Complementary Learning of Aspect Terms for Aspect-based Sentiment Analysis

Han Qin♠*, **Yuanhe Tian**♥*, **Fei Xia**♥, **Yan Song**♠†

♠The Chinese University of Hong Kong (Shenzhen)   ♥University of Washington

♠hanqin@link.cuhk.edu.cn   ♥{yhtian, fxia}@uw.edu   ♠songyan@cuhk.edu.cn

Abstract

Aspect-based sentiment analysis (ABSA) aims to predict the sentiment polarity towards a given aspect term in a sentence on the fine-grained level, which usually requires a good understanding of contextual information, especially appropriately distinguishing of a given aspect and its contexts, to achieve good performance. However, most existing ABSA models pay limited attention to the modeling of the given aspect terms and thus result in inferior results when a sentence contains multiple aspect terms with contradictory sentiment polarities. In this paper, we propose to improve ABSA by complementary learning of aspect terms, which serves as a supportive auxiliary task to enhance ABSA by explicitly recovering the aspect terms from each input sentence so as to better understand aspects and their contexts. Particularly, a discriminator is also introduced to further improve the learning process by appropriately balancing the impact of aspect recovery to sentiment prediction. Experimental results on five widely used English benchmark datasets for ABSA demonstrate the effectiveness of our approach, where state-of-the-art performance is observed on all datasets.‡

Keywords: aspect-based sentiment analyses, complementary learning

1. Introduction

Aspect-based sentiment analysis (ABSA) is an important task in natural language processing (NLP). It aims to detect the sentiment polarity of a given aspect term in an input sentence. Normally, a good understanding of the information concerning the aspect term, including the boundaries of the aspect term, its context, and the words contained in the aspect term, becomes essential for ABSA. For instance, the example sentence in Figure 1 has two different aspect terms (i.e., "*price*" and "*service*"), where the context word "*reasonable*" shapes "*price*" to be positive while "*poor*" determines "*service*" to be negative. In this case, with the same sentential context, the aspect terms have contradictory sentiment polarities, which raise the bar for ABSA and often cause inferior performance of models without careful treatment to the aspect terms and their contextual information.

To incorporate such information into an ABSA model, most previous studies (Song et al., 2019; Zeng et al., 2019; Phan and Ogunbona, 2020; Yang and Zeng, 2020; Veyseh et al., 2020; Chen et al., 2020a) concatenate aspect term(s) directly to the end of an input sentence with a special token serving as the separator (e.g., "[SEP]" for BERT-based models (Devlin et al., 2019)) and feed the resulted sentence+aspect pair into an encoder. This simple and straightforward approach has been proved to be rather effective; however, it only emphasizes the aspect term to the ABSA model without further understanding of the boundary and context of a given aspect term. Therefore, an appropriate approach
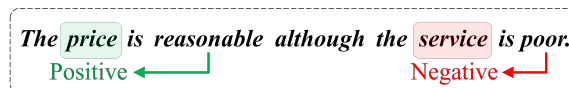

Figure 1: An example sentence contains two aspect terms with contradictory sentiment polarities (i.e., positive for "*price*" and negative for "*service*").

to model the given aspect term(s) in an input sentence is expected to enhance the ABSA model with a better aspect-aware context comprehension.

In this paper, we propose an approach to enhance ABSA through complementary learning of aspects, which is a supportive auxiliary task of recovering aspect terms for improving ABSA. Specifically, our approach has two training stages. In the first training stage, in addition to the main sentiment classifier, our approach models the aspect term and its context through an extra decoder to re-construct the input sentence, especially the aspects, and optimize the entire model accordingly. Moreover, a discriminator is added to control the impact of the complementary learning to the main sentiment classifier in a discriminative manner and ensure our model focus on ABSA. In the second training stage, the decoder and the discriminator are removed, and only the sentiment classifier is trained following the standard procedure of training a supervised ABSA model. Therefore, through the first training stage, the model is able to explicitly learn the aspect terms as well as their context, which allows the model to take the original sentence as the input. Then, the second training stage ensures the learning target (i.e., ABSA) being emphasized. Compared with previous studies (Liang et al., 2019) that apply multi-task learning to aspect term recognition and ABSA though one single training stage, our approach not only enhances

---

*Equal contribution.
†Corresponding author.
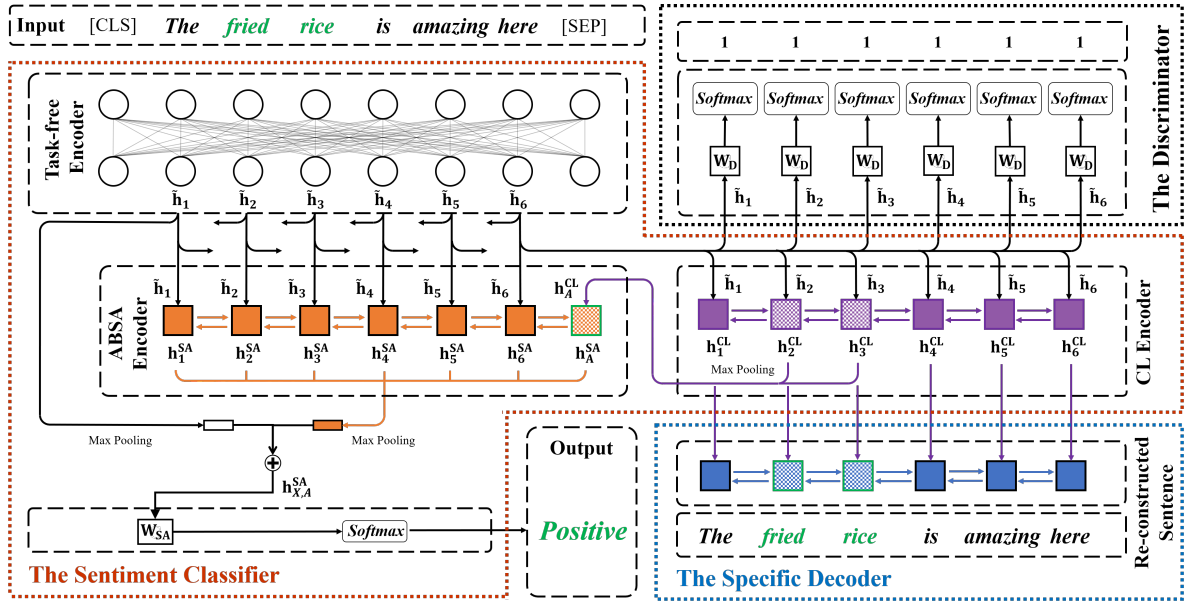‡Our code is available at https://github.com/synlp/ASA-CLD.

Figure 2: The overall architecture of the proposed model with complementary learning of aspects for ABSA with an example input sentence and an aspect term (i.e., "*fried rice*" highlighted in green). The sentiment classifier, the specific decoder, and the discriminator are illustrated in red, blue, and black dash-line boxes, respectively, all of which are trained at the first training stage and the sentiment classifier is trained alone at the second stage.

the multi-task learning process by a discriminator to control the impact of the complementary learning, but also uses a two-stage-training design to force the model to focus mainly on the target ABSA task itself. Experimental results on five English benchmark datasets, i.e., LAP14, REST14 (Pontiki et al., 2014), REST15 (Pontiki et al., 2015), REST16 (Pontiki et al., 2016), and MAMS (Jiang et al., 2019), demonstrate the effectiveness of our approach to ABSA, where state-of-the-art performance is observed on all datasets.

2. The Approach

The architecture of our approach is illustrated in Figure 2, where the main sentiment classifier for ABSA is highlighted in the red dash-line box and the specific decoder for complementary learning (CL) is illustrated in the blue dash-line box at the bottom right part, respectively, with the task-free encoder (TFE) for both tasks shown at the top left part. A discriminator, which takes the output of the task-free encoder and determines whether the specific decoder is able to correctly recover the words in the input sentence, is illustrated on the bottom right part. Overall, our approach has **two training stages**. In the **first training stage**, we train the entire model with the sentiment classifier, the specific decoder, and the discriminator together, which can be formalized by learning the process of

$$\widehat{y}^{\text{SA}}, \widehat{\mathbf{y}}^{\text{CL}} = \text{ABSA-CL}(\mathcal{X}, \mathcal{A}) \qquad (1)$$

with ABSA-CL referring to the joint function of the classifier, decoder, and discriminator; $\mathcal{X} = x_1, x_2 \cdots x_n$ is the sequence of the input sentence with n words; $\mathcal{A} = a_1, a_2 \cdots a_m$ is the given aspect term with m words (which is a sub-sequence of $\mathcal{X}$); $\widehat{y}^{\text{SA}}$ is

the predicted sentiment polarity towards the given aspect term $\mathcal{A}$; $\widehat{\mathbf{y}}^{\text{CL}}$ is the re-constructed sequence for the input sentence. In the **second training stage**, we further train the sentiment classifier alone following the standard supervised ABSA training, without involving other components. This process can be formalized as learning

$$\widehat{y}^{\text{SA}} = \text{ABSA}(\mathcal{X}, \mathcal{A}) \qquad (2)$$

with ABSA referring to the classification of sentiment polarities. In doing so, the sentiment classifier is further enhanced by focusing on sentiment polarity prediction after the first training stage. The following texts elaborate CL for ABSA and then describe the details of the discriminator.

2.1. ABSA with Complementary Learning

In general, a good understanding of the running text is highly important for NLP tasks (Pennington et al., 2014; Song et al., 2018; Song and Shi, 2018; Baldini Soares et al., 2019; Zhang et al., 2019b; Xu et al., 2019; Diao et al., 2020; Li et al., 2020; Helwe et al., 2020; Diao et al., 2021; Song et al., 2021). For ABSA, in particular, the modeling of the given aspect term and its surrounding context is the key for a model to achieve promising performance. In achieving this understanding goal, motivated by previous studies to learn extra knowledge and features through multi-task learning (Liu et al., 2017; Liu et al., 2019; Qin et al., 2021a; Qin et al., 2022), we propose CL for ABSA to re-construct the input sentence, especially the words in the aspect term appearing in the sentence. Through this process, the model is able to better analyze the input sentence so as to facilitate the sentiment prediction with aspect-aware contextual information. Specifically,

our approach follows the encoding-decoding paradigm, where the TFE is shared by both the main sentiment classifier and the CL task, and encodes the input sentence $\mathcal{X}$ by

$$\widetilde{\mathbf{h}}_1, \widetilde{\mathbf{h}}_2 \cdots \widetilde{\mathbf{h}}_n = \text{TFE}(\mathcal{X}) \tag{3}$$

with $\widetilde{\mathbf{h}}_i$ ($1 \leq i \leq n$) denoting the task-free hidden vector for x_i. Then, we apply max pooling onto these hidden vectors to obtain a task-free sentence representation (i.e. $\widetilde{\mathbf{h}}_{\mathcal{X}}$) by

$$\widetilde{\mathbf{h}}_{\mathcal{X}} = \text{MaxPooling}(\{\widetilde{\mathbf{h}}_1, \widetilde{\mathbf{h}}_2 \cdots \widetilde{\mathbf{h}}_n\}) \tag{4}$$

Such word- and sentence-level representations are further leveraged by all components including the main sentiment classifier, the specific decoder for CL, and the discriminator.

The Sentiment Classifier for ABSA ABSA is generally formalized as a classification task with a given input sentence $\mathcal{X}$ and a specified aspect term $\mathcal{A}$.[1] In our approach, we firstly feed the task-free hidden vectors (i.e., $\widetilde{\mathbf{h}}_1, \widetilde{\mathbf{h}}_2 \cdots \widetilde{\mathbf{h}}_n$) obtained from the TFE and the representation[2] of the aspect term (denoted by $\mathbf{h}_{\mathcal{A}}^{\text{CL}}$) learned from the CL task together to the ABSA encoder (denoted as SAE), to compute the ABSA-specific hidden vectors by

$$\mathbf{h}_1^{\text{SA}} \cdots \mathbf{h}_n^{\text{SA}}, \mathbf{h}_{\mathcal{A}}^{\text{SA}} = \text{SAE}(\widetilde{\mathbf{h}}_1, \cdots, \widetilde{\mathbf{h}}_n, \mathbf{h}_{\mathcal{A}}^{\text{CL}}) \tag{5}$$

where $\mathbf{h}_i^{\text{SA}}$ ($i \in [1, n]$) is the ABSA-specific hidden vector for x_i and $\mathbf{h}_{\mathcal{A}}^{\text{SA}}$ is the hidden vector for the aspect term $\mathcal{A}$. Next, we apply max pooling onto these hidden vectors to extract ABSA-specific sentence representation with the complementary aspect term information (i.e., $\mathbf{h}_{\mathcal{X},\mathcal{A}}^{\text{SA}}$), through

$$\mathbf{h}_{\mathcal{X},\mathcal{A}}^{\text{SA}} = \text{MaxPooling}(\{\mathbf{h}_1^{\text{SA}}, \cdots, \mathbf{h}_n^{\text{SA}}, \mathbf{h}_{\mathcal{A}}^{\text{SA}}\}) \tag{6}$$

Then, we concatenate ($\oplus$) $\mathbf{h}_{\mathcal{X},\mathcal{A}}^{\text{SA}}$ with $\widetilde{\mathbf{h}}_{\mathcal{X}}$ to compute $\mathbf{h}^{\text{SA}}$ for final prediction by

$$\mathbf{h}^{\text{SA}} = \widetilde{\mathbf{h}}_{\mathcal{X}} \oplus \mathbf{h}_{\mathcal{X},\mathcal{A}}^{\text{SA}} \tag{7}$$

and

$$\mathbf{o}^{\text{SA}} = \text{softmax}\left(\mathbf{W}_{\text{SA}} \cdot \mathbf{h}^{\text{SA}} + \mathbf{b}_{\text{SA}}\right) \tag{8}$$

where $\mathbf{W}_{\text{SA}}$ and $\mathbf{b}_{\text{SA}}$ are the trainable matrix and bias vector, respectively; each dimension of $\mathbf{o}^{\text{SA}}$ represents the predicted probability of a particular sentiment polarity y^{SA} given $\mathcal{X}$ and $\mathcal{A}$. Normally, the loss for the sentiment classifier (i.e., $\mathcal{J}_{\text{SA}}$) is computed with negative log likelihood function

$$\mathcal{J}_{\text{SA}} = -\log p(y^{\text{SA}^*}|\mathcal{X}, \mathcal{A}) \tag{9}$$

where $p(y^{\text{SA}^*}|\mathcal{X}, \mathcal{A})$ denotes the predicted probability of the ground truth sentiment polarity y^{SA^*} for a given aspect term.

---

[1]If a sentence has multiple aspect terms, it is paired with each aspect term at a time to form separate instances.

[2]We use the hidden vector $\mathbf{h}_{\mathcal{A}}^{\text{CL}}$ to represent the entire aspect term no matter how many words it contains.

The Complementary Learning Task CL in our approach follows the standard sequence-to-sequence paradigm, whose object is to re-construct the input sequence. In CL, we firstly collect task-free hidden vectors $\widetilde{\mathbf{h}}_1 \cdots \widetilde{\mathbf{h}}_n$ obtained from the TFE, and then feed them into the CL encoder to obtain the CL-specific hidden vector, i.e., $\mathbf{h}_i^{\text{CL}}$, for each x_i. Next, $\mathbf{h}_i^{\text{CL}}$ are further fed into the specific decoder f_d to re-construct the input sentence through

$$\widehat{y}_t^{\text{CL}} = f_d(\mathbf{h}_1^{\text{CL}} \cdots \mathbf{h}_{t-1}^{\text{CL}}, x_1 \cdots x_{t-1}) \tag{10}$$

where x_{t-1} represents the $(t\text{-}1)$-th word in the re-constructed sentence and $\widehat{y}_t^{\text{CL}}$ the t-th word generated by the decoder. Finally, we apply the negative log likelihood loss function to the output sequence and compute the loss for CL ($\mathcal{J}_{\text{CL}}$) by

$$\mathcal{J}_{\text{CL}} = -\sum_{t=1}^{n} \log p(x_t|x_{t-1} \cdots x_1) \tag{11}$$

where the term $p(x_t|x_{t-1} \cdots x_1)$ denotes the predicted probability of the ground truth word, namely, x_t. In such process, the CL encoder is able to learn more detailed aspect-aware information. Therefore, we extract the output hidden vectors of the aspect term $\mathcal{A}$ by

$$\mathbf{h}_{\mathcal{A}}^{\text{CL}} = \text{MaxPooling}(\{\mathbf{h}_i^{\text{CL}}|x_i \in \mathcal{A}\}) \tag{12}$$

which is further used in the sentiment classifier (i.e., Eq. (5)) to guide the model to predict the sentiment polarity towards to the given aspect term $\mathcal{A}$ in the input sentence.

2.2. The Discriminator

Although the aforementioned CL task offers the main sentiment classifier with explicit aspect-aware contextual information, it cannot automatically adjust its appropriate contribution to ABSA, which may potentially lead the main sentiment classifier to being under- and over-fit by such information. To tackle this problem, we propose to add a discriminator to the TFE to control the impact of CL on the main sentiment classifier. The discriminator is designed to take the output of the TFE (i.e., $\widetilde{\mathbf{h}}_1, \widetilde{\mathbf{h}}_2 \cdots \widetilde{\mathbf{h}}_n$) as its input and predict whether the specific decoder for CL is able to successfully recover particular words in the input sentence. In doing so, for each word x_i, the discriminator performs a binary classification, where the prediction is denoted by $\widehat{y}_i^{\text{D}}$ and the ground truth (denoted by $y_i^{\text{D}^*}$) of x_i is defined by

$$y_i^{\text{D}^*} = \begin{cases} 0 & \widehat{y}_i^{\text{CL}} \neq x_i \\ 1 & \widehat{y}_i^{\text{CL}} = x_i \end{cases} \tag{13}$$

with $\widehat{y}_i^{\text{CL}}$ referring to the i-th recovered word in the input sentence. Therefore, for each word x_i in $\mathcal{X}$, the discriminator maps $\widetilde{\mathbf{h}}_i$ to a two dimensional vector $\mathbf{o}_i^D$ through

$$\mathbf{o}_i^{\text{D}} = \text{softmax}\left(\mathbf{W}_{\text{D}} \cdot \widetilde{\mathbf{h}}_i + \mathbf{b}_{\text{D}}\right) \tag{14}$$

Dataset		Pos. #	Neu. #	Neg. #
LAP14	Train	994	464	870
	Test	341	169	128
REST14	Train	2,164	637	807
	Test	728	196	182
REST15	Train	907	36	254
	Test	326	34	207
REST16	Train	1,229	69	437
	Test	469	30	114
MAMS	Train	3,380	5,042	2,764
	Dev	403	604	325
	Test	400	607	329

Table 1: The number of aspect terms with "*positive*" (Pos.), "*neutral*" (Neu.), and "*negative*" (Neg.) sentiment polarities in the train/test sets of all five datasets.

where $\mathbf{W}_D$ and $\mathbf{b}_D$ denote the trainable matrix and bias vector, respectively. Herein, the values at the first and second dimension of $\mathbf{o}_i^D$ are the probabilities of classifying x_i to be class 0 and 1 (defined by Eq. (13)), respectively. Afterwards, we compute the loss of the discriminator (i.e., $\mathcal{J}_D$) by

$$\mathcal{J}_D = -\sum_{i=1}^{n} \log p(y_i^{D^*}|\mathcal{X}) \qquad (15)$$

Finally, we use this loss to control the effect of CL by $\mathcal{J}_D \times \mathcal{J}_{CL}$. Therefore, the object of our approach with CL to minimize the total loss $\mathcal{J}$ is defined by

$$\mathcal{J} = \mathcal{J}_{SA} + \mathcal{J}_D \times \mathcal{J}_{CL} \qquad (16)$$

where $\mathcal{J}_{SA}$, $\mathcal{J}_{CL}$ and $\mathcal{J}_D$ are losses from sentiment classifier, specific decoder for CL, and discriminator, respectively.

Through this process, the effect of CL is dynamically controlled by the discriminator, which is further explained as follows. On the one hand, when the specific decoder for CL successfully recovers x_i and the discriminator predicts that the decoder is able to do so (i.e., $y_i^{D^*} = 1$ and $\hat{y}_i^D = 1$) (which means the TFE in the main sentiment classifier has already have a good modeling to the words and their contexts), both $\mathcal{J}_D$ and $\mathcal{J}_{CL}$ are relatively small. Therefore, the loss from the specific decoder for CL should be reduced and alleviate the influence of CL to the main sentiment classifier. On the other hand, when the specific decoder for CL makes incorrect predictions and the discriminator predicts that the decoder cannot recover the corresponding words (i.e., $y_i^{D^*} = 0$ and $\hat{y}_i^D = 0$) (which means the decoder cannot identify an aspect term and its context correctly), $\mathcal{J}_D$ is relatively small even though $\mathcal{J}_{CL}$ is rather large. As a result, the discriminator controls and adjusts the effect of the CL task on the main sentiment classifier and prevents it from being dominated by the CL task. In the rest cases where $\hat{y}_i^D \neq y_i^{D^*}$ (which

Hyper-parameters	Values
Learning Rate	$5e-6, \mathbf{1e-5}, 3e-5, 5e-5$
Warmup Rate	$\mathbf{0.06}, 0.1$
Dropout Rate	$\mathbf{0.1}$
Batch Size	$4, \mathbf{8}$

Table 2: The hyper-parameters used in tuning our models and the best one used in our final experiments are highlighted in boldface.

means it is difficult to model the aspect term and its context correctly), $\mathcal{J}_D$ is relatively large. Therefore, regardless of how large $\mathcal{J}_{CL}$ is, the discriminator further enhances $\mathcal{J}_{CL}$ and thus forces the entire model (including the main sentiment classifier) to learn more aspect-aware information from CL.

3. Experimental Settings

3.1. Datasets

We use five English benchmark datasets for ABSA, i.e., LAP14 and REST14 (Pontiki et al., 2014), REST15 (Pontiki et al., 2015), REST16 (Pontiki et al., 2016), and MAMS[3] (Jiang et al., 2019). Particularly, LAP14 contains laptop computer reviews, REST14, REST15, REST16, and MAMS is collected from online reviews of restaurants. For LAP14, REST14, and REST16, we follow previous studies (Chen et al., 2017; He et al., 2018a; Chen et al., 2020a) to remove the aspect terms with "*conflict*" sentiment polarity[4] and the sentences without an aspect term. For all datasets, we use their official train/dev/test splits[5] and report the statistics (i.e., the numbers of aspect terms with "*positive*", "*negative*", and "*neutral*" sentiment polarities) of the five datasets in Table 1.

3.2. Implementation

For the TFE in our approach, we use BERT-large[6] (Devlin et al., 2019), which achieves state-of-the-art performance in many NLP tasks (Ohashi et al., 2020; Tian et al., 2020b; Tabassum et al., 2020; Nie et al., 2020; Mass et al., 2020; Tian et al., 2020a; Herzig and Berant, 2021; Qin et al., 2021c; Barnes et al., 2021) with the default setting (i.e., 24 layers of multi-head self-attention with 1024-dimensional hidden vectors). For the ABSA-specific encoder, as well as the encoder and decoder for CL, we try two popular architectures, namely, BiLSTM and Transformer, with randomly initialized parameters. Following previous studies (Tang et al., 2016a; Chen et al., 2017; Zhang et al.,

---

[3]We use the ATSA part of MAMS obtained from `https://github.com/siat-nlp/MAMS-for-ABSA`.

[4]"*Conflict*" is a sentiment polarity used to identify the aspect terms that have contradictory sentiment polarities in the same sentence in LAP14, REST14/16.

[5]It is worth noting that LAP14, REST14, REST115, and REST16 do not have their official development sets.

[6]We obtain the BERT models from `https://github.com/huggingface/pytorch-pretrained-BERT`.

	LAP14		REST14		REST15		REST16		MAMS	
	ACC	F1	ACC	F1	ACC	F1	ACC	F1	ACC	F1
BiLSTM	78.99	74.47	83.57	74.40	79.18	68.32	75.38	72.89	80.21	79.27
+ CL	**82.92**	76.15	85.14	80.90	**86.39**	72.32	91.33	78.17	83.90	83.44
+ CL-D	81.67	**78.14**	**87.14**	**81.02**	86.33	**72.38**	**93.14**	**80.33**	**84.13**	**83.49**
Transformer	78.34	75.32	85.26	77.26	84.10	69.01	90.58	73.42	81.55	81.94
+ CL	77.59	77.39	86.25	79.94	81.72	69.04	93.14	81.80	82.71	82.28
+ CL-D	**83.23**	**80.42**	**88.30**	**83.07**	**87.31**	**74.48**	**93.30**	**81.96**	**83.98**	**83.54**

Table 3: Experimental results (accuracy and F1 scores) of baselines and our approaches with different settings on five benchmark datasets. "+ CL-D" refers to our approach with both complementary learning and the discriminator.

2019a; Tang et al., 2020; Chen et al., 2020a; Tian et al., 2021b), we evaluate all models via accuracy and macro-averaged F1 scores over all sentiment polarities. It is worth noting that for all datasets without official development sets, we randomly pick 10% of the training set serving as the development set so as to find the best hyper-parameters (which are illustrated in Table 2), which are then applied to our models when learning on the entire training set.

4. Results and Analyses

4.1. Overall Results

In the experiments, we run baselines and our models with BERT-large TFE, CL, and the discriminator, where either BiLSTM or Transformer is used as the ABSA-specific encoder and the CL decoder. We report the experimental results on the test set of all datasets in Table 3. There are two observations. First, although the BERT-large baselines have already achieved outstanding performance, our approach with CL (i.e., "+CL") can still consistently outperform the baselines on all datasets. This observation confirms that, compared with the baselines that use the combined sentence and aspect term as the input, our approach to learning aspect information through CL is able to better understand aspect terms and their contexts thus achieve higher ABSA performance. Second, further improvement is observed when our model is enhanced by the discriminator (i.e., "+CL-D"), which confirms that the discriminator is able to appropriately balance the impact of sentence recovery and ABSA so that prevent the main sentiment classifier from being dominated by the auxiliary CL task.

4.2. Comparison with Previous Studies

To further demonstrate the effectiveness of our approach, we compare our best performing model (i.e., BERT-large TFE with Transformer ABSA encoder, CL, and the discriminator) with previous studies on all datasets. The results are reported in Table 4. Overall, it is observed that our approach outperforms previous studies on all datasets with respect to F1 scores and achieves state-of-the-art performance on four of them. Particularly, it is promising that our approach that does

not rely on extra input outperforms previous studies (marked by "⋆") (He et al., 2018a; Sun et al., 2019; Zhang et al., 2019a; Huang and Carley, 2019; Wang et al., 2020; Tang et al., 2020; Zhang and Qian, 2020; Meng et al., 2020) that leverage dependency information to model the the aspect term with its context, which further confirms the effectiveness of the proposed CL as an informative context understanding process.

4.3. Effect of Complementary Learning

In general, a good comprehension of the given aspect term and its surrounding context is highly important for ABSA. However, it is sometimes challenging for a model to appropriately understand the aspect term and its context, especially in cases where a sentence has multiple aspects with different sentiment polarities, where the same sentential context shared by different aspect terms may introduce noise when the model tries to predict the sentiment polarity towards a particular aspect term. To explore whether our approach is able to successfully deal with such challenge, we extract sentences having at least two aspect terms from the test sets of all datasets. We then evaluate baselines and our best performing model (i.e., BERT-large TFE with Transformer ABSA encoder, CL, and the discriminator) on the extracted subsets and report the accuracy and F1 scores of different models in Table 5. It is observed that our model outperforms all baselines, which further confirms the effectiveness of our model in modeling the aspect terms and their context.

4.4. Effect of the Discriminator

Since the main sentiment classifier and the CL task share the same TFE, it is possible that the main sentiment classifier is overwhelmed by the auxiliary CL task. Therefore, we investigate how the discriminator controls the balance of CL to the classifier. To explore whether the discriminator functionalizes as expected, we extract the intermediate models (checkpoints) obtained after the first training stage from our best performing models with and without the discriminator (i.e., Transformer+CL-D and Transformer+CL) and evaluate such intermediate models on ABSA and CL tasks. We report the results in Table 6, where ac-

Models	LAP14 ACC	LAP14 F1	REST14 ACC	REST14 F1	REST15 ACC	REST15 F1	REST16 ACC	REST16 F1	MAMS ACC	MAMS F1
⋆Chen et al. (2017)	74.49	71.35	80.23	70.80	-	-	-	-	-	-
Ma et al. (2017)	72.10	-	78.60	-	-	-	-	-	-	-
Fan et al. (2018)	75.39	72.47	81.25	71.94	-	-	-	-	-	-
Gu et al. (2018)	74.12	-	81.16	-	-	-	-	-	-	-
⋆He et al. (2018a)	72.57	69.13	80.63	71.32	81.67	66.05	64.61	67.45	-	-
He et al. (2018b)	71.15	67.46	79.11	69.73	81.30	68.74	85.58	69.76	-	-
Huang and Carley (2018)	70.06	-	79.20	-	-	-	-	-	-	-
Li et al. (2018)	76.54	71.75	80.69	71.27	-	-	-	-	-	-
Chen and Qian (2019)	73.87	71.10	79.55	71.41	-	-	-	-	-	-
Du et al. (2019)	76.80	73.29	81.79	73.40	-	-	-	-	-	-
Hu et al. (2019)	-	-	84.28	74.45	78.58	54.72	-	-	-	-
Mao et al. (2019)	75.84	72.49	82.49	72.10	-	-	-	-	-	-
Song et al. (2019)	79.93	76.31	83.12	73.76	-	-	-	-	-	-
Xu et al. (2019)	78.07	75.08	84.95	76.96	-	-	-	-	-	-
Jiang et al. (2019)	-	-	85.93	-	-	-	-	-	83.39	-
⋆Sun et al. (2019)	77.19	72.99	82.30	74.02	-	-	85.58	69.93	-	-
⋆Zhang et al. (2019a)	75.55	71.05	81.22	72.94	79.89	61.89	88.99	67.48	-	-
⋆Huang and Carley (2019)	80.10	-	83.00	-	-	-	-	-	-	-
Liang et al. (2019)	75.86	-	82.95	-	-	-	-	-	-	-
⋆Wang et al. (2020)	78.21	74.07	86.60	81.35	-	-	-	-	-	-
⋆Tang et al. (2020)	79.8	75.6	86.3	80.0	84.0	71.0	91.9	79.0	-	-
⋆Meng et al. (2020)	80.96	76.95	86.71	79.12	85.39	66.26	91.35	75.19	-	-
⋆Zhang and Qian (2020)	74.59	71.84	81.97	73.48	81.16	64.79	88.96	70.84	-	-
Xu et al. (2020)	82.86	73.78	77.64	74.23	80.82	61.59	89.51	75.92	-	-
Phan and Ogunbona (2020)	80.52	77.13	86.71	80.31	-	-	-	-	-	-
⋆Veyseh et al. (2020)	82.8	80.2	87.2	82.5	-	-	-	-	-	-
⋆Chen et al. (2020a)	81.98	78.81	86.43	80.30	86.35	70.76	92.53	79.24	-	-
Yan et al. (2021)	-	76.76	-	75.56	-	73.91	-	-	-	-
Our Best Model	**83.23**	**80.42**	**88.30**	**83.07**	**87.31**	**74.48**	**93.30**	**81.96**	**83.98**	**83.54**

Table 4: The comparison of our best model (i.e., BERT-large TFE with Transformer + CL-D) with previous studies on all datasets. "⋆" mark the models that leverages extra dependency information;

Datasets	TF Acc	TF F1	TF+CL-D Acc	TF+CL-D F1
LAP14	76.48	72.39	**82.34**	**79.62**
REST14	84.32	75.15	**87.77**	**81.20**
REST15	85.44	68.18	**86.58**	**70.21**
REST16	90.32	78.87	**94.55**	**79.78**
MAMS	77.27	76.38	**83.86**	**83.40**

Table 5: Experimental results of the baseline (i.e., Transformer (TF)) and our approach (i.e., Transformer + CL-D) on the subsets of test sets of five datasets, where the subsets consist of all test sentences with multiple aspect terms.

curacy[7] and BLEU-2 scores[8] are used to evaluate the model performance on the CL task. It is noted that among different datasets, the intermediate model with a discriminator works better on ABSA while performs slightly worse on the CL task, compared to the one without a discriminator. Given that our approach with the discriminator outperform the one without after the second training stage on ABSA, this observation confirms that the discriminator is able to control the effect of the CL task and force the model to focus more on the main sentiment classifier, which prevents the model from being introduced with unnecessary knowledge for ABSA.

4.5. Case Study

To further explore the effect of our approach, we conduct a case study on two example sentences, which are illustrated in Figure 3. The first sentence (i.e., Figure 3(a)) has four aspect terms (i.e., ("*drinks*", "*brunch*", "*spot*", and "*waiting*") and the second sentence (i.e., Figure 3(b)) has three aspect terms (i.e., "*decor*", "*bar*", and "*atmosphere*"), where the positive, negative, and neutral gold standard sentiment polarities toward these aspect terms are represented by green, red, and grey background colors in the sentence, respectively. The predictions of the baseline (i.e., Transformer), our approach with CL (i.e., Transformer+CL), and our full model (i.e., Transformer+CL-D) for all as-

[7]Accuracy is measured by the percentage of words successfully recovered by the specific decoder (i.e., $\widehat{y}_i^{CL} = x_i$).

[8]We use `torchtext.data.metrics.bleu_score` in the `torchtext` framework to compute the BLEU scores.

Figure 3: A case study of different models on two example sentences with multiple aspect terms that are highlighted in different colored boxes. The gold standard and the predicted "*positive*", "*negative*", and "*neutral*" sentiment polarities for different aspect terms are assigned with green, red, and grey background colors, respectively.

Datasets	D.	ABSA		CL	
		Acc	F1	Acc	BLEU-2
LAP14	×	76.49	72.16	**38.65**	**62.28**
	√	**81.19**	**78.50**	36.09	47.91
REST14	×	83.21	75.61	**42.18**	**88.61**
	√	**85.01**	**77.47**	35.15	74.81
REST15	×	**82.84**	66.17	**40.38**	**78.95**
	√	**82.84**	**68.31**	38.78	74.62
REST16	×	89.54	74.44	**51.59**	**74.19**
	√	**90.69**	**75.35**	49.71	67.07
MAMS	×	82.18	81.59	**58.49**	**94.17**
	√	**83.08**	**83.26**	47.71	81.91

Table 6: Experimental results on ABSA and the CL task of the intermediate models obtained after the first training stage. "√" and "×" refer to the models with and without the discriminator (denoted by "D.").

pect terms are also illustrated following the same background color scheme. It is observed that our full model (i.e., Transformer+CL-D) correctly predicts all sentiment polarities whereas other models fail to do so. The analyses are as follows.

In the first example, the sentiment of the entire sentence tends to be positive. Therefore, it is possible that the baseline is misled by the sentence-level sentiment polarity when it predicts the sentiment towards "*spot*" and "*waiting*", because it is limited for the baseline model to appropriately distinguish the aspect terms and their contexts from the sentential context through concatenating the sentence and the aspect term. In contrast, Transformer+CL is able to do so through CL and thus identifies the difference between the sentence level sentiment polarity and that towards "*waiting*". Yet, Transformer+CL still fails to correctly predict the sentiment polarity towards "*spot*", which might result from the domination of CL on ABSA. When the discriminator is added to CL, the impact of CL is dynamically controlled, which adjusts the contribution of CL to the main sentiment classifier and hence allows Transformer+CL-D to predict the correct sentiment polarities towards all aspect terms.

Similarly, the sentential sentiment of the second sentence tend to be negative, which misled the baseline

model to make incorrect predictions. On the contrary, our approach (i.e., Transformer + CL-D) can better model the aspect terms and their context and thus is able to distinguish the sentiment polarity towards a particular aspect term from the sentential one, and lead to correct predictions.

5.   Related Work

Contextual information in the input sentence, especially the aspect term and its contexts, is highly important to ABSA. Many previous studies tried advanced encoders (e.g., BiLSTM and Transformer) to better model contextual information for ABSA (Tang et al., 2016a; Tang et al., 2016b; Wang et al., 2016; Ma et al., 2017; Fan et al., 2018; Li et al., 2018; Mao et al., 2019; Xu et al., 2019; Jiang et al., 2019; Xu et al., 2020; Qin et al., 2021b). Among these studies, most ones utilized extra syntactic information, especially the dependency parses, of the input sentence to further capture long-distance or syntactic relevant contextual information to improve ABSA (Dong et al., 2014; Chen et al., 2017; He et al., 2018a; Huang and Carley, 2019; Sun et al., 2019; Zhang et al., 2019a; Wang et al., 2020; Chen et al., 2020b; Tang et al., 2020; Meng et al., 2020; Zhang and Qian, 2020; Tian et al., 2021a), where advanced architectures (e.g., GCN and GAT) are used to model aspect terms and their surrounding words. There are also studies tried to incorporate external knowledge through multi-task learning, where the information from the other tasks is supposed to offer instructions for ABSA (He et al., 2018b; Chen and Qian, 2020; Zhang et al., 2020). In addition, some other studies tried to explicitly concatenate the aspect term and the original sentence and considered such sentence-aspect pair as new model input to capture contextual information at the word- and sentence-level (Song et al., 2019; Zeng et al., 2019; Yang and Zeng, 2020; Phan and Ogunbona, 2020; Veyseh et al., 2020; Chen et al., 2020a). Compared with the approaches that directly adopt the original sentences as input, this straightforward method has been proved to be rather effective in promoting ABSA performance. However, it only focuses on the morphology level of the aspect term and lacks further understanding of its boundary and correspondent contextual information. To address the issue, Liang et al. (2019) propose a multi-task learning approach to learn the as-

pect term information by recovering the given aspect terms and predict the sentiment polarity at the same time.

Compared with previous studies, especially the ones that use multi-task learning to learn aspect term information through one single training stage, this paper offers an alternative to model aspect terms and their context. Specifically, our approach not only applies a discriminator to the first training stage to control the impact of complementary learning, but also use the two-stage-training strategy to force the main sentiment classifier to focus on the target ABSA task. Furthermore, since the second training stage exactly follows the standard supervised ABSA training, our model does not require extra input features (e.g., dependency parses) and thus is more efficient when processing large data.

6. Conclusion

In this paper, we propose an approach to enhance ABSA through complementary learning of aspect terms. Specifically, our approach has two training stages, where CL is applied to the first stage and serves as a supportive auxiliary task to enhance ABSA. CL improves the main sentiment classifier through modeling the aspect term along with its context through a decoding process that recovers the input sentence so as to implicitly return important contextual information from back-propagation. Furthermore, a discriminator is introduced to control the effect of CL on the main sentiment classifier to prevent the ABSA learning from being overwhelmed by the CL task. Experimental results and further analyses on five English benchmark datasets for ABSA illustrate the validity and effectiveness of the proposed approach, where our model outperforms strong baselines and achieves state-of-the-art on all datasets.

7. Acknowledgements

8. Bibliographical References

Baldini Soares, L., FitzGerald, N., Ling, J., and Kwiatkowski, T. (2019). Matching the Blanks: Distributional Similarity for Relation Learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2895–2905.

Barnes, J., Kurtz, R., Oepen, S., Øvrelid, L., and Velldal, E. (2021). Structured Sentiment Analysis as Dependency Graph Parsing. 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 3387–3402, Online, August.

Chen, Z. and Qian, T. (2019). Transfer capsule network for aspect level sentiment classification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 547–556.

Chen, Z. and Qian, T. (2020). Relation-aware collaborative learning for unified aspect-based sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3685–3694.

Chen, P., Sun, Z., Bing, L., and Yang, W. (2017). Recurrent attention network on memory for aspect sentiment analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 452–461.

Chen, C., Teng, Z., and Zhang, Y. (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.

Chen, G., Tian, Y., and Song, Y. (2020b). Joint Aspect Extraction and Sentiment Analysis with Directional Graph Convolutional Networks. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 272–279, December.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Diao, S., Bai, J., Song, Y., Zhang, T., and Wang, Y. (2020). ZEN: Pre-training Chinese Text Encoder Enhanced by N-gram Representations. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4729–4740, November.

Diao, S., Xu, R., Su, H., Jiang, Y., Song, Y., and Zhang, T. (2021). Taming Pre-trained Language Models with N-gram Representations for Low-Resource Domain Adaptation. 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 3336–3349, Online, August.

Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., and Xu, K. (2014). Adaptive recursive neural network for target-dependent twitter sentiment classification. In *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers)*, pages 49–54.

Du, C., Sun, H., Wang, J., Qi, Q., Liao, J., Xu, T., and Liu, M. (2019). Capsule network with interactive

attention for aspect-level sentiment classification. 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 5489–5498.

Fan, F., Feng, Y., and Zhao, D. (2018). Multi-grained attention network for aspect-level sentiment classification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3433–3442.

Gu, S., Zhang, L., Hou, Y., and Song, Y. (2018). A position-aware bidirectional attention network for aspect-level sentiment analysis. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 774–784.

He, R., Lee, W. S., Ng, H. T., and Dahlmeier, D. (2018a). Effective attention modeling for aspect-level sentiment classification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1121–1131.

He, R., Lee, W. S., Ng, H. T., and Dahlmeier, D. (2018b). Exploiting Document Knowledge for Aspect-level Sentiment Classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 579–585.

Helwe, C., Dib, G., Shamas, M., and Elbassuoni, S. (2020). A Semi-Supervised Bert Approach for Arabic Named Entity Recognition. In *Proceedings of the Fifth Arabic Natural Language Processing Workshop*, pages 49–57.

Herzig, J. and Berant, J. (2021). Span-based Semantic Parsing for Compositional Generalization. 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 908–921, Online, August.

Hu, M., Zhao, S., Zhang, L., Cai, K., Su, Z., Cheng, R., and Shen, X. (2019). CAN: Constrained Attention Networks for Multi-Aspect 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)*, pages 4593–4602.

Huang, B. and Carley, K. (2018). Parameterized convolutional neural networks for aspect level sentiment classification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1091–1096.

Huang, B. and Carley, K. M. (2019). Syntax-aware aspect level sentiment classification with graph attention 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 5472–5480.

Jiang, Q., Chen, L., Xu, R., Ao, X., and Yang, M. (2019). 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)*, pages 6280–6285.

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

Li, Y., Li, Z., and Zhang, M. (2020). Semi-supervised Domain Adaptation for Dependency Parsing via Improved Contextualized Word Representations. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3806–3817.

Liang, Y., Meng, F., Zhang, J., Xu, J., Chen, Y., and Zhou, J. (2019). A Novel Aspect-Guided Deep Transition Model 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)*, pages 5569–5580.

Liu, P., Qiu, X., and Huang, X. (2017). Adversarial Multi-task Learning for Text Classification. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1–10, Vancouver, Canada, July.

Liu, X., He, P., Chen, W., and Gao, J. (2019). Multi-Task Deep Neural Networks for Natural Language Understanding. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4487–4496, Florence, Italy, July.

Ma, D., Li, S., Zhang, X., and Wang, H. (2017). Interactive attention networks for aspect-level sentiment classification. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 4068–4074.

Mao, Q., Li, J., Wang, S., Zhang, Y., Peng, H., He, M., and Wang, L. (2019). Aspect-based sentiment classification with attentive neural turing machines. In *IJCAI*, pages 5139–5145.

Mass, Y., Carmeli, B., Roitman, H., and Konopnicki, D. (2020). Unsupervised FAQ Retrieval with Question Generation and BERT. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 807–812, Online, July.

Meng, F., Feng, J., Yin, D., Chen, S., and Hu, M. (2020). A structure-enhanced graph convolutional network for sentiment analysis. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 586–595.

Nie, Y., Tian, Y., Wan, X., Song, Y., and Dai, B. (2020). Named Entity Recognition for Social Me-

dia Texts with Semantic Augmentation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1383–1391, Online, November.

Ohashi, S., Takayama, J., Kajiwara, T., Chu, C., and Arase, Y. (2020). Text Classification with Negative Supervision. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 351–357, Online, July.

Pennington, J., Socher, R., and Manning, C. (2014). GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, October.

Phan, M. H. and Ogunbona, P. O. (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.

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., and Manandhar, S. (2014). SemEval-2014 task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35.

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., and Androutsopoulos, I. (2015). Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pages 486–495.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., et al. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *International workshop on semantic evaluation*, pages 19–30.

Qin, H., Chen, G., Tian, Y., and Song, Y. (2021a). Improving Arabic Diacritization with Regularized Decoding and Adversarial Training. In *Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*.

Qin, H., Chen, G., Tian, Y., and Song, Y. (2021b). Improving Federated Learning for Aspect-based Sentiment Analysis via Topic Memories. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3942–3954, Online and Punta Cana, Dominican Republic, November.

Qin, H., Tian, Y., and Song, Y. (2021c). Relation Extraction with Word Graphs from N-grams. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2860–2868, Online and Punta Cana, Dominican Republic, November.

Qin, H., Tian, Y., and Song, Y. (2022). Enhancing Relation Extraction via Adversarial Multi-task Learn-

ing. In *Proceedings of the 13th Language Resources and Evaluation Conference*.

Song, Y. and Shi, S. (2018). Complementary Learning of Word Embeddings. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 4368–4374, 7.

Song, Y., Shi, S., and Li, J. (2018). Joint Learning Embeddings for Chinese Words and Their Components via Ladder Structured Networks. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 4375–4381.

Song, Y., Wang, J., Jiang, T., Liu, Z., and Rao, Y. (2019). Attentional encoder network for targeted sentiment classification. *arXiv preprint arXiv:1902.09314*.

Song, Y., Zhang, T., Wang, Y., and Lee, K.-F. (2021). ZEN 2.0: Continue Training and Adaption for N-gram Enhanced Text Encoders. *arXiv preprint arXiv:2105.01279*.

Sun, K., Zhang, R., Mensah, S., Mao, Y., and Liu, X. (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.

Tabassum, J., Maddela, M., Xu, W., and Ritter, A. (2020). Code and Named Entity Recognition in StackOverflow. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4913–4926, Online, July.

Tang, D., Qin, B., Feng, X., and Liu, T. (2016a). Effective LSTMs for Target-Dependent Sentiment Classification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3298–3307.

Tang, D., Qin, B., and Liu, T. (2016b). Aspect level sentiment classification with deep memory network. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 214–224.

Tang, H., Ji, D., Li, C., and Zhou, Q. (2020). Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6578–6588.

Tian, Y., Song, Y., Xia, F., and Zhang, T. (2020a). Improving Constituency Parsing with Span Attention. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1691–1703, November.

Tian, Y., Song, Y., Xia, F., Zhang, T., and Wang, Y. (2020b). Improving Chinese Word Segmentation with Wordhood Memory Networks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8274–8285, July.

Tian, Y., Chen, G., and Song, Y. (2021a). Aspect-based Sentiment Analysis with Type-aware Graph

Convolutional Networks and Layer Ensemble. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2910–2922, Online, June.

Tian, Y., Chen, G., and Song, Y. (2021b). Enhancing Aspect-level Sentiment Analysis with Word Dependencies. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3726–3739, Online, April.

Veyseh, A. P. B., Nouri, N., Dernoncourt, F., Tran, Q. H., Dou, D., and Nguyen, T. H. (2020). Improving Aspect-based Sentiment Analysis with Gated Graph Convolutional Networks and Syntax-based Regulation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 4543–4548.

Wang, Y., Huang, M., Zhu, X., and Zhao, L. (2016). Attention-based LSTM for Aspect-level Sentiment Classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615.

Wang, K., Shen, W., Yang, Y., Quan, X., and Wang, R. (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.

Xu, H., Liu, B., Shu, L., and Yu, P. (2019). BERT post-training for review reading comprehension and aspect-based sentiment analysis. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2324–2335.

Xu, L., Bing, L., Lu, W., and Huang, F. (2020). Aspect sentiment classification with aspect-specific opinion spans. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3561–3567.

Yan, H., Dai, J., Qiu, X., Zhang, Z., et al. (2021). A Unified Generative Framework for Aspect-Based Sentiment Analysis. *arXiv preprint arXiv:2106.04300*.

Yang, H. and Zeng, B. (2020). Enhancing Fine-grained Sentiment Classification Exploiting Local Context Embedding. *arXiv preprint arXiv:2010.00767*.

Zeng, B., Yang, H., Xu, R., Zhou, W., and Han, X. (2019). Lcf: A local context focus mechanism for aspect-based sentiment classification. *Applied Sciences*, 9(16):3389.

Zhang, M. and Qian, T. (2020). Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3540–3549.

Zhang, C., Li, Q., and Song, D. (2019a). 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 4560–4570.

Zhang, H., Bai, J., Song, Y., Xu, K., Yu, C., Song, Y., Ng, W., and Yu, D. (2019b). Multiplex Word Embeddings for Selectional Preference Acquisition. 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 5247–5256, Hong Kong, China, November.

Zhang, C., Li, Q., Song, D., and Wang, B. (2020). A Multi-task Learning Framework for Opinion Triplet Extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 819–828.