentence  $X = \{w_1, w_2, \dots, w_n\}$  with  $n$  words, the goal of our model is to output a set of triplets  $\mathcal{T} = \{(a, o, s)_m\}_{m=1}^{|\mathcal{T}|}$  from the sentence  $X$ , where  $a$  and  $o$  denote aspect term and opinion term, respectively. The sentiment polarity  $s$  of the given aspect belongs to a sentiment label set  $\mathcal{S} = \{POS, NEU, NEG\}$ . That is, the sentiment label set comprises of three sentiment polarities: positive, neutral and negative. The sentence  $X$  has a total number of  $|\mathcal{T}|$  triplets.

#### 3.2 Relation Definition and Table Filling

We define ten types of relations between words in a sentence for ASTE. These relations are shown in Table 1. Specifically, four relations or labels,  $\{B-A, I-A, B-O, I-O\}$  aim to extract aspect terms and opinion terms. Compared with GTS (Wu et al., 2020a), the relations we defined introduce more accurately boundary information into our model. The

	The	gourmet	food	is	delicious	but	the	service	is	poor
The	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
gourmet	⊥	B-A	A	⊥	POS	⊥	⊥	⊥	⊥	⊥
food	⊥	A	I-A	⊥	POS	⊥	⊥	⊥	⊥	⊥
is	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
delicious	⊥	POS	POS	⊥	B-O	⊥	⊥	⊥	⊥	⊥
but	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
the	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
service	⊥	⊥	⊥	⊥	⊥	⊥	⊥	B-A	⊥	NEG
is	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
poor	⊥	⊥	⊥	⊥	⊥	⊥	⊥	NEG	⊥	B-O

Figure 3: Table filling for triplet extraction in a sentence is illustrated. Each cell denotes a word pair with a relation or label. Refer Table 1 for definitions of relations.

$B$  and  $I$  denote the beginning of and inside of the term respectively, while  $-A$  and  $-O$  subtags aim to determine the role of the term, i.e., an aspect or an opinion. The  $A$  and  $O$  relations in Table 1 are used to detect whether the word pair formed by two different words belongs to the same aspect or opinion term, respectively. The goal of the three sentiment relations  $\{POS, NEU, NEG\}$  is not only to detect whether a word-pair matches or not, but also judge the sentiment polarity of the aspect-opinion pair. Thus, we can construct a relation table for each labelled sentence with table filling method (Miwa and Sasaki, 2014; Gupta et al., 2016). In Figure 3, we show the word pairs and their relations in an example sentence. Here, each cell corresponds to a word pair with a relation.

### 3.3 Triplet Decoding

The decoding details of the ASTE task are shown in Algorithm 1. For simplicity, we use the upper triangular table to decode triplets. Firstly, we use the predicted relations of all word pairs  $(w_i, w_j)$  only based on the main diagonal, to extract aspect terms and opinion terms. Secondly, we need to judge whether the extracted aspect terms and opinion terms match. Particularly, for an aspect term  $a$  and an opinion term  $o$ , we count predicted relations of all word pairs  $(w_i, w_j)$ , where  $w_i \in a$  and  $w_j \in o$ . If there exists any sentiment relation in predicted relations, the aspect term and the opinion term are considered to be paired, otherwise these two are not paired. Finally, for judging the sentiment polarity of the aspect-opinion pair, the most predicted

### Algorithm 1 Triplet Decoding for ASTE

**Input:** The predicted results  $\mathbf{P}$  of a sentence  $X$  with length  $n$ .  $\mathbf{P}(w_i, w_j)$  denotes the predicted label of the word pair  $(w_i, w_j)$ .

**Output:** Triplets  $\mathcal{T}$  of the given sentence.

```

1: Initialize  $\mathcal{D} = \emptyset, \mathcal{A} = \{\}, \mathcal{O} = \{\}, \mathcal{T} = \{\}$ .
2: while  $i \leq n$  do
3:    $\mathcal{D}.append(\mathbf{P}(w_i, w_i)), i \leftarrow i + 1$ 
4: end while
5:  $\mathcal{A} \leftarrow \text{GetAspect}(\mathcal{D}), \mathcal{O} \leftarrow \text{GetOpinion}(\mathcal{D})$ 
6: while  $a \in \mathcal{A}$  and  $o \in \mathcal{O}$  do
7:    $\mathcal{S} = \{\}$ 
8:   while  $w_i \in a$  and  $w_j \in o$  do
9:     if  $i < j$  then  $label = \mathbf{P}(w_i, w_j)$ 
10:    else  $label = \mathbf{P}(w_j, w_i)$ 
11:    if  $label \in \{POS, NEU, NEG\}$ 
12:      then  $\mathcal{S} \leftarrow \mathcal{S} \cup \{label\}$ 
13:    end while
14:   if  $\mathcal{S} \neq \{\}$ 
15:     then The most counted sentiment label denoted as  $s$ ,
16:      $\mathcal{T} \leftarrow \mathcal{T} \cup \{a, o, s\}$ 
17:   end while
18: end while

```

sentiment relation  $s \in \mathcal{S}$  is regarded as sentiment polarity. Thus, we collect a triplet  $(a, o, s)$ .

### 3.4 EMC-GCN Model

#### 3.4.1 Input and Encoding Layer.

BERT (Devlin et al., 2019) has demonstrated its effectiveness in various tasks. We utilize BERT as the sentence encoder to extract hidden contextual representations. Given an input sentence  $X = \{w_1, w_2, \dots, w_n\}$  with  $n$  tokens, the encoding layer outputs the hidden representation sequence  $H = \{h_1, h_2, \dots, h_n\}$  at the last Transformer block.

#### 3.4.2 Biaffine Attention Module

We utilize a biaffine attention module to capture the relation probability distribution of each word pair in a sentence, since the biaffine attention has been proven effective in syntactic dependency parsing (Dozat and Manning, 2017). The biaffine attention process is formulated as,

$$h_i^a = \text{MLP}_a(h_i) \quad (1)$$

$$h_j^o = \text{MLP}_o(h_j) \quad (2)$$

$$g_{i,j} = h_i^{aT} U_1 h_j^o + U_2 (h_i^a \oplus h_j^o) + b \quad (3)$$

$$r_{i,j,k} = \frac{\exp(g_{i,j,k})}{\sum_{l=1}^m \exp(g_{i,j,l})} \quad (4)$$

$$R = \text{Biaffine}(\text{MLP}_a(H), \text{MLP}_o(H)) \quad (5)$$

where multi-layer perceptron is used. The **score vector**  $r_{i,j} \in \mathbb{R}^{1 \times m}$  models relations between  $w_i$  and  $w_j$ ,  $m$  is the number of relation types and  $r_{i,j,k}$  denotes the score of the  $k$ -th relation type for word pair  $(w_i, w_j)$ . The adjacency tensor  $R \in \mathbb{R}^{n \times n \times m}$



models relations between words, and each channel corresponds to a relation type.  $U_1$ ,  $U_2$  and  $b$  are trainable weights and bias.  $\oplus$  denotes concatenation. Eq. (5) collects process of Eqs. (1) to (4).

### 3.4.3 Multi-Channel GCN

Motivated by CNN, GCN is an efficient CNN variant that operates directly on graphs (Kipf and Welling, 2017). A graph contains nodes and edges and GCN applies the convolution operation on those nodes connected directly by edges to aggregate relevant information. Given a sentence with  $n$  words, the general approach is to use the syntactic dependency tree to construct an adjacency matrix  $A \in \mathbb{R}^{n \times n}$  representing a graph for the sentence (Zhang et al., 2019; Sun et al., 2019). The element  $A_{ij}$  denotes the edge of node pair  $(w_i, w_j)$ . Specifically,  $A_{ij} = 1$  if the  $i$ -th node is directly connected to the  $j$ -th node, and  $A_{ij} = 0$  otherwise. A few studies (Guo et al., 2019; Chen et al., 2020a; Li et al., 2021) construct soft edges by attention mechanism for graph. The edge of any node pair  $(w_i, w_j)$  is a probability that indicates the correlation degree between nodes  $w_i$  and  $w_j$ .

To model various relations between words, our EMC-GCN extend the vanilla GCN with a multi-channel adjacency tensor  $R^{ba} \in \mathbb{R}^{n \times n \times m}$  which is constructed by the aforementioned biaffine attention module. Each channel of the adjacency tensor represents the modeling of a relation between words defined in Table 1. Then, we utilize a GCN to aggregate information along each channel for each node. We formulate the process as follows,

$$\tilde{H}_k^{ba} = \sigma \left( R_{:::,k}^{ba} H W_k + b_k \right) \quad (6)$$

$$\hat{H}^{ba} = f(\tilde{H}_1^{ba}, \tilde{H}_2^{ba}, \dots, \tilde{H}_m^{ba}) \quad (7)$$

where  $R_{:::,k}^{ba} \in \mathbb{R}^{n \times n}$  denotes the  $k$ -th channel slice of  $R^{ba}$ .  $W_k$  and  $b_k$  are the learnable weight and bias.  $\sigma$  is an activation function (e.g., ReLU). An average pooling function  $f(\cdot)$  is applied over the node hidden representations of all channels.

### 3.4.4 Linguistic Features

To enhance our EMC-GCN model, we introduce four types of linguistic features for each word pair, shown in Figure 4, including the part-of-speech combination, syntactic dependency type, tree-based distance, and relative position distance. For syntactic dependency type, we add a *self* dependency type for each word pair  $(w_i, w_i)$ . In particular, we randomly initialize four adjacency tensors

Figure 4 illustrates four types of linguistic features for the sentence "The food is delicious". At the top, a dependency tree shows the words and their grammatical roles: "The" (DT), "food" (NN), "is" (VBZ), and "delicious" (JJ). The tree structure is: "The" (DT) --det--> "food" (NN) --nsubj--> "is" (VBZ) --cop--> "delicious" (JJ).

The four tables below represent different feature types:

- Part-of-Speech Combination:** A 4x4 matrix where rows and columns are labeled with the words "The", "food", "is", and "delicious". The cells contain two-letter codes representing the part-of-speech of the pair (e.g., "DT-DT" for "The-The", "NN-DT" for "food-The", etc.).
- Syntactic Dependency Type:** A 4x4 matrix where rows and columns are labeled with the words. The cells contain the syntactic dependency type between the words (e.g., "self", "det", "nsubj", "cop", "-").
- Tree-based Distance:** A 4x4 matrix where rows and columns are labeled with the words. The cells contain the distance between the words in the dependency tree (e.g., 0 for self, 1 for "The-food", 2 for "The-is", etc.).
- Relative Position Distance:** A 4x4 matrix where rows and columns are labeled with the words. The cells contain the relative position distance between the words (e.g., 0 for self, 1 for "The-food", 2 for "The-is", etc.).

Figure 4: Four types of features for a sentence.

based on these features, namely  $R^{psc}$ ,  $R^{dep}$ ,  $R^{tbd}$  and  $R^{rpd}$ . Take syntactic dependency type feature as an example. If a dependency arc exists between  $w_i$  and  $w_j$  and the dependency type is *nsubj*, then  $R_{i,j,:}^{dep}$  is initialized to the embedding of *nsubj* by looking up a trainable embedding table; otherwise we initialize  $R_{i,j,:}^{dep}$  with an  $m$ -dimensional zero vector. Subsequently, the graph convolution operation is repeated using these adjacency tensors to obtain node representations  $\hat{H}^{psc}$ ,  $\hat{H}^{dep}$ ,  $\hat{H}^{tbd}$  and  $\hat{H}^{rpd}$ . Finally, we respectively apply the average pooling function and concatenation operation to all node representations and all edges formally as,

$$\mathbf{H} = f \left( \hat{H}^{ba}, \hat{H}^{psc}, \hat{H}^{dep}, \hat{H}^{tbd}, \hat{H}^{rpd} \right) \quad (8)$$

$$\mathbf{R} = R^{ba} \oplus R^{psc} \oplus R^{dep} \oplus R^{tbd} \oplus R^{rpd} \quad (9)$$

where  $\mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n\}$  and  $\mathbf{R} = \{\mathbf{r}_{1,1}, \mathbf{r}_{1,2}, \dots, \mathbf{r}_{n,n}\}$  denote node representations and edge representations of word pairs.

### 3.4.5 Relation Constraint

In order to precisely capture the relations between words, we impose a constraint on the adjacent tensor obtained from biaffine module, i.e.,

$$\mathcal{L}_{ba} = - \sum_i^n \sum_j^n \sum_{c \in \mathcal{C}} \mathbb{I}(y_{ij} = c) \log(r_{i,j|c}) \quad (10)$$

where  $\mathbb{I}(\cdot)$  denotes the indicator function,  $y_{ij}$  is the ground truth of word-pair  $(w_i, w_j)$ , and  $\mathcal{C}$  denotes

the relation set. Likewise, we impose the relation constraint on four adjacent tensors produced by linguistic features. The constraint costs denote as  $\mathcal{L}_{psc}$ ,  $\mathcal{L}_{dep}$ ,  $\mathcal{L}_{tbd}$  and  $\mathcal{L}_{rpd}$ .

### 3.4.6 Refining Strategy and Prediction Layer

To obtain the representation of word pair  $(w_i, w_j)$  for label prediction, we concatenate their node representations  $\mathbf{h}_i$ ,  $\mathbf{h}_j$  and their edge representation  $\mathbf{r}_{ij}$ . Moreover, motivated by the classifier chains (Read et al., 2011) method in multi-label classification task, we devise an effective refining strategy, which consider the implicit results of aspect and opinion extraction when judging whether word pairs match. Specifically, assuming that  $w_i$  is a word in an aspect term and  $w_j$  is a word in an opinion term, word pair  $(w_i, w_j)$  is more likely to be predicted as an sentiment relation, i.e., *POS*, *NEU* or *NEG*. Otherwise, they are unlikely to match. Thus, we introduce the  $\mathbf{r}_{ii}$  and  $\mathbf{r}_{jj}$  to refine the representation  $s_{ij}$  of word pair  $(w_i, w_j)$ , i.e.,

$$s_{ij} = \mathbf{h}_i \oplus \mathbf{h}_j \oplus \mathbf{r}_{ij} \oplus \mathbf{r}_{ii} \oplus \mathbf{r}_{jj} \quad (11)$$

Finally, we feed the word pair representation  $s_{ij}$  into a linear layer, followed by a softmax function to produce a label probability distribution  $p_{ij}$ , i.e.,

$$p_{ij} = \text{softmax}(W_p s_{ij} + b_p) \quad (12)$$

where  $W_p$  and  $b_p$  are the learnable weight and bias.

## 3.5 Loss Function

Our goal is to minimize the objective function as,

$$\mathcal{L} = \mathcal{L}_p + \alpha \mathcal{L}_{ba} + \beta (\mathcal{L}_{psc} + \mathcal{L}_{dep} + \mathcal{L}_{tbd} + \mathcal{L}_{rpd}) \quad (13)$$

where coefficients  $\alpha$  and  $\beta$  are for adjusting the influence of corresponding relation constraint loss. The standard cross-entropy loss  $\mathcal{L}_p$  is used for the ASTE task, i.e.,

$$\mathcal{L}_p = - \sum_i^n \sum_j^n \sum_{c \in \mathcal{C}} \mathbb{I}(y_{ij} = c) \log(p_{i,j|c}). \quad (14)$$

## 4 Experiments

### 4.1 Datasets

We evaluate our method on two ABSA datasets. Both of them are from the SemEval ABSA Challenges (Pontiki et al., 2014, 2015, 2016). The first dataset  $\mathcal{D}_1$ <sup>3</sup> comes from Wu et al. (2020a). The

<sup>3</sup><https://github.com/NJUNLP/GTS>

Dataset		14res		14lap		15res		16res	
		#S	#T	#S	#T	#S	#T	#S	#T
$\mathcal{D}_1$	train	1,259	2,356	899	1,452	603	1,038	863	1,421
	dev	315	580	225	383	151	239	216	348
	test	493	1,008	332	547	325	493	328	525
$\mathcal{D}_2$	train	1266	2338	906	1460	605	1013	857	1394
	dev	310	577	219	346	148	249	210	339
	test	492	994	328	543	322	485	326	514

Table 2: Statistics for two groups of experiment datasets.

second dataset  $\mathcal{D}_2$ <sup>4</sup> is annotated by Xu et al. (2020), which is a corrected version of dataset proposed by Peng et al. (2020). Statistics for these two groups of datasets are shown in Table 2.

### 4.2 Baselines

We compare our EMC-GCN with state-of-the-art baselines. These models are briefly grouped into three categories. **1) Pipeline methods:** CMLA+, RINANTE+, Li-unified-R, and Peng-two-stage are proposed by Peng et al. (2020). Peng-two-stage+IOG and IMN+IOG are proposed by Wu et al. (2020a). **2) End-to-end methods:** GTS-CNN, GTS-BiLSTM, GTS-BERT (Wu et al., 2020a), OTE-MTL (Zhang et al., 2020), JET-BERT (Xu et al., 2020), S<sup>3</sup>E<sup>2</sup> (Chen et al., 2021b) and BART-ABSA (Yan et al., 2021). **3) MRC-based methods:** BMRC (Chen et al., 2021a) is a multi-turn MRC-based model, which is end-to-end in the training phase, but works in pipeline during the inference phase.

### 4.3 Implementation Details

We use the BERT-base-uncased version<sup>5</sup> as our sentence encoder. AdamW optimizer (Loshchilov and Hutter, 2018) is used with a learning rate of  $2 \times 10^{-5}$  for BERT fine-tuning and a learning rate of  $10^{-3}$  for the other trainable parameters. The dropout rate is set to 0.5. The hidden state dimensionality of BERT and GCN are set to 768 and 300, respectively. The EMC-GCN model is trained in 100 epochs with a batch size of 16. To control the influence of relation constraint, we set the hyperparameter  $\alpha$  and  $\beta$  to 0.1 and 0.01, respectively. Note that the number of channels equals to the number of relations we defined, which is immutable due to the relation constraint we proposed. All sentences are parsed by Stanza (Qi et al., 2020). We save

<sup>4</sup><https://github.com/xuuuluuu/SemEval-Triplets-data/tree/master/ASTE-Data-V2-EMNLP2020>

<sup>5</sup><https://github.com/huggingface/transformers>

Model	14res			14lap			15res			16res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Peng-two-stage+IOG	58.89	60.41	59.64	48.62	45.52	47.02	51.70	46.04	48.71	59.25	58.09	58.67
IMN+IOG	59.57	63.88	61.65	49.21	46.23	47.68	55.24	52.33	53.75	-	-	-
GTS-CNN	70.79	61.71	65.94	55.93	47.52	51.38	60.09	53.57	56.64	62.63	66.98	64.73
GTS-BiLSTM	67.28	61.91	64.49	59.42	45.13	51.30	63.26	50.71	56.29	66.07	65.05	65.56
S <sup>3</sup> E <sup>2</sup>	69.08	64.55	66.74	59.43	46.23	52.01	61.06	56.44	58.66	71.08	63.13	66.87
GTS-BERT	70.92	69.49	70.20	57.52	51.92	54.58	59.29	58.07	58.67	68.58	66.60	67.58
BMRC	-	-	70.01	-	-	57.83	-	-	58.74	-	-	67.49
Our EMC-GCN	71.85	72.12	<b>71.98</b>	61.46	55.56	<b>58.32</b>	59.89	61.05	<b>60.38</b>	65.08	71.66	<b>68.18</b>

Table 3: Experimental results on  $\mathcal{D}_1$  (Wu et al., 2020a). All baseline results are from the original papers.

Model	14res			14lap			15res			16res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
CMLA+ <sup>‡</sup>	39.18	47.13	42.79	30.09	36.92	33.16	34.56	39.84	37.01	41.34	42.10	41.72
RINANTE+ <sup>‡</sup>	31.42	39.38	34.95	21.71	18.66	20.07	29.88	30.06	29.97	25.68	22.30	23.87
Li-unified-R <sup>‡</sup>	41.04	67.35	51.00	40.56	44.28	42.34	44.72	51.39	47.82	37.33	54.51	44.31
Peng-two-stage <sup>‡</sup>	43.24	63.66	51.46	37.38	50.38	42.87	48.07	57.51	52.32	46.96	64.24	54.21
OTE-MTL <sup>†</sup>	62.00	55.97	58.71	49.53	39.22	43.42	56.37	40.94	47.13	62.88	52.10	56.96
JET-BERT <sup>‡</sup>	70.56	55.94	62.40	55.39	47.33	51.04	64.45	51.96	57.53	70.42	58.37	63.83
GTS-BERT <sup>†</sup>	68.09	69.54	68.81	59.40	51.94	55.42	59.28	57.93	58.60	68.32	66.86	67.58
BMRC <sup>†</sup>	75.61	61.77	67.99	70.55	48.98	57.82	68.51	53.40	60.02	71.20	61.08	65.75
BART-ABSA <sup>†</sup>	65.52	64.99	65.25	61.41	56.19	58.69	59.14	59.38	59.26	66.60	68.68	67.62
Our EMC-GCN	71.21	72.39	<b>71.78</b>	61.70	56.26	<b>58.81</b>	61.54	62.47	<b>61.93</b>	65.62	71.30	<b>68.33</b>

Table 4: Experimental results on  $\mathcal{D}_2$  (Xu et al., 2020). The “<sup>‡</sup>” denotes that results are retrieved from Xu et al. (2020). The “<sup>†</sup>” means that we reproduce the models using released code with original parameters on the dataset.

the model parameters according to the best performance of the model on the development set. The reported results are the average on five runs with different random seeds.

#### 4.4 Main Results

The main experimental results are reported in Tables 3 and 4. Under the F1 metric, our EMC-GCN model outperforms all pipeline, end-to-end and MRC-based methods on the two groups of datasets. We observe that end-to-end and MRC-based methods achieve more significant improvements than pipeline methods do, as they establish the correlations between these subtasks and alleviate the problem of error propagation by jointly training multiple subtasks. Note that the tagging schemes of OTE-MTL and GTS-BERT are similar to table filling. Compared with GTS-BERT, our EMC-GCN significantly surpasses its performance by an average of 1.96% and 2.61% F1-score on  $\mathcal{D}_1$  and  $\mathcal{D}_2$ , respectively. This improvement is attributed to that our EMC-GCN can leverage the relations between words and linguistic knowledge for word representation learning. Another finding is that

Model	14res	14lap	15res	16res
EMC-GCN	<b>71.78</b>	<b>58.81</b>	<b>61.93</b>	<b>68.33</b>
w/o Ten Relations	70.68	57.71	59.85	66.48
w/o Linguistic Features	71.22	58.38	60.62	67.15
w/o Relation Constraint	70.59	57.28	59.83	67.89
w/o Refining Strategy	70.62	56.72	60.23	67.31

Table 5: F1 scores of ablation study on  $\mathcal{D}_2$ .

those methods with BERT encoder, such as JET-BERT, GTS-BERT and BMRC, generally achieve better performance than other methods with BiLSTM encoder. We suppose the reason is that BERT has been pre-trained on large-scale data and can provide a strong language understanding ability.

#### 4.5 Model Analysis

##### 4.5.1 Ablation Study

To investigate the effectiveness of different modules in EMC-GCN, we conduct ablation study on the second dataset  $\mathcal{D}_2$ . The experimental results are shown in Table 5. w/o Ten Relations denotes that EMC-GCN uses the same tagging schema as GTS (Wu et al., 2020a) with six labels. Without

Model	14res			14lap		
	POS	NEU	NEG	POS	NEU	NEG
EMC-GCN	74.69	<b>19.65</b>	<b>62.43</b>	<b>67.74</b>	<b>19.14</b>	<b>56.20</b>
w/o Refining Strategy	<b>74.98</b>	17.39	59.87	67.31	16.08	52.74

Table 6: F1 scores of three sentiment relations on  $\mathcal{D}_2$ .

the four relations  $\{B-A, I-A, B-O, I-O\}$ , EMC-GCN loses boundary information of terms, the performance drops significantly. *w/o* Linguistic Features means that we remove the four types of features from EMC-GCN. Without the enhancement of linguistic features, the performance of our EMC-GCN is slightly degraded on 14res and 14lap, but decreased by 1.31% and 1.18% on 15res and 16res, respectively. As 15res and 16res contain less training data, the linguistic features can provide additional information when the training data is insufficient, which is helpful to the prediction of the model. *w/o* Relation Constraint indicates that we remove the relation constraint loss between the adjacency tensor  $R^{ba}$  and the golden label. Thus, each channel in the adjacency tensor cannot precisely describe the relation dependency between words. As a result, the performance of EMC-GCN *w/o* Relation Constraint on four sub datasets is significantly dropped. *w/o* Refining Strategy denotes that we remove the implicit results of aspect and opinion extraction  $\mathbf{r}_{ii}$  and  $\mathbf{r}_{jj}$  from word pair representation  $s_{ij}$ . Since the adjacency tensor has a relation constraint with the golden label, we can suppose  $\mathbf{r}_{ii}$  as a predicted label or relation probability distribution of word pair  $(w_i, w_i)$  on the main diagonal. Thus, we leverage the aspect and opinion extraction implicit results as prior information to help predict the label of word pair  $(w_i, w_j)$ . To sum up, each module of our EMC-GCN contributes to the entire performance on the ASTE task.

#### 4.5.2 Effect of Refining Strategy

The purpose of refining strategy is to facilitate the word pair matching process based on the aspect and opinion extraction implicit results. To verify the idea, we conduct comparative experiments of three sentiment relations  $\{POS, NEU, NEG\}$  on 14rest and 14lap of  $\mathcal{D}_2$ . The results of are shown in Table 6. Note that the function of the three sentiment relations is to detect whether a word-pair matches or not and identify the sentiment polarity of the aspect-opinion pair. The results show that the performance of *w/o* Refining Strategy has declined markedly and the refinement strategy works as we



Figure 5: Visualization of  $POS$  and  $NEG$  relation channels of adjacency tensor  $R^{ba}$  obtained from the biaffine attention.

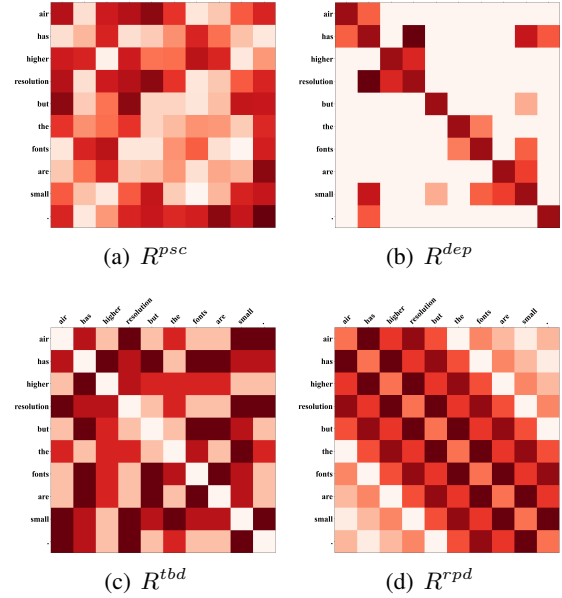


Figure 6: Visualization of adjacency tensors of four linguistic features.

expected.

#### 4.5.3 Channel Visualization

To investigate the effect of relations between words, we visualize the channel slice of adjacency tensor  $R^{ba}$  corresponding to a specific relation. Consider the sample sentence, “air has higher resolution but the fonts are small.” from 14lap dataset. This sentence comprises two triplets,  $\{(resolution, higher, POS), (fonts, small, NEG)\}$ . As shown in the left of Figure 5, the visualized adjacency information of “higher” and “resolution” corresponds to the  $POS$  relation channel. In the visualization, “higher” and “resolution” are highly related to each other. As a result, they convey their own information to each other. Similarly, in the right of Figure 5, “fonts” can receive the node representation and negative sentiment of “small” in the  $NEG$  relation channel. Meanwhile, “small” can also obtain the information of the opinion target it describes. Thus, our EMC-GCN model can readily predict the correct labels of word pairs (“fonts”, “small”) and (“resolution”, “higher”).



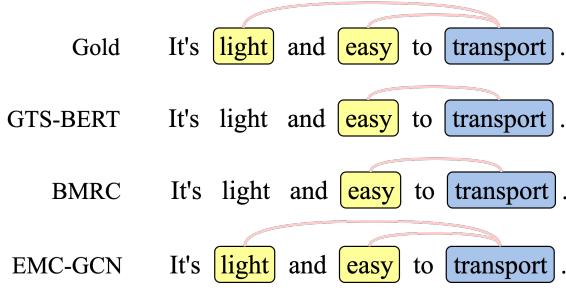


Figure 7: Different models outputs for a given sentence.

#### 4.5.4 Linguistic Feature Visualization

To further analyze the role of linguistic features on ASTE task, we visualize adjacency tensors of four linguistic features. We use the  $l_2$  norm of feature vector in the adjacency tensor to represent the relevance score of the corresponding word pair. In Figure 6, the first one is visualization of adjacency tensor  $R^{psc}$  from part-of-speech combination feature and we observe that the score between adjective and noun is higher, because adjective and noun easily form an aspect-opinion pair, while the score between adjectives is lower, since the two adjectives are usually not related and are likely to be bring noise to each other. In visualization of  $R^{dep}$ , we find that each word only has a score with the words it directly depends on, and computes different relevance scores according to different syntactic dependency types. The visualization of  $R^{tbd}$  shows that the relevance score calculated for each word with other words at different tree-based distances. The visualization of  $R^{rpd}$  demonstrates that the relevance of two adjacent words is greater than that of long-distance word pairs. In summary, all linguistic features we devised contribute to ASTE task.

#### 4.5.5 Case Study

A case study is given in Figure 7. In this example, the aspect terms and opinion terms are highlighted in blue and yellow, respectively. The red line indicates the aspect term and opinion term match, and form a triplet with positive sentiment. The golden opinion term “light” is hard to identify by GTS-BERT and BMRC, while “easy” is predicted correctly by all methods, since “light” is farther from “transport” than “easy”. Thus, they ignore the triplet (“transport”, “light”, positive), while our EMC-GCN can precisely extract it. We argue the key factor is that “light” and “transport” can establish significant connections through sentiment relation and linguistic features.

## 5 Conclusions and Future Work

In this paper, we propose an EMC-GCN architecture for ASTE task. To exploit relations between words, we first devise a multi-channel graph structure for modeling different relation type of each word pair. Then, we utilize graph convolution operation over all channels to learn relation-aware node representations. Furthermore, we consider linguistic features to enhance the GCN-based model. Finally, we design an effective refining strategy on EMC-GCN for better extracting triplets. Extensive experiments on benchmark datasets show that our EMC-GCN model consistently outperforms all baselines. In the future, we will analyse roles of linguistic features and effects of their combinations.

## Acknowledgements

This work was supported in part by the National Key R&D Program of China under Grant 2019YFF0303300 and Subject II under Grant 2019YFF0303302, in part by the National Natural Science Foundation of China under Grants 61906018 and 62076032, in part by the 111 Project under Grant B08004, and in part by the Fundamental Research Funds for the Central Universities under Grant 2021RC36. We appreciate constructive feedback from the anonymous reviewers.

## References

- Chenhua Chen, Zhiyang Teng, and Yue Zhang. 2020a. [Inducing target-specific latent structures for aspect sentiment classification](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5596–5607, Online. Association for Computational Linguistics.
- Shaowei Chen, Jie Liu, Yu Wang, Wenzheng Zhang, and Ziming Chi. 2020b. [Synchronous double-channel recurrent network for aspect-opinion pair extraction](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6515–6524, Online. Association for Computational Linguistics.
- Shaowei Chen, Yu Wang, Jie Liu, and Yuelin Wang. 2021a. [Bidirectional machine reading comprehension for aspect sentiment triplet extraction](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14):12666–12674.
- Zhexue Chen, Hong Huang, Bang Liu, Xuanhua Shi, and Hai Jin. 2021b. [Semantic and syntactic enhanced aspect sentiment triplet extraction](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1474–1483, Online. Association for Computational Linguistics.

- Zhuang Chen and Tieyun Qian. 2020. [Enhancing aspect term extraction with soft prototypes](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2107–2117, Online. Association for Computational Linguistics.
- Hongliang Dai and Yangqiu Song. 2019. [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*, pages 5268–5277, Florence, Italy. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Zi-Yi Dou. 2017. [Capturing user and product information for document level sentiment analysis with deep memory network](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 521–526, Copenhagen, Denmark. Association for Computational Linguistics.
- Timothy Dozat and Christopher D. Manning. 2017. [Deep biaffine attention for neural dependency parsing](#). In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. Open-Review.net.
- Zhifang Fan, Zhen Wu, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. 2019. [Target-oriented opinion words extraction with target-fused neural sequence labeling](#). 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)*, pages 2509–2518, Minneapolis, Minnesota. Association for Computational Linguistics.
- Zhijiang Guo, Yan Zhang, and Wei Lu. 2019. [Attention guided graph convolutional networks for relation extraction](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 241–251, Florence, Italy. Association for Computational Linguistics.
- Pankaj Gupta, Hinrich Schütze, and Bernt Andrassy. 2016. [Table filling multi-task recurrent neural network for joint entity and relation extraction](#). In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2537–2547, Osaka, Japan. The COLING 2016 Organizing Committee.
- Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2019. [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*, pages 504–515, Florence, Italy. Association for Computational Linguistics.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.
- Thomas N. Kipf and Max Welling. 2017. [Semi-supervised classification with graph convolutional networks](#). In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017*.
- Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore. 2011. Twitter sentiment analysis: The good the bad and the omg! In *Fifth International AAAI conference on weblogs and social media*.
- Ruifan Li, Hao Chen, Fangxiang Feng, Zhanyu Ma, Xiaojie Wang, and Eduard Hovy. 2021. [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)*, pages 6319–6329, Online. Association for Computational Linguistics.
- Xin Li, Lidong Bing, Wai Lam, and Bei Shi. 2018. [Transformation networks for target-oriented sentiment classification](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 946–956, Melbourne, Australia. Association for Computational Linguistics.
- Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2019a. [A unified model for opinion target extraction and target sentiment prediction](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):6714–6721.
- Xin Li, Lidong Bing, Wenxuan Zhang, and Wai Lam. 2019b. [Exploiting BERT for end-to-end aspect-based sentiment analysis](#). In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, pages 34–41, Hong Kong, China. Association for Computational Linguistics.
- Ilya Loshchilov and Frank Hutter. 2018. [Fixing weight decay regularization in adam](#).
- Chenyang Lyu, Jennifer Foster, and Yvette Graham. 2020. [Improving document-level sentiment analysis with user and product context](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6724–6729, Barcelona, Spain (Online). International Committee on Computational Linguistics.

- Dehong Ma, Sujian Li, and Houfeng Wang. 2018. [Joint learning for targeted sentiment analysis](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4737–4742, Brussels, Belgium. Association for Computational Linguistics.
- Dehong Ma, Sujian Li, Fangzhao Wu, Xing Xie, and Houfeng Wang. 2019. [Exploring sequence-to-sequence learning in aspect term extraction](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3538–3547, Florence, Italy. Association for Computational Linguistics.
- Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI’17*, page 4068–4074. AAAI Press.
- Yue Mao, Yi Shen, Chao Yu, and Longjun Cai. 2021. [A joint training dual-mrc framework for aspect based sentiment analysis](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(15):13543–13551.
- Makoto Miwa and Yutaka Sasaki. 2014. [Modeling joint entity and relation extraction with table representation](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1858–1869, Doha, Qatar. Association for Computational Linguistics.
- Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. [Knowing what, how and why: A near complete solution for aspect-based sentiment analysis](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):8600–8607.
- Minh Hieu Phan and Philip O. Ogunbona. 2020. [Modelling context and syntactical features for aspect-based sentiment analysis](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3211–3220, Online. Association for Computational Linguistics.
- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. [SemEval-2016 task 5: Aspect based sentiment analysis](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California. Association for Computational Linguistics.
- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. [SemEval-2015 task 12: Aspect based sentiment analysis](#). In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495, Denver, Colorado. Association for Computational Linguistics.
- Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Haris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. [SemEval-2014 task 4: Aspect based sentiment analysis](#). In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.
- Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. [Stanza: A python natural language processing toolkit for many human languages](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online. Association for Computational Linguistics.
- Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. 2011. Classifier chains for multi-label classification. *Machine learning*, 85(3):333–359.
- Aliaksei Severyn and Alessandro Moschitti. 2015. [Twitter sentiment analysis with deep convolutional neural networks](#). In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR ’15*, page 959–962, New York, NY, USA. Association for Computing Machinery.
- Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. 2019. [Aspect-level sentiment analysis via convolution over dependency tree](#). 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)*, pages 5679–5688, Hong Kong, China. Association for Computational Linguistics.
- Duyu Tang, Bing Qin, and Ting Liu. 2016. [Aspect level sentiment classification with deep memory network](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 214–224, Austin, Texas. Association for Computational Linguistics.
- Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020. [Relational graph attention network for aspect-based sentiment analysis](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3229–3238, Online. Association for Computational Linguistics.
- Wenya Wang and Sinno Jialin Pan. 2019. [Transferable interactive memory network for domain adaptation in fine-grained opinion extraction](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):7192–7199.
- Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016. [Recursive neural conditional random fields for aspect-based sentiment analysis](#). In



- Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 616–626, Austin, Texas. Association for Computational Linguistics.
- Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2017. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI’17, page 3316–3322. AAAI Press.
- Zhenkai Wei, Yu Hong, Bowei Zou, Meng Cheng, and Jianmin Yao. 2020. Don’t eclipse your arts due to small discrepancies: Boundary repositioning with a pointer network for aspect extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3678–3684, Online. Association for Computational Linguistics.
- Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020a. Grid tagging scheme for aspect-oriented fine-grained opinion extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2576–2585, Online. Association for Computational Linguistics.
- Zhen Wu, Fei Zhao, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. 2020b. Latent opinions transfer network for target-oriented opinion words extraction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):9298–9305.
- Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu. 2018. Double embeddings and CNN-based sequence labeling for aspect extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 592–598, Melbourne, Australia. Association for Computational Linguistics.
- Lu Xu, Yew Ken Chia, and Lidong Bing. 2021. Learning span-level interactions for aspect sentiment triplet extraction. 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)*, pages 4755–4766, Online. Association for Computational Linguistics.
- Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. Position-aware tagging for aspect sentiment triplet extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2339–2349, Online. Association for Computational Linguistics.
- Hang Yan, Junqi Dai, Tuo Ji, Xipeng Qiu, and Zheng Zhang. 2021. 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)*, pages 2416–2429, Online. Association for Computational Linguistics.
- Bishan Yang and Claire Cardie. 2012. Extracting opinion expressions with semi-Markov conditional random fields. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1335–1345, Jeju Island, Korea. Association for Computational Linguistics.
- Bishan Yang and Claire Cardie. 2013. Joint inference for fine-grained opinion extraction. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1640–1649, Sofia, Bulgaria. Association for Computational Linguistics.
- Bishan Yang and Claire Cardie. 2014. Context-aware learning for sentence-level sentiment analysis with posterior regularization. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 325–335, Baltimore, Maryland. Association for Computational Linguistics.
- Yichun Yin, Furu Wei, Li Dong, Kaimeng Xu, Ming Zhang, and Ming Zhou. 2016. Unsupervised word and dependency path embeddings for aspect term extraction. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI’16, page 2979–2985. AAAI Press.
- Chen Zhang, Qiuchi Li, and Dawei Song. 2019. Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4568–4578, Hong Kong, China. Association for Computational Linguistics.
- Chen Zhang, Qiuchi Li, Dawei Song, and Benyou Wang. 2020. A multi-task learning framework for opinion triplet extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 819–828, Online. Association for Computational Linguistics.